

The Trendline manual for data collection and analysis of road safety performance indicators

Draft 5, July 2026



Co-funded by the European Union

The Trendline manual for data collection and analysis of road safety performance indicators

Wouter Van den Berghe, Jiří Ambros & Jan Elgner

Please refer to this manual as follows:

Van den Berghe, W., Ambros, J., Elgner, J. (2026). *The Trendline manual for data collection and analysis of road safety performance indicators*. Trendline programme, co-funded by the European Union. Brno (Czechia): Transport Research Centre

Main Author(s):

Wouter Van den Berghe, Tilkon Research & Consulting (Belgium)
Jiří Ambros, CDV – Transport Research Centre (Czechia)
Jan Elgner, CDV – Transport Research Centre (Czechia)

Acknowledgements

The Trendline programme is co-funded by the European Union; the authors are grateful to the European Commission for its support throughout the programme.

This manual builds on the work of the whole Trendline and Baseline communities, involving over 200 people across Europe. The authors thank the national data collection teams, the members of the KPI Expert Groups, and the partner organisations of the Trendline consortium for the involvement in the development and application of the methodological guidelines, the pilot studies and the practical experience on which this manual is based.

Key contributors to the finalisation of this manual:

Eva Kšicová, CDV (Czechia)
George Yannis, Anastasios Dragomanovits, Alexandra Laiou & Katerina Folla, NTUA (Greece)
Heike Martensen & Agnieszka Stelling, SWOV (The Netherlands)
Uta Meesmann & Nathalie Moreau, Vias institute (Belgium).

Other key contributors to the development of the methodologies discussed in the manual:

Peter Silverans, Sofie Boets, Naomi Wardenier & Nina Nuyttens, VIAS institute (Belgium)
Wendy Weijermars, Sander van der Kint & Letty Aarts, SWOV (The Netherlands)
Veronika Valentova, Lukáš Musil & Martin Šípek, CDV (Czechia)

All the co-authors of the methodological guidelines in Trendline are included in the References section.

Use of generative AI

Generative AI tools were used by the authors to support the drafting and editing of parts of this manual, in particular for restructuring and harmonising text drawn from the underlying methodological guidelines, for language editing, and for improving consistency of terminology. All methodological content, decisions and recommendations are those of the authors and contributors. The authors reviewed and verified the text and take full responsibility for its content.

Legal Disclaimer

All information in this document is provided "as is" and no guarantee or warranty is given that the information is fit for any particular purpose. The user, therefore, uses the information at its sole risk and liability. For the avoidance of all doubts, the European Commission has no liability in respect of this document, which is merely representing the authors' view.

ISBN 978-80-88583-01-1 (Online; pdf)
© Centrum dopravního výzkumu, v. v. i.
Brno 2026

Table of contents

Foreword	9
PART A: GENERAL FRAMEWORK	11
1 Introduction	12
1.1 <i>Why SPIs are important for road safety policy</i>	12
1.2 <i>The Trendline programme</i>	13
1.3 <i>The Trendline manual</i>	14
2 What are Safety Performance Indicators (SPIs)?	17
2.1 <i>Definition and conceptual framework</i>	17
2.2 <i>The focus on risky behaviour</i>	19
2.3 <i>The building blocks of an SPI</i>	20
2.4 <i>Selection criteria for SPIs</i>	21
2.5 <i>Scope of SPIs</i>	22
3 Terminology and definitions.....	23
3.1 <i>Overview</i>	23
3.2 <i>Core SPI concepts</i>	25
3.3 <i>Statistical and measurement terms</i>	26
3.4 <i>Data collection and sampling terms</i>	26
3.5 <i>Road network and vehicle terms</i>	27
3.6 <i>Temporal terms</i>	29
3.7 <i>Survey terms</i>	30
3.8 <i>Common abbreviations</i>	30
4 Using SPIs	32
4.1 <i>Reporting SPI results</i>	32
4.2 <i>Using SPIs in the policy cycle</i>	36
PART B – BEHAVIOURAL SPIs	40
5 Introduction to behavioural SPIs	41
5.1 <i>Why direct observation or roadside measurement is needed</i>	41
5.2 <i>Positioning SPI-oriented studies within behavioural research</i>	41
5.3 <i>General characteristics of behavioural SPI studies</i>	42
5.4 <i>Advantages and limitations of using human observers</i>	44
5.5 <i>Recurring challenges for SPI-oriented behavioural studies</i>	45
5.6 <i>Focus groups as a qualitative complement for behavioural SPIs</i>	45
6 Sampling strategy for behavioural SPIs	47
6.1 <i>Stratification and sampling of observations</i>	47
6.2 <i>Organisation and planning of observation sessions</i>	49
6.3 <i>Minimum numbers of observations</i>	51

7	Fieldwork: data collection and data recording.....	53
7.1	<i>Preparation of the fieldwork</i>	53
7.2	<i>Using observers for roadside measurements</i>	53
7.3	<i>Observation procedure</i>	55
7.4	<i>Road-side data recording</i>	57
7.5	<i>Camera-based and technology-based observation</i>	59
7.6	<i>Ethical and legal requirements</i>	60
7.7	<i>Collaboration with the police and other authorities</i>	62
8	Data analysis for behavioural SPIs	65
8.1	<i>Creation of the database with observation data</i>	65
8.2	<i>Weighting process</i>	68
8.3	<i>Estimation of SPI values</i>	74
8.4	<i>Estimation of confidence intervals</i>	75
8.5	<i>Using the data templates for data reporting</i>	77
9	Using self-reported data for SPIs	79
9.1	<i>Conceptual framework</i>	79
9.2	<i>Survey methodology</i>	82
9.3	<i>Sampling design for surveys</i>	83
9.4	<i>Question phrasing and indicator content</i>	84
9.5	<i>Data analysis</i>	87
10	SPIs for speeding	89
10.1	<i>Introduction</i>	89
10.2	<i>Indicator definition and theoretical population</i>	89
10.3	<i>Position within the behavioural framework</i>	91
10.4	<i>Scope</i>	91
10.5	<i>Sampling and stratification</i>	92
10.6	<i>Data collection</i>	95
10.7	<i>Variables, analysis and reporting</i>	96
10.8	<i>Minimum sampling and reporting requirements</i>	98
10.9	<i>Alternative measurement methods</i>	98
11	SPIs for restraint systems (seat belts and CRS)	101
11.1	<i>Introduction</i>	101
11.2	<i>Indicator definition and theoretical population</i>	101
11.3	<i>Position within the behavioural framework</i>	103
11.4	<i>Scope</i>	103
11.5	<i>Sampling and stratification</i>	104
11.6	<i>Data-collection specifics</i>	105
11.7	<i>Supplementary procedure: in-vehicle inspection of CRS (Method B)</i>	107
11.8	<i>Variables, analysis and reporting</i>	108
11.9	<i>Minimum sampling and reporting requirements</i>	110
12	SPIs for helmet use	112
12.1	<i>Introduction</i>	112
12.2	<i>Sampling and stratification</i>	113

12.3	<i>Data-collection specifics</i>	113
12.4	<i>Specific considerations by subcategory</i>	115
12.5	<i>Minimum sampling and reporting requirements</i>	119
12.6	<i>Using alternative measurement methods</i>	119
13	SPIs for distraction (by a handheld mobile device)	122
13.1	<i>Introduction</i>	122
13.2	<i>Sampling and stratification</i>	124
13.3	<i>Data-collection specifics</i>	125
13.4	<i>Variables, analysis and reporting</i>	126
13.5	<i>Minimum sampling and reporting requirements</i>	127
13.6	<i>Automatic detection via roadside cameras</i>	128
14	SPIs for driving under the influence of alcohol or drugs	130
14.1	<i>Introduction</i>	130
14.2	<i>Roadside testing by the police</i>	132
14.3	<i>Self-reported surveys</i>	135
14.4	<i>Random saliva testing without police involvement</i>	136
14.5	<i>Variables, analysis and reporting</i>	136
14.6	<i>Minimum sampling and reporting requirements</i>	137
14.7	<i>Beyond the legal limit: complementary indicators</i>	139
15	SPIs for compliance at pedestrian crossings and intersections	141
15.1	<i>Introduction</i>	141
15.2	<i>Indicator definition and theoretical population</i>	141
15.3	<i>Position within the behavioural framework</i>	143
15.4	<i>Scope</i>	143
15.5	<i>Sampling and stratification</i>	143
15.6	<i>Data-collection specifics: observation approach for each of the four SPIs</i>	145
15.7	<i>Variables, analysis and reporting</i>	148
15.8	<i>Minimum sampling and reporting requirements</i>	149
16	SPIs on use of lighting by cyclists in the dark.....	150
16.1	<i>Introduction</i>	150
16.2	<i>Sampling and stratification</i>	151
16.3	<i>Data-collection specifics</i>	152
16.4	<i>Variables, analysis and reporting</i>	154
16.5	<i>Minimum sampling and reporting requirements</i>	155
	PART C – DATABASE-BASED SPIs	156
17	Common framework for database-based SPIs.....	157
17.1	<i>Introduction</i>	157
17.2	<i>General principles</i>	158
17.3	<i>Data sources and quality issues</i>	159
17.4	<i>Data processing</i>	161
17.5	<i>Metadata requirements</i>	162
17.6	<i>Alternative approaches</i>	162

18	SPIs for infrastructure safety	164
18.1	<i>Introduction</i>	164
18.2	<i>Indicator definition and scope</i>	164
18.3	<i>Data requirements and sources</i>	165
18.4	<i>Estimation of the SPI</i>	167
18.5	<i>Minimum reporting requirements</i>	168
18.6	<i>Complementary indicators and alternative approaches</i>	170
19	SPIs on 30 km/h roads in urban areas	172
19.1	<i>Introduction</i>	172
19.2	<i>Indicator definition and scope</i>	172
19.3	<i>Data requirements and sources</i>	173
19.4	<i>Estimation of the SPI</i>	174
19.5	<i>Minimum reporting requirements</i>	174
19.6	<i>Complementary indicators and alternative approaches</i>	175
20	SPIs for vehicle safety	177
20.1	<i>Introduction</i>	177
20.2	<i>Indicator definition and scope</i>	177
20.3	<i>Data requirements and sources</i>	178
20.4	<i>Estimation of the SPI</i>	179
20.5	<i>Minimum reporting requirements</i>	180
20.6	<i>Complementary indicators and alternative approaches</i>	181
21	SPIs for post-crash care	183
21.1	<i>Introduction</i>	183
21.2	<i>Indicator definition and scope</i>	183
21.3	<i>Data requirements and sources</i>	184
21.4	<i>Estimation of the SPI</i>	184
21.5	<i>Minimum reporting requirements</i>	186
21.6	<i>Complementary indicators and additional information</i>	186
22	SPIs for enforcement of traffic regulations	188
22.1	<i>Introduction</i>	188
22.2	<i>Indicator definition and scope</i>	188
22.3	<i>Data requirements and sources</i>	189
22.4	<i>Estimation of the SPIs</i>	190
22.5	<i>Minimum reporting requirements</i>	191
22.6	<i>Complementary indicators and alternative approaches</i>	191
	PART D: IMPLEMENTATION GUIDANCE.....	193
23	Planning and managing an SPI study	194
23.1	<i>The SPI study as a project</i>	194
23.2	<i>Subcontracting and the division of work</i>	195
23.3	<i>Planning, timeline and resources</i>	197
23.4	<i>Quality assurance</i>	199
23.5	<i>Reporting and delivering the data to Trendline</i>	200

23.6	<i>The appendices and how to use them</i>	201
REFERENCES		203
	<i>Methodological guidelines developed in Baseline and the first phase of Trendline</i>	203
	<i>Scientific publications referred to in the Manual</i>	204
APPENDICES		207
Appendix A — Data templates and guidance		208
	<i>SPI-specific template structure</i>	208
	<i>Structure of roadside-based templates</i>	208
	<i>Structure of desk-based templates</i>	209
	<i>Completing the template</i>	209
	<i>Common errors to avoid</i>	209
Appendix B – Worked out example		210
	<i>Data structure</i>	210
	<i>Survey design specification</i>	210
	<i>SPI values and confidence intervals</i>	211
24 Appendix C – Checklists		213
	<i>Checklist 1 — Selecting and target-setting SPIs</i>	213
	<i>Checklist 2 — Sampling and selection of observation locations</i>	214
	<i>Checklist 3 — Observation-site requirements</i>	216
	<i>Checklist 4 — Observer selection, training and quality control</i>	218
	<i>Checklist 5 — Questionnaire and survey design</i>	219
	<i>Checklist 6 — Ethics, privacy and legal requirements</i>	220
	<i>Checklist 7 — Data processing and analysis</i>	221
	<i>Checklist 8 — Data sources and data access</i>	222
	<i>Checklist 9 — Data linkage and coverage</i>	223
	<i>Checklist 10 — Metadata and data delivery</i>	224
	<i>Checklist 11 — Reporting and communicating results</i>	226
	<i>Checklist 12 — Subcontracting and tender specification</i>	227
Appendix D — Frequently asked questions for field observers		228
	<i>A. Before you start</i>	228
	<i>B. What to record and what counts</i>	228
	<i>C. Borderline cases</i>	228
	<i>D. Difficult conditions</i>	229
	<i>E. When you are not sure</i>	229
	<i>F. Talking to the public and the police</i>	229
	<i>G. After the session</i>	229
Appendix E – Equipment specifications		230
	<i>A. Introduction</i>	230
	<i>B. Speed-measurement equipment</i>	231
	<i>C. Breath-testing equipment for alcohol (Random Breath Testing)</i>	232
	<i>D. Oral-fluid (saliva) testing equipment for drugs</i>	233

<i>E. Observation cameras and automated image analysis</i>	233
<i>F. Light-level measurement for the cyclist-lighting indicator</i>	235
<i>G. Data-collection (coding) devices</i>	235
<i>H. Traffic-counting equipment</i>	235
<i>I. Observer safety and personal protective equipment</i>	235

Foreword

Road safety is no longer measured solely by counting fatalities and serious injuries. Indeed, crash data alone are insufficient for proactive road-safety management. Fatality counts react slowly to policy interventions, are statistically volatile in smaller jurisdictions or for specific road-user groups, and do not by themselves explain why crashes occur. Within the Safe System framework road safety is increasingly managed proactively, on the basis of indicators that describe the quality of the processes, behaviours and system characteristics that lead to crash outcomes. These indicators are commonly referred to as Safety Performance Indicators, or SPIs. SPIs complement crash data by capturing measurable risk factors before these factors translate into changes in injury statistics. They also typically require fewer resources to collect than detailed crash investigations, and their results are intuitively understandable to non-specialist audiences, which makes them well suited for use in communication and awareness work. Within the EU monitoring framework SPIs are also called Key Performance Indicators (KPIs); the two terms are used interchangeably in this manual.

This manual provides a harmonised methodological framework for measuring and reporting SPIs, more specifically in the context of the Trendline programme of the European Union. Although participants in the Trendline programme are the primary target group, most parts of it will also be useful beyond the Trendline programme and outside the European Union for anyone who is interested in designing SPIs, collecting data and analysing it.

The manual integrates previously separate methodological guidelines of the Trendline programme into a single coherent framework structured around two complementary pillars of measurement. The first pillar covers observed behavioural SPIs, which measure the real-time behaviour of road users in traffic through roadside observations and direct measurements. These indicators include, for example, seat belt use, distraction by mobile phone, alcohol-impaired driving, and compliance with traffic rules. Because such behaviours cannot be observed for every road user, representative sampling and statistical weighting are required. The manual provides harmonised procedures for sampling design, fieldwork organisation, data coding, and statistical estimation, followed by indicator-specific chapters for each behavioural SPI. This pillar also includes methods for collecting and analysing self-reported data on behaviour and attitudes.

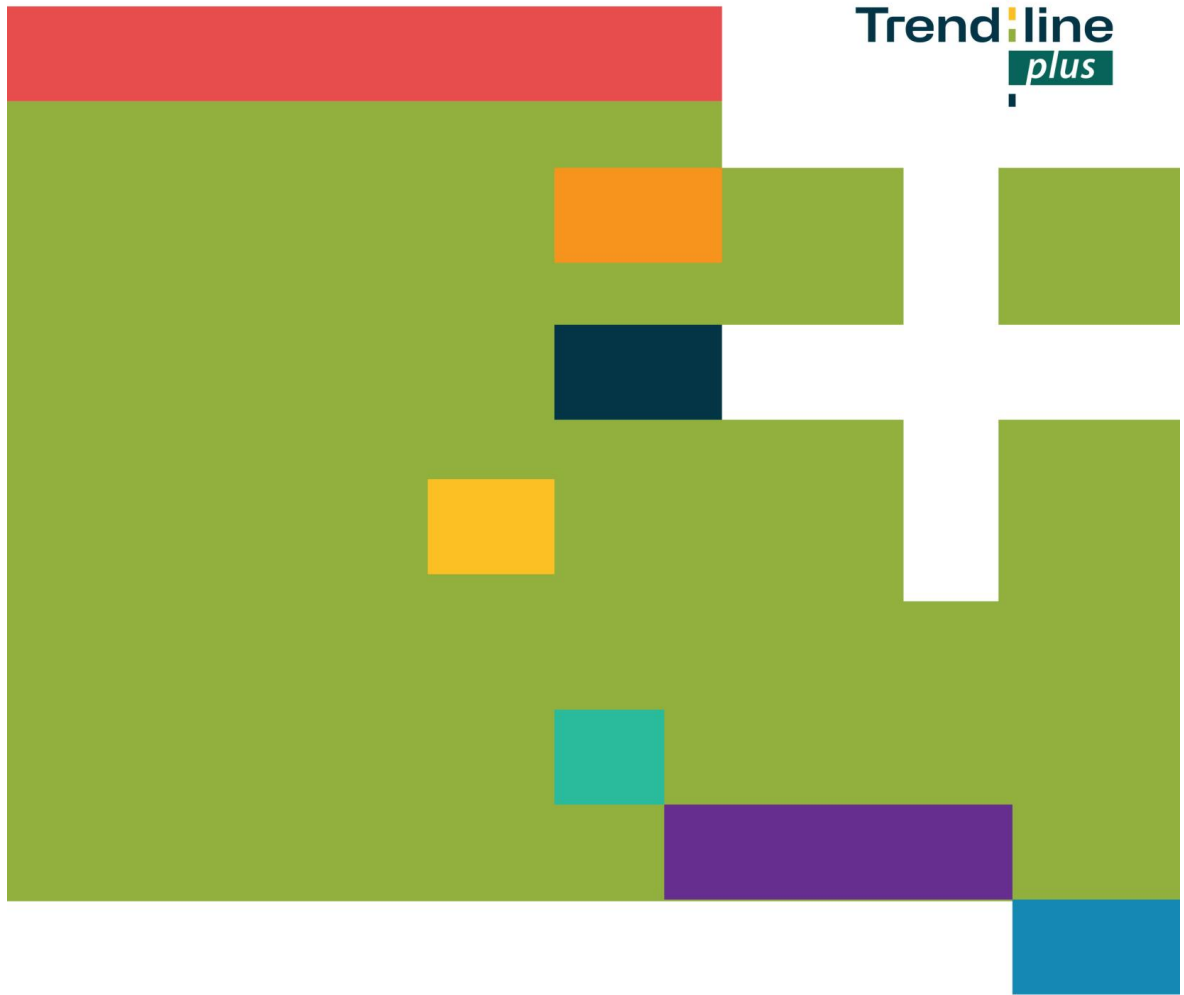
The second pillar addresses database and register-based SPIs, which describe structural and systemic safety conditions rather than individual behaviour. These indicators are derived from administrative or operational datasets, such as road infrastructure inventories, vehicle registration systems and enforcement records. Examples of such SPIs include the share of safe roads in urban environments, emergency response times after crashes, and the level of enforcement. Here, the key methodological challenges are not sampling but data coverage, harmonisation of definitions, linkage of data sources, and transparent documentation of limitations. The manual establishes a common framework for preparing, validating, and aggregating such data before presenting detailed procedures for each SPI.

Across both pillars, the manual emphasises several overarching principles: representativeness, standardisation, repeatability over time, and transparent metadata. Statistical uncertainty is addressed differently depending on the SPI type: sampling variability for observational and survey SPIs, and coverage and definition bias for database SPIs. The manual provides guidance on reporting results, interpreting differences between countries and over time, and using SPIs within the road safety policy cycle.

This manual builds on the work of the whole Trendline and Baseline communities, involving over 200 people across Europe. The authors thank the national data collection teams, the members of the KPI Expert Groups, and the partner organisations of the Trendline consortium for the involvement in the development and application of the methodological guidelines, the pilot studies and the practical experience on which this manual is based.

This is the first version of the manual, replacing the 17 existing Methodological Guidelines for SPIs as well as the statistical note. It is released at the beginning of the second phase of Trendline. An updated version of the manual is planned to be published at the end or just after the second phase (end of 2027 or early 2028) to include experience gained, in particular with the new experimental indicators. This approach ensures that the Trendline framework remains responsive to new developments and needs.

Wouter Van den Berghe, Jiří Ambros, & Jan Elgner, July 2026



PART A: GENERAL FRAMEWORK

1 Introduction

1.1 Why SPIs are important for road safety policy

The methodology presented in this manual represents a contemporary European approach to road safety monitoring. Historically, the evaluation of road safety performance was predominantly characterized by a reactive approach, where the primary risk metrics were the counts of fatalities and serious injuries resulting from road traffic collisions. While these "lagging indicators" remain the ultimate measure of the social and economic impact of road crashes, they are increasingly recognized as insufficient for the proactive management of modern transport systems. (Rao and Bharadwaj, 2026)

Traditional reliance on crash data as the sole instrument for road safety management presents several critical limitations that the Trendline framework can mitigate:

- **Statistical Volatility:** In many jurisdictions, particularly those with smaller populations or lower absolute numbers of fatalities, annual fluctuations in crash data are often driven by random chance rather than systemic changes in safety quality. In addition, crash data are often underreported, particularly for minor injuries and property damage incidents. This "noise" can lead to erroneous conclusions regarding the effectiveness of safety interventions.
- **Absence of Diagnostic Insight:** Mortality and morbidity data provide evidence that a failure occurred within the system, but they rarely offer immediate insight into the specific behavioural or infrastructural deficiencies that preceded the event.
- **Delayed Policy Response:** By definition, result indicators are available only after a crash has occurred. In a proactive safety culture, the objective is to identify and rectify systemic vulnerabilities before they result in physical harm.

The transition toward the use of Safety Performance Indicators (SPIs) is fundamentally rooted in the Safe System philosophy (ETSC, 2001). This approach, which serves as the theoretical foundation for the European Union's targets (the elimination of all road deaths and serious injuries by 2050), requires that the transport system must be engineered to account for human fallibility. The Safe System dictates that if one component of the system fails, such as a driver making a judgmental error, other layers of protection, such as vehicle safety features or forgiving infrastructure, must intervene to prevent a fatal outcome. To monitor the integrity of these protective layers, policymakers require a robust set of intermediate metrics.

As established in the EU Road Safety Policy Framework 2021–2030 (European Parliament, 2022), SPIs serve as the primary mechanism for monitoring progress toward intermediate safety targets. They function as a "thermometer" for the system, providing a real-time assessment of safety levels long before the consequences are reflected in fatality statistics.

A note on the SPI versus KPI terminology

The terms 'Safety Performance Indicator' (SPI) and 'Key Performance Indicator' (KPI) are used in essentially the same sense in this manual and in the wider road-safety literature. The term 'Safety Performance Indicator' is the older of the two and is the one used in the SafetyNet project and in much of the academic literature. The term 'Key Performance Indicator' was introduced by the European Commission in its EU Road Safety Policy Framework (European Commission, 2020), and is therefore the term used in the Baseline and Trendline programmes that operationalise that framework. The two terms refer to the same class of indicators and may be treated as interchangeable. Where this manual uses SPI, it can equally well be read as KPI.

1.2 The Trendline programme

The current framework is the direct successor to the **Baseline** project (<https://baseline.vias.be>, 2020–2022), which served as the initial attempt to implement a common methodology for eight core KPIs across 18 EU Member States:

- Speeding
- Use of seat belts and child restraint systems
- Use of protective helmets by riders of bicycles and powered two-wheelers
- Driving under the influence of alcohol
- Distraction by a mobile device
- Vehicle safety
- Road infrastructure safety
- Post-crash response.

For each KPI, a specific KPI Expert Group (KEG) was established which developed the methodological guidelines and also verified and interpreted the collected KPIs. More information is available in the Baseline final report (Silverans and Vanhove, 2023).

While the Baseline project successfully demonstrated the feasibility of cross-border data harmonization, it also identified a need for more granular definitions and broader indicator coverage to reflect the complexity of modern mobility. The **Trendline** programme (<https://trendlineproject.eu>, 2022–2027, now in its second phase) builds upon the foundation of Baseline by:

- refining the statistical rigor of the original eight indicators
- developing and testing "experimental" indicators to address emerging safety concerns
- expanding the consortium to include 25 EU Member States in Phase 1 and 23 in Phase 2
- analysing and discussing the integration of SPIs in the policy cycle.

In addition, ten complementary "experimental" indicators and methodologies were developed and tested within Trendline (2022–2025):

- Driving under the influence of drugs
- 30km/h on urban roads
- Compliance with traffic rules on signalized pedestrian crossings and intersections
- Compliance with traffic rules on unsignalized pedestrian crossings and intersections
- Helmet wearing by personal mobility devices (PMD) riders
- Self-report behaviour
- Attitudes
- Light use by cyclists in the dark
- Enforcement of traffic regulations
- Alternative speeding KPIs.

As in Baseline, for each of the original eight KPIs and the experimental KPIs, a KEG was established, consisting of European experts. The main role of the KEGs was to draft the common methodological guidelines, to answer methodological questions, and to supervise the pilot tests of the new methodologies. More information is available in the Trendline final report (Van den Berghe and Stelling, 2025). In the second phase of Trendline, almost all the experimental indicators for which a methodology was developed in the first phase, were taken on board and became eligible for EU grants.

In its current, second phase (2026–2027), the Trendline programme covers twelve indicator clusters, of which 10 are eligible for EU grants (those not eligible for funding are marked in *italics*):

- Speeding
- Use of seat belts and child restraint systems
- *Helmet use*

- Driving under the influence
- Distraction by a mobile device
- Compliance with traffic rules at pedestrian crossings and intersections
- *Use of lighting by cyclists in the dark*
- Vehicle safety
- Post-crash care
- 30 km/h on urban roads
- Enforcement
- *Infrastructure safety*

The methodologies for all these SPIs are described in this manual.

The infrastructure indicator occupies a particular place in this set. Unlike the others it rests on an established regulatory framework, deriving its scope and assessment method from the Road Infrastructure Safety Management Directive (2019/1936/EC) and the Network-Wide Road Safety Assessment methodology endorsed by the Expert Group on Road Infrastructure Safety (EGRIS). Because that framework covers mainly motorways and primary roads and excludes most urban roads, the safety of the urban network is addressed separately, in particular through the experimental 30 km/h indicator.

The objectives of the Trendline programme are:

- to collect, harmonise and report a common set of safety performance indicators across the EU
- to consolidate and refine the methodologies for the existing indicators and to develop, pilot and mainstream new indicators for emerging safety concerns
- to reduce methodological divergence and increase the comparability of national data
- to support the integration of SPIs into the road safety policy cycle at national and European level, while building capacity for SPI data collection, including beyond the European Union.

A key secondary objective of Trendline is to reduce methodological divergence and increase comparability between Member States. Historically, national data on safety behaviours were often incomparable due to variations in sampling techniques, road definitions, and observational protocols. Using the methodologies that are described in this manual, will contribute to this goal. These methodologies have been based on good national practice and are rooted in European initiatives such as the SafetyNet project (Thomas et al., 2009).

Although the methodologies established under Trendline were primarily targeted at EU countries, they have already demonstrated significant utility beyond the borders of the European Union. The framework has been successfully applied within the Western Balkans via the Westbelt project (Transport Community, 2024) and has provided a template for SPI data collection in nations such as Tajikistan and Moldova. This global relevance underscores the framework's robustness and its alignment with international best practices in safety management.

1.3 The Trendline manual

1.3.1 Purpose, scope and structure of the manual

This manual is designed as the primary technical reference for the collection, analysis, and reporting of safety performance data. It provides a standardized protocol to be followed by all participating entities to ensure the integrity of the resulting European database. The manual addresses the technical and operational needs of several distinct user groups within the road safety ecosystem: national road safety authorities, fieldwork and research organizations, and data analysts and statisticians. For the European Commission the manual serves as the benchmark against which national data submissions are validated. As stated in the Foreword, the manual can also be used outside the European Union as a reference document for anyone who is interested in designing SPIs, collecting data and analysing it.

To facilitate ease of use, the manual is structured into four parts, each focusing on a specific dimension of the safety performance monitoring process (Table 1-1). Practitioners can use individual chapters as standalone references for specific tasks, such as designing a sampling plan, training observers, or analysing roadside observation data. The document as a whole provides an integrated framework that ties these tasks back to the underlying purpose of producing comparable, reliable, policy-relevant SPIs.

Table 1-1. Structure of the Trendline manual

Part	Title	Operational scope
A	General framework	Establishes the concepts, including fundamental definitions of road categories, vehicle types, and the principles of statistical representativeness.
B	Behavioural SPIs	Provides detailed protocols for indicators derived from direct observation of risky behaviour of different road users. Requires roadside observers or automated sensors (radar, lidar, camera). The main challenge is ensuring free-flow conditions for speed and representative sampling across road types and time periods. Self-reported data can also be considered in certain cases.
C	Database-based SPIs	Outlines methodologies for safety indicators derived from administrative and technical registers. The primary challenges are coverage of databases, the linkage between different databases and international comparability.
D	Implementation guidance	Offers practical guidance on the execution of the project cycle, from data collection planning to the final delivery of results and metadata using standardized templates.

1.3.2 Some underlying principles and strategic choices

The development of the Trendline framework necessitated several strategic choices to balance the competing requirements of scientific precision, logistical feasibility, and international comparability.

The first choice is that the mandatory level of reporting for the Trendline programme is the national level. The primary goal is to provide values for indicators that characterizes the safety state of the entire Member State for a particular domain. However, recognizing that road safety dynamics often vary significantly between urbanized and rural regions, Member States are encouraged to perform **regional stratification** (typically at the NUTS-2 or NUTS-3 levels). In such cases, the requirements for sample size and stratification of locations need to be applied to each region for which the indicator value is calculated (see Chapter 6).

To ensure a sufficient level of comparability across countries, a second choice is that for every SPI the manual specifies a set of **mandatory minimum requirements**. These requirements represent the "floor" of methodological compliance and include: typically:

- mandatory road stratification
- minimum vehicle types or road user categories
- minimum sample sizes and minimum number of observation sites
- requirements regarding the timing of observations.

However, countries can also include additional elements, such as more granular vehicle classifications or increased sample sizes, provided that the core mandatory data remains extractable and harmonized.

A third principle was that the methodologies must be technically and legally **implementable for all EU Member States**, including countries with little SPI experience and severe budget constraints. This has led to

- relatively low minimum sample sizes (compared to good practice in countries with a long experience)
- the possibility to use alternative methodologies or SPI definitions in certain cases (e.g. using self-reported data for driving under the influence of alcohol, instead of breathalyser-based observations)

- acceptance by the EU of small deviations from the methodology (e.g. to ensure consistency with existing time series or to accommodate for small samples).

By focusing on indicators that are practical to collect on a recurring annual or biennial basis, the framework ensures the sustainability of the monitoring program over the long term.

A final component of the Trendline methodological choices is the focus on standardized data delivery. To ensure that data from diverse sources can be integrated into a single European database, the project utilizes standardized Excel reporting templates, which should be considered as an Annex to this Manual. They can be found on <https://trendlineproject.eu/templates>. These templates serve several critical functions, such as data validation, inclusion of metadata and data necessary for weighting. Underlying these templates are also harmonised definitions of road categories, vehicle categorisations and temporal characteristics (such as “night” and “weekend”); these are provided in Chapter 3.

2 What are Safety Performance Indicators (SPIs)?

This chapter is based on a number of sources, in particular a thematic report from the European Commission (2022), a book chapter (Van den Berghe et al., 2025) and a report (Hakkert & Gitelman, 2007). The reader who is interested in learning more about SPIs in road safety is referred to these sources.

2.1 Definition and conceptual framework

Safety Performance Indicators (SPIs; also frequently referred to as Key Performance Indicators (KPIs) in the context of the European Commission's monitoring framework) are measurable variables that represent the operational safety performance of the road traffic system (Hakkert and Gitelman, 2007). An SPI is not a measure of the final outcome (deaths or injuries) but a measure of the quality of the processes and conditions that lead to those outcomes. This distinction is fundamental: while crash statistics record failures that have already occurred, SPIs reveal the state of the system before those failures materialise, offering a window for preventive action.

An SPI may quantify one of four broad phenomena:

- It may measure the prevalence of a risk-related behaviour, for example, the percentage of drivers exceeding the legal blood-alcohol concentration (BAC) limit.
- It may capture the presence of a protective behaviour or safety feature, for example, the percentage of cyclists wearing helmets.
- It may express the safety quality of a system component, for example, the percentage of vehicle-kilometres travelled on roads with a high safety rating.
- It may reflect the performance of post-crash response systems, for example, the median emergency response time from the moment of collision to the arrival of professional medical care.

Whatever the domain, an SPI must possess three essential attributes: a clearly defined population to which it applies, a clearly defined safety-relevant characteristic that it measures, and a consistent measurement or calculation method that can be replicated over time and across jurisdictions.

The conceptual framework underpinning the Trendline programme sees road safety as a “production process.” The model consists of inputs, throughputs, and outputs (see also Figure 2-1):

- Inputs are the actions and resources committed to safety. Examples are setting speed limits, lowering the legal blood-alcohol-concentration limit, building new pedestrian crossings, running an awareness campaign, or mandating new vehicle-safety technology.
- Throughputs or processes, are processes like speed compliance, seat belt use, infrastructure quality, vehicle safety that increase or reduced safety risks. SPIs measure the performance of these processes. Examples are a reduction in average travel speed, an increase in the share of motorcyclists wearing a helmet, or a shortening of the time between emergency call and ambulance arrival.
- Outputs or outcomes, are the final safety results and their impact. Examples are a reduction in the number of speed-related crashes, in alcohol-related fatalities, in head injuries among riders, or in preventable deaths at the roadside.

The three stages are causally linked, even though the strength of the link depends on the specific action and on the wider context in which it is implemented.

Indicators can be defined at each of the three stages, and a country's monitoring system typically uses indicators at all three. Examples of input indicators include the costs for improving the safety of roads and the number of speed checks per kilometre. Throughput indicators are the SPIs themselves, in the narrow technical

sense used throughout this Manual. Output indicators are the road-safety result indicators: fatalities, serious injuries, and road crashes.

Figure 2-1. Inputs, processes and outputs as a basis for road safety indicators.



Different authors have used the term “safety performance indicator” with slightly different boundaries. Most experts (see Van den Berghe et al., 2025), restrict the term to indicators of the causal factors of crashes such as road-user behaviour, infrastructure safety, and vehicle safety. Others take a broader view and include some elements of the policy and implementation layers, in other words, input factors, in their definition. This manual predominantly follows the convention adopted by the SafetyNet project and by the European Commission: SPIs (or, equivalently, KPIs) describe the operational conditions of the road-traffic system that influence its safety performance. They are throughput factors, sitting between policy actions on the one side and ultimate impact on the other. Yet it is pointed out that some of the SPIs included in this manual can be considered input indicators (30 km/h in urban areas; enforcement activities) rather than process indicators.

The phenomena measured by SPIs should have a proven causal links to crash frequency or severity. For the SPIs included in Trendline, this has amply been proven. For example, for speeding the Power Model has demonstrated that a 1% increase in mean speed typically results in a 2% increase in injury crashes, a 3% increase in severe crashes, and a 4% increase in fatal crashes (OECD/ITF, 2018). For protective equipment, the use of seat belts and helmets does not prevent the occurrence of a crash but drastically alters the energy transfer to the human body, shifting an outcome from fatal to minor injury. The causal mechanism here is biomechanical rather than behavioural, but it is no less strong for that. For alcohol, the risk of a crash increases exponentially with Blood Alcohol Concentration (BAC), particularly as it crosses the 0.5 g/l threshold (Hels et al., 2011). The relationship is steep enough that even modest reductions in the share of drivers exceeding the legal limit produce measurable reductions in alcohol-related crashes.

Policymakers are a prime target group of SPIs. SPIs allow them to:

- monitor whether risk factors (e.g. speeding, alcohol, distraction) are increasing or decreasing
- track the implementation and effectiveness of safety interventions
- identify emerging safety problems before they are fully reflected in crash data.

SPIs are also useful because they are often directly linkable to interventions and policy measures. A speed enforcement campaign, for instance, finds its natural performance measure in a speed compliance SPI; drink-driving enforcement is tracked through an alcohol SPI; and improvements in emergency medical services through a post-crash response SPI. Thus, SPIs function as intermediate performance measures in the road safety policy cycle, bridging the gap between the adoption of a policy and its eventual effect on deaths and injuries. Without such intermediate measures, policymakers would be forced to wait years for crash data to confirm whether their interventions had worked, by which time opportunities for course correction might have been lost.

Because they sit upstream of crashes, SPIs provide a “leading” rather than a “lagging” perspective on road safety. Fatality data tells us what happened in the past; SPI data identifies where the system is currently

failing, allowing policymakers to intervene proactively before a crash occurs. A sustained rise in the percentage of drivers exceeding the speed limit on a particular road type, for instance, signals heightened risk months before any corresponding rise in crash figures becomes statistically visible. This predictive quality makes SPIs indispensable for modern safety management.

2.2 The focus on risky behaviour

The most frequently used SPIs measure risky behaviour. Risky behaviour in traffic refers to actions or omissions of road users that increase the likelihood or the severity of a crash, whether intentionally or unintentionally. It involves deviating from safety norms or legal requirements such as not respecting the speed limit, using a handheld mobile phone while driving, driving under the influence of alcohol, or failing to wear a seat belt or a helmet. Such behaviours raise the level of exposure to danger not only for the individual concerned but also for other road users, and some may have more serious consequences for third parties than for the person taking the risk.

Risky behaviour of road users typically arises from one of three families of causes, and the design of behavioural SPIs needs to take all three into account:

- Intentional risky behaviour: drivers who knowingly speed, who drive aggressively, or who choose to drive after drinking.
- Unintentional risky behaviour, where the road user does not deliberately violate the rules but loses attention, is dazzled by oncoming lights, makes a perceptual error, or is unfamiliar with a particular vehicle function.
- System-level risk, where external factors such as inconsistent infrastructure, missing or confusing road signs, or inadequate maintenance push the road user towards a risky action even when intentions are good.

A comprehensive SPI system must be sensitive to all three sources of risk, because the appropriate policy response differs in each case: intentional behaviour may call for enforcement, unintentional behaviour for design improvements, and system-level risk for infrastructure investment.

There is a direct, well-documented link between several of these behaviours and the number and severity of crashes. The link is particularly strong for speeding, for driving with a substantial alcohol intake, and for not wearing a seat belt or a helmet. For other behaviours the relationship is weaker or more variable: a limited alcohol intake (for example, one drink), or use of a hands-free phone while driving, both elevate risk but do so to a lesser extent than their more extreme counterparts. SPI design must therefore distinguish between behaviours with a strong, consistent causal link to crashes and those where the link is more contingent, because the strength of the link determines how confidently an SPI can be used to predict safety outcomes.

It is important to distinguish risky behaviour from illegal behaviour. Not all risky behaviour is covered by traffic regulations or can be enforced. Driving while extremely tired, for instance, is dangerous but in most jurisdictions cannot be enforced directly. Conversely, behaviour that is legal can sometimes be risky: driving at the speed limit during heavy rain or fog can be perfectly legal yet hazardous, and a hands-free phone conversation while driving is legal in most countries even though it carries a measurable distraction risk. There are also situations in which illegal behaviour is only marginally risky, such as a pedestrian who is crossing at a red light when there is no traffic. SPIs concentrate on those forms of risky behaviour for which there is a documented causal link with crash occurrence or severity, and for which a reliable measurement methodology exists. This pragmatic focus ensures that SPIs remain both scientifically defensible and operationally useful.

Risky behaviours of road users are typically grouped by the road-user category to which they apply. The risky behaviours most frequently seen in Europe are shown in Table 2-1. Many of these risks can be measured through SPIs and are included in this manual.

Table 2-1. Typical risky behaviours per road-user group

Road-user group	Typical risky behaviours
Drivers of cars, vans, lorries and buses	Speeding; aggressive driving; risky overtaking; failure to use appropriate lights at night; red-light running; driving while tired; driving under the influence of alcohol or drugs; using a handheld mobile phone; not wearing a seat belt
Motorcyclists (PTW riders)	Speeding; not wearing a helmet (rider or passenger); risky lane-filtering
Cyclists	Riding on the pavement; not wearing a helmet; not using lights in the dark
Pedestrians	Crossing on a red light; walking on the carriageway; walking at night without being visible

Risky behaviour cannot be eliminated by a single type of intervention. Effective road-safety policy combines four broad families of measures:

- road infrastructure design and maintenance (for example, speed-calming features and well-designed intersections)
- vehicle design (for example, audible seat belt reminders and automatic emergency braking)
- education and awareness
- traffic regulations in combination with their enforcement.

Behavioural SPIs are the principal instrument for monitoring how well this combination of measures is working in practice for the specific behaviours that drive crash risk. They provide a regular, standardised health check on the behavioural component of the road traffic system.

2.3 The building blocks of an SPI

Each SPI is constructed from four building blocks that together specify what is being measured and how the measurement is to be interpreted: the topic, the metric, the target, and the indicator value.

- The **topic** is the risk-related phenomenon that the SPI describes. Examples are speed compliance, seat belt use, mobile-phone distraction, or driving under the influence of alcohol. A single topic can in principle be addressed by several different SPIs, each capturing a different facet of the same underlying risk.
- The **metric** is the physical dimension or unit in which the indicator value is expressed. Alternative terms that can be used instead of “metric” are measure, yardstick, or gauge. Metrics may be absolute counts (for example, the number of fatalities or crashes), fractions or percentages (for example, the percentage of drivers exceeding the speed limit), or other physical quantities (for example, vehicle-kilometres travelled). For a single topic, several metrics are usually possible. Speed, for instance, can be expressed as the average speed of vehicles, as the percentage of drivers exceeding the legal speed limit, or as the speed difference between the fastest and slowest vehicles in a given location. The choice of metric has direct consequences for what the indicator can detect and for how it is interpreted. An average-speed metric may mask dangerous extremes, while a percentage-over-limit metric highlights precisely those extremes but says nothing about overall traffic flow.
- The **target** is the value or the range of values against which the indicator value is compared. Alternative terms that can be used instead of “target” are norm, standard, or benchmark. The target has the same dimension as the metric and is often built directly into the definition of the indicator. The example “percentage of drivers driving faster than 30 km/h in urban areas” illustrates this clearly: the topic is speed in urban areas, the metric is the percentage of drivers exceeding the norm, and 30 km/h is the target.

Without targets it is difficult to assess progress. For example, whether a change from 80% to 85% is an excellent result, depends on the target: if it is 84%, it is a great result but if the target is 98%, the result is poor.

- The **indicator value** is the number that is calculated for the topic in a given context. An example is: “94% of front-seat passengers are wearing a seat belt.” This value is generally the (weighted) average of a range of cases or observations and shares the same dimension as the metric and the target. Breakdowns of the indicator value into subpopulations (by road type, vehicle type, gender, age group, region, time of day, part of the week) are often essential to understand the underlying phenomenon and to identify where intervention is most needed. A national seat belt-wearing rate of 94% may conceal a rate of 78% among young male drivers on rural roads at night, and that breakdown points directly to the population and context where a targeted campaign would have the greatest effect.

2.4 Selection criteria for SPIs

Several criteria can be used to identify good and useful Safety Performance Indicators. These criteria are listed and described in Table 2-2.

Table 2-2. Criteria for identifying good and useful SPIs

Criterion	What it requires
Causal link with road crashes and injuries	Documented scientific evidence that the behaviour or condition has a proven association with crash occurrence or severity, with no or limited confounding effects. If an indicator value improves, there must be a corresponding decrease in the severity of crashes or injuries involving that road user group.
Quantifiability and sensitivity	The phenomenon can be captured numerically; the measurement is sensitive enough that real changes in the underlying behaviour are reflected as observable changes in the SPI value.
Relevance for policymaking	SPIs must provide actionable intelligence. The risk factor should be influenceable by policy interventions and enforcement. It should be possible to formulate SPI targets that are meaningful in the policy context.
Representativeness	The methodology must cover adequately the geographic area or period considered, typically through random sampling and weighting by exposure. SPI measurements must represent the real-world distribution of exposure. Failure to ensure representativeness leads to biased SPI values that cannot be compared over time or between countries.
Reliability	The measurement depends on accurate registration, sufficient sample sizes, reproducible methodology, and quality assurance during data collection and analysis. This requires Inter-Rater Reliability (IRR) testing and equipment calibration protocols. SPI methods should be designed so that the same methodology can be applied in future years and observed trends reflect real change rather than methodological change.
Comparability	Indicator values are comparable over time, with other geographical areas, and where relevant with other indicators. This requires standardized vehicle categories, standardized road types, and standardized weighting by traffic volume.
Feasibility	Operational feasibility requires that measurements are not blocked by legal obstacles; relevant authorities are willing to cooperate; adequate measurement technology is available, access to the necessary data is possible; the methodology can be repeated.

The relative importance of each criterion depends on the type of SPI. For database-based SPIs such as the share of safe road infrastructure or vehicle-fleet safety, feasibility and representativeness are often dominant

constraints. For behavioural SPIs representativeness and reliability are often the most demanding criteria, because they depend directly on the design of the sampling frame and on the consistency of observer behaviour.

It is also important to acknowledge that no indicator is ever perfect. By definition, an indicator is a reduction of reality and therefore conceals internal variation in the underlying data. A complex issue can never be fully captured by a single value, and someone who opposes the use of an indicator can always find an argument against it. The role of the methodologist is to ensure that the simplification is honest, transparent, and accompanied by clear documentation of its limitations.

2.5 Scope of SPIs

An SPI must be precisely scoped along four dimensions: the risk factor it addresses, the geographic area it covers, the road users it concerns, and the part of the road network to which it applies. Without a clear specification of all four dimensions, the indicator cannot be reliably defined and its values cannot be unambiguously interpreted or compared with other SPI values:

- The **risk factor**. This is the SPI topic, for example speeding, helmet use or post-crash care.
- The **geographic area** or jurisdiction. Within Trendline this is by default the whole country, but in a different context it may cover a region or even a municipality. Behaviour can vary considerably between regions, and the choice of geographic areas determines whether differences are masked or made visible.
- The **road-user group** and **vehicle type**. This dimension has historically been dominated by car drivers but is now increasingly extended to motorcyclists, cyclists, users of personal mobility devices, and pedestrians.
- The **road network**. Most countries have a national road categorisation based on road hierarchy, for example, motorways/expressways, primary roads, secondary roads, etc. and may further subdivide these by speed limit or other characteristics. The distinction between urban and rural roads, which is required for SPIs, is often not present in road categorisation by default. Nevertheless, it needs to be made clear to which part of the road network the SPI applies. For some SPIs (for example, speed compliance), further categorisation by speed limit is essential.

Each SPI chapter specifies these four scope dimensions.

A practical question that arises in many countries is whether the whole network is to be covered or only its busiest part: it is often the case that 10 to 15 % of the road length carries more than 75 % of the traffic, and there are economic and policy arguments for concentrating SPI measurement on the busier part of the network. This is promoted at UN level in the global voluntary road safety performance indicators (Van den Berghe et al., 2020). It is, however, not the case for Trendline SPIs, with the exception of the infrastructure SPI.

3 Terminology and definitions

3.1 Overview

This chapter brings together the key terms and concepts used throughout the manual and fixes a single agreed meaning for each. Most of the road, vehicle and temporal definitions are taken from official sources such as EU Directives, the CARE database and the harmonised European Road Safety Observatory (ERSO) classifications. Statistical and indicator-related terms follow established international good practice. The purpose is twofold: to ensure that every part of the manual uses the same words in the same way, and to provide a reference that data providers can consult so that figures delivered to the Trendline Secretariat rest on common definitions. Where a national classification cannot be mapped exactly onto a definition given here, the deviation is documented in the metadata.

The terms are grouped into thematic categories, each introduced by a short note that explains what the category covers and why it matters. The categories are ordered to follow the logic of an SPI study: the core indicator concepts come first, followed by the statistical and measurement vocabulary, the practical language of data collection and sampling, the classifications of the road network and the vehicle fleet, the temporal definitions, and finally the survey terms used for self-reported indicators. A short list of common abbreviations closes the chapter. Table 3-1 is an alphabetical index of all defined terms and the category under which each is explained.

Table 3-1. Alphabetical index of defined terms.

Term	Category
Annual average	Core SPI concepts
Annual average daily traffic (AADT)	Road network and vehicle terms
Baseline year	Temporal terms
Bicycle	Road network and vehicle terms
Breakdown	Core SPI concepts
Built-up area	Road network and vehicle terms
Bus / coach	Road network and vehicle terms
Child restraint system (CRS)	Road network and vehicle terms
Confidence interval (95%)	Statistical and measurement terms
Coverage	Statistical and measurement terms
Crash outcome indicator	Core SPI concepts
Darkness	Temporal terms
Daytime	Temporal terms
Design effect (Deff)	Statistical and measurement terms
Euro NCAP safety rating	Road network and vehicle terms
Exposure	Core SPI concepts
Expressway	Road network and vehicle terms
Free-flow traffic	Data collection and sampling terms
Headway	Data collection and sampling terms
Heavy goods vehicle (HGV)	Road network and vehicle terms

Inter-rater reliability (IRR)	Statistical and measurement terms
Interaction (at crossings)	Data collection and sampling terms
Key performance indicator (KPI)	Core SPI concepts
Korn–Graubard confidence interval	Statistical and measurement terms
Light goods vehicle (LGV)	Road network and vehicle terms
Measurement period	Temporal terms
Metadata	Core SPI concepts
Motorway	Road network and vehicle terms
Nighttime	Temporal terms
Observation session	Data collection and sampling terms
Observation site	Data collection and sampling terms
Passenger car	Road network and vehicle terms
Period prevalence	Survey terms
Personal mobility device (PMD)	Road network and vehicle terms
Point estimate	Statistical and measurement terms
Point prevalence	Core SPI concepts
Post-stratification weight	Statistical and measurement terms
Powered two-wheeler (PTW)	Road network and vehicle terms
Prevalence	Core SPI concepts
Recall period	Survey terms
Reference year	Temporal terms
Road segment	Road network and vehicle terms
Rural road	Road network and vehicle terms
Safety performance indicator (SPI)	Core SPI concepts
Sampling frame	Data collection and sampling terms
Sampling stratum	Statistical and measurement terms
Section speed	Data collection and sampling terms
Self-reported behaviour	Survey terms
Speed pedelec	Road network and vehicle terms
Spot speed	Data collection and sampling terms
Standard error	Statistical and measurement terms
Stratification	Data collection and sampling terms
Survey fieldwork period	Survey terms
Traffic count	Data collection and sampling terms
Trend period	Temporal terms
Trip-based prevalence	Survey terms
Trip-based sampling	Data collection and sampling terms
Urban area	Road network and vehicle terms
Urban road	Road network and vehicle terms
Vehicle kilometres travelled (VKT)	Road network and vehicle terms
Vehicle type	Road network and vehicle terms

Weekday daytime	Temporal terms
Weekday night-time	Temporal terms
Weekend daytime	Temporal terms
Weekend night-time	Temporal terms
Weighted estimate	Statistical and measurement terms

3.2 Core SPI concepts

This first group defines the indicators themselves and the vocabulary used to describe what they measure. A Safety Performance Indicator is an intermediate measure: it sits between the conditions and behaviours that influence safety and the crash outcomes that ultimately matter, and it is valued precisely because it can be measured more frequently and more sensitively than fatalities or serious injuries. The terms below distinguish an SPI from the outcome it predicts, set out the related notions of exposure and prevalence on which most indicators rest, and name the supporting concepts — metadata, breakdowns, trend periods — that recur in every later chapter.

Table 3-2. Core SPI concepts

Term	Definition	Source
Safety performance indicator (SPI)	An intermediate road safety indicator describing behaviour, system safety or protection levels.	ERSO / European Commission
Key performance indicator (KPI)	A measurable intermediate factor known to influence crash outcomes, used to monitor road safety performance; used interchangeably with KPI in EU road safety policy.	European Commission Road Safety Policy Framework
Crash outcome indicator	A result indicator measuring final safety outcomes such as fatalities or serious injuries, as opposed to intermediate performance indicators.	ERSO Road Safety Performance Framework
Exposure	The amount of travel or traffic presence (e.g. vehicles, trips, distance) forming the denominator for risk or performance measurement.	ERSO / transport safety theory
Prevalence	The proportion of a defined population engaging in a specified behaviour at a given time or during a defined period.	Epidemiological statistics terminology
Point prevalence	The proportion of road users exhibiting a behaviour at the moment of observation; the basis of the observed behavioural SPIs.	Epidemiological statistics terminology
Metadata	Structured documentation describing the data source, definitions, coverage, methodology and limitations associated with an SPI; an integral part of every reported value.	Trendline reporting framework
Breakdown	A subdivision of an SPI by category, such as road type, vehicle type, age group, week period or region.	Trendline reporting framework
Annual average	An indicator value calculated to represent an average condition over a full calendar year, even where it is based on samples collected during shorter periods.	Statistical reporting practice
Trend period	A sequence of consecutive reference years used to analyse changes in SPI values over time.	Monitoring and evaluation practice

3.3 Statistical and measurement terms

Most SPI values are estimates, not censuses. This group of terms defines the language used to express how good an estimate is and how it was obtained. The first terms (point estimate, standard error, confidence interval) are the standard vocabulary of statistical inference. The remainder concern the features of the Trendline sampling designs that often make the simple textbook formulae inadequate: observations are clustered within sessions and locations, so the variance of an estimate is inflated by a design effect and must be corrected for; and because indicator proportions can lie close to zero or one hundred per cent, a method is needed that remains valid at the extremes. The Korn–Graubard interval and the supporting weighting terms are gathered here so that the indicator-specific chapters can refer to them without restating the statistics. Inter-rater reliability, which governs whether two observers code the same scene the same way, is included because it is a measurement property rather than a feature of the field protocol.

Table 3-3. Statistical and measurement terms.

Term	Definition	Source
Point estimate	The single best numerical estimate of an SPI value derived from the sample data.	Statistical estimation theory
Standard error	A measure of the variability of an estimate that is due to sampling.	Statistical inference
Confidence interval (95%)	A statistical range around an estimate within which the true population value is expected to lie with 95 per cent probability.	Statistical inference
Sampling stratum	A subgroup of the population defined by characteristics such as road type or time period, within which observations are sampled.	Sampling theory
Post-stratification weight	A correction factor applied after data collection to align the sample distribution with a known population or traffic structure.	Survey sampling methodology
Weighted estimate	An SPI estimate adjusted using weights so that the result reflects the national traffic or population structure.	Survey statistics
Design effect (Deff)	The factor by which the variance of an estimate is inflated, relative to simple random sampling, because observations are clustered (here, vehicles within sessions and sessions within locations); it must be taken into account when computing confidence intervals for behavioural SPIs.	Survey sampling methodology
Korn–Graubard confidence interval	The recommended method for computing confidence intervals for SPI proportions under a complex sampling design, constructed from a Beta distribution with an effective sample size derived from the design-adjusted variance; it remains valid when the proportion is close to 0 or 100 per cent.	Korn & Graubard (1998)
Inter-rater reliability (IRR)	The degree of agreement between independent observers coding the same road users; checked by percentage agreement in the field (a minimum of 90 per cent on the primary variable is recommended before deployment) and by Cohen's κ in analysis (a value below 0.80 indicates that retraining or replacement is required).	Observational study methodology
Coverage	The proportion of the target system represented in an administrative dataset (e.g. the percentage of roads carrying a safety rating); the data-base-based counterpart of sampling completeness.	Trendline reporting framework

3.4 Data collection and sampling terms

This group covers the practical units and instruments of behavioural data collection: where an observation is made, over what period, and from what frame the sites are drawn. It also defines the conditions that an observation must satisfy to be valid, in particular free-flow traffic and the headway rule that operationalises it.

The encounter-based unit of observation used at unsignalised pedestrian crossings is defined here as well, because it is a feature of how the data are collected rather than of the indicator itself.

Table 3-4. Data collection and sampling terms.

Term	Definition	Source
Observation site	A physical road location selected for behavioural measurement.	Roadside survey methodology
Observation session	A defined time period at a specific location during which behavioural observations are conducted.	Observational study methodology
Sampling frame	The list or system from which observation sites or respondents are selected.	Sampling theory
Stratification	The division of a sampling frame into subgroups (strata) to ensure balanced representation across road types, time periods or other relevant characteristics.	Sampling theory
Traffic count	The number of relevant road users passing during a specified counting period, used for exposure weighting.	Traffic engineering statistics
Trip-based sampling	A survey method in which trips, rather than persons, are sampled to estimate behavioural prevalence.	Travel survey methodology
Free-flow traffic	Traffic conditions in which drivers can freely choose their speed and are not restricted by congestion, by infrastructure such as speed bumps, or by road works. For speed measurement, a vehicle is treated as free-flowing only if its headway to the vehicle in front is at least the distance covered in five seconds at the prevailing speed limit.	Trendline speeding guideline
Headway	The time or distance gap between a vehicle and the vehicle in front; used to decide whether a vehicle is travelling in free-flow conditions.	Traffic engineering
Spot speed	The instantaneous speed of a vehicle at a defined point. The speeding SPIs are based on spot speed.	Traffic engineering
Section speed	The average speed of a vehicle over a length of road. It is not used for the speeding SPIs, because averaging over a stretch of road cannot guarantee free-flow conditions for each individual vehicle.	Traffic engineering
Interaction (at crossings)	At unsignalised crossings, the unit of observation is not a single road user but the encounter between a pedestrian and a driver involving a mutual reaction (a change of speed); compliance is assigned to the interaction as a whole.	Trendline crossings guideline

3.5 Road network and vehicle terms

The classification of the road network is the most fundamental layer of the SPI methodology. Trendline adopts a functional classification whereby all roads must be stratified into three primary types: motorways, rural roads and urban roads. Countries may introduce secondary strata (for example distinguishing main from local urban roads, urban from other motorways, or motorways from expressways), but for European reporting these must be collapsed back into the three primary categories using the weighting procedures described in Chapter 8. Where a country historically uses categories that differ from those below, the national typology should be mapped onto these categories using characteristics such as the signs that demarcate the inside-versus-outside-built-up-area boundary, the speed regime and the number of lanes; any deviation is explained in the methodology section of the report.

Table 3-5. Road network types.

Term	Definition	Source
Motorway	A road specially designed and built for motor traffic, which does not serve adjacent properties; which is provided, except at special points, with separate carriageways for the two directions of traffic; which does not cross at level with any road, railway, tramway, bicycle path or footpath; and which is specifically designated as a motorway.	Directive (EU) 2019/1936
Expressway	A road specially built for motor traffic, accessible only from interchanges or controlled junctions, specifically signposted as an expressway and reserved for specific categories of motor vehicles, on which stopping and parking on the running carriageway are prohibited.	Directive (EU) 2019/1936
Rural road	A public road outside the urban boundary signs, excluding motorways and expressways.	ERSO road classification
Urban road	A public road located within the urban boundary signs.	ERSO road classification
Urban area (built-up area)	An area within officially designated urban boundary signs. 'Built-up area' is used as a synonym.	Trendline 30 km/h guideline

The second part of this group classifies the vehicle categories and the road user types that are used with the indicators. The passenger-car category is the minimum scope of most behavioural SPIs and is therefore defined once here and referred to in the indicator chapters. The micromobility and powered-two-wheeler categories are defined with care because the boundaries between them (between a personal mobility device and an e-bike, or between a speed pedelec and an ordinary bicycle) determine which subjects fall within which indicator.

Table 3-6. Vehicle and traffic terms

Term	Definition	Source
Vehicle type	The classification of vehicles into categories such as passenger car, light goods vehicle, heavy goods vehicle and bus/coach.	CARE / UNECE vehicle classes
Passenger car	A road motor vehicle, other than a moped or motorcycle, intended for the carriage of passengers and designed to seat no more than nine persons including the driver (UNECE category M1). The category includes microcars, taxis and other hired passenger cars with fewer than ten seats, and motor caravans; it excludes light goods vehicles, buses, coaches and mini-buses.	Eurostat / UNECE category M1
Light goods vehicle (LGV)	A goods vehicle with a maximum gross weight of less than 3.5 tonnes, used for the transport of goods (sometimes called a 'van').	CARE
Heavy goods vehicle (HGV)	A goods vehicle with a maximum gross weight exceeding 3.5 tonnes, used for the transport of goods (sometimes called a 'truck' or 'lorry').	CARE
Bus / coach	A passenger-carrying vehicle with more than eight seats in addition to the driver. A coach is distinguished by the presence of a luggage hold separate from the passenger cabin.	CARE / UNECE
Powered two-wheeler (PTW)	A motorised two- or three-wheeled vehicle, including motorcycles and mopeds.	UNECE vehicle classes
Bicycle / Pedal cycle	A road vehicle with at least two wheels, without engine. In some cases it can also use electric power (e-bike).	CARE

Speed pedelec	A power-assisted cycle formally classified as an L1e-B vehicle, and hence treated as a powered two-wheeler rather than a bicycle. Speed pedelecs are excluded from the cyclist subcategory of the helmet and lighting SPIs.	UNECE vehicle classes
Personal mobility device (PMD) or Motorised micro-mobility device (MMD)	A motorised, micro-mobility device such as an e-scooter, an e-step, a segway, a monowheel or a self-balancing unicycle. The device should have at least one wheel, be designed for one person, and have an electric motor that can achieve a maximum speed of up to 25 km/h.	CARE
Child restraint system (CRS)	A protective device for the carriage of children, used in place of, or together with, the adult safety belt according to the child's age, height and/or weight as set by national law. Its correct use can normally be assessed only by in-vehicle inspection.	Trendline restraint guideline
Vehicle kilometres travelled (VKT)	The total distance travelled by vehicles over a given road section or network.	Eurostat transport statistics
Annual average daily traffic (AADT)	The total number of motor vehicles passing a point on a road over a year, divided by 365.	Traffic engineering statistics
Euro NCAP safety rating	A star-based vehicle safety performance rating issued by the European New Car Assessment Programme.	Euro NCAP assessment protocol

3.6 Temporal terms

Behaviour and traffic composition vary strongly with the time of day and the part of the week, so most SPIs are reported separately for daytime and night-time and for weekdays and weekends. This group fixes those temporal boundaries using the harmonised ERSO classification. Two points deserve emphasis because they sit on the boundaries that matter most for the driving-under-the-influence indicator: Friday night and Sunday night both count as weekend night-time, not weekday night-time. The terms 'weeknight' and 'weekday night-time' are synonyms; the manual uses 'weekday night-time' throughout. This group also defines the reference points used for trend reporting — the baseline year and the reference year — and the lighting-related notion of darkness used by the cyclist-lighting SPI.

Table 3-7. Temporal terms.

Term	Definition	Source
Daytime	The period from 06:00 to 21:59.	ERSO temporal classification
Nighttime	The period from 22:00 to 05:59.	ERSO temporal classification
Weekday daytime	Monday to Friday, 06:00–21:59.	ERSO temporal classification
Weekday night-time	Monday to Thursday, 22:00–05:59. (Synonym: 'weeknight'.)	ERSO temporal classification
Weekend daytime	Saturday and Sunday, 06:00–21:59.	ERSO temporal classification
Weekend night-time	Friday to Sunday nights, 22:00–05:59. Friday night and Sunday night both count as weekend night-time.	ERSO temporal classification
Darkness	For the cyclist-lighting SPI, a natural light density below 26 lux, corresponding in most countries to roughly half an hour before sunrise and half an hour after sunset; ideally verified with a lux meter.	Trendline lighting guideline
Measurement period	The actual time span during which data were collected.	Statistical reporting standards

Reference year	The calendar year to which an SPI value refers, regardless of when the data were processed or reported.	Official statistics standards
Baseline year	The reference year against which future SPI values are compared.	Monitoring and evaluation practice

3.7 Survey terms

Some indicators, most notably the alternative measure of driving under the influence, are derived not from roadside observation but from self-reported survey data (in Europe mostly the ESRA survey series). Self-reported measures use a different notion of prevalence from the point prevalence of observed behaviour: respondents report whether they have engaged in a behaviour over a recall period, so the resulting figure is a period prevalence. This group defines that vocabulary and the related survey-operational terms.

Table 3-8. Survey terms

Term	Definition	Source
Self-reported behaviour	Behaviour measured through survey responses rather than direct observation or administrative data.	Survey research methodology
Recall period	The time frame survey respondents are asked to consider when reporting behaviour (e.g. the last 30 days).	Survey methodology standards
Period prevalence	The proportion of respondents reporting at least one occurrence of a behaviour within a recall period.	Epidemiological statistics
Trip-based prevalence	Behaviour prevalence measured per trip rather than per person or per observation moment.	Travel behaviour survey methodology
Survey fieldwork period	The time span during which a questionnaire survey is conducted.	Survey operations standards

3.8 Common abbreviations

The following abbreviations are used throughout the manual.

Table 3-9. Common abbreviations.

Abbreviation	Meaning
AADT	Annual average daily traffic
CARE	The EU road accident database (Community database on Accidents on the Roads in Europe)
CRS	Child restraint system
Deff	Design effect
ESRA	E-Survey of Road users' Attitudes
EU	European Union
HGV	Heavy goods vehicle
IRR	Inter-rater reliability
KPI	Key performance indicator
LGV	Light goods vehicle

MS	Member State
PMD	Personal mobility device
PTW	Powered two-wheeler
RBT	Random breath testing of drivers' blood alcohol concentration, conducted without prior suspicion
SPI	Safety performance indicator
VKT	Vehicle-kilometres travelled
VRU	Vulnerable road user

4 Using SPIs

4.1 Reporting SPI results

4.1.1 Purpose of reporting

The purpose of reporting SPI values is to provide clear, comparable and policy-relevant information about the safety performance of the road transport system. Whatever the measurement approach (roadside observation, camera measurement, database analysis or survey), an indicator is useful only once it is reported in a form that is clear, comparable and honest about its uncertainty. Reporting should make four things possible for every indicator:

- the monitoring of trends over time
- comparisons between and within countries
- the identification of the areas that most need intervention
- the evaluation of the effect of policy measures already taken.

An SPI should always be presented as an intermediate safety outcome and never as a final outcome such as the number of fatalities. This distinction governs how a result may legitimately be interpreted, and it is the single point most often lost when results reach a non-technical audience. Also, when different methods for a single indicator (e.g., driving under the influence of alcohol) produce different values, those values measure different things and should never be combined into one series or compared as if equivalent.

This chapter is concerned with principles. The operational specifics of delivering data to Trendline, such as the levels of data delivery, the standardised submission template, the full metadata record, and the consolidated list of minimum requirements, concern the execution of a study rather than the principles of reporting; these are set out in the practical guidance of Part D. Each indicator chapter in Parts B and C also closes with its own minimum reporting requirements. What follows is the common ground beneath all of them.

4.1.2 Core reporting elements

Every SPI report, whatever its length, should contain a minimum set of elements. These elements are what a reader needs in order to understand what was measured, how far to trust it, and whether it may be compared with another figure. These are listed in Table 4-1 (next page).

Metadata are what make these elements comparable: without them, two values that look alike may rest on different definitions and mislead. Every reported value therefore carries a metadata record, and transparency is placed above the appearance of comparability — where a national classification cannot be mapped exactly onto the Trendline categories, the deviation is documented rather than concealed. The detailed content of that record, and the levels at which data are delivered to the Trendline coordination team, are set out in Part D.

4.1.3 Uncertainty, comparability and trends

Different types of indicator differ in how their uncertainty arises, in what limits their comparability, and in how their trends must be read. The words that are often used, such as “confidence”, “coverage” and “comparable”, carry different meaning for each. Every SPI report should state the main source of uncertainty for each indicator it presents, because the appropriate caveat differs (see Table 4-2).

Table 4-1. Core elements of an SPI report.

Element	What it states
SPI definition	The exact operational definition: the behaviour or condition, the population, and the inclusion rules
Reference year / period	The year or period the value refers to, and the data-collection or extraction period
National estimate	The headline SPI value (weighted, for sampled indicators)
Measure of confidence	A 95% confidence interval for sampled indicators; a coverage level for database indicators
Sample size	The number of observations or respondents behind each estimate
Breakdowns	Values by road type, user group, region, time period and other strata, where the data allow
Metadata summary	Data source, method, definitions used and limitations

Table 4-2. Main source of uncertainty by type of indicator.

SPI type	What the indicator reflects	Main source of uncertainty
Observed behavioural	Real-world behaviour at a point in time	Sampling variability (clustered design)
Self-reported	Norms, perceptions and stated behaviour	Sampling error together with self-reporting bias
Database / system	Structural or system safety conditions	Coverage bias

For sampled indicators, the confidence interval should be computed by a method that respects the sampling design. Confidence intervals that assume simple random sampling are too narrow and are not acceptable for Trendline reporting (see Section 8.4). A wide confidence interval should be reported as such, because it implies that trends will be more difficult to observe. For example, a long-running national alcohol-observation series may test eight to ten thousand drivers per round but still leave most year-on-year differences inside the margin of error, so that no trend can be claimed from the values alone. For database indicators the equivalent caveat concerns coverage and classification: an indicator drawn from a source that covers only part of the network, or that classifies records inconsistently, must say so, because the gap is the analogue of a wide interval. Breakdowns are reported only where the subgroup data are sufficient and consistent; reduced precision or partial coverage should be flagged explicitly.

Comparison and trend reporting follow from this. Cross-country comparison requires equivalent definitions, comparable coverage levels and similar methodologies. Differences between countries may reflect methodological or contextual factors rather than genuine differences in safety; an SPI report should say so where there is doubt. A trend can be read only across rounds measured consistently, so trend reporting requires a stable methodology and stable definitions. Any change in the methodology (e.g., a new weighting basis, a revised classification, a different device or a new data source) should be documented and its likely effect on comparability explained. Charts with SPI values should show a break in the series rather than smoothed over. Movements that fall within the confidence intervals, or within the bounds of a coverage change, are described as such and not presented as change.

4.1.4 Communicating SPI results

Technically correct data that are poorly communicated may not influence policy. A robust communication approach requires a translation of technical content into language a non-specialist can absorb. This covers explaining the confidence measure in plain terms at least once in every document, explaining weighting in a simple way ("rural roads carry less traffic, so observations there have a smaller influence on the national average"), restating the difference between an SPI and a crash statistic in every presentation to a policy audience; and presenting changes in absolute terms wherever possible ("up from 72 % to 78 %" rather than "an eight per cent improvement").

Different audiences need different products. Table 4-3 sets out the principal audiences and the format and message each requires.

Table 4-3. Formats and messages for different audiences.

Audience	Message focus	Format
Ministry / senior officials	Are we on track to meet the targets, and what is the priority for action?	Two-page brief plus infographic
Road safety department	Full SPI values, methodology, confidence or coverage, trend analysis	Full technical report
Enforcement agencies	Which behaviours, roads, times and groups are most non-compliant	Operational summary with maps and charts
Media and public	One headline finding per SPI; the country against its neighbours	Press release plus simple infographic
International partners / donors	Methodology, comparability, progress against targets	Technical report plus executive summary

4.1.5 Structure of a national SPI report

A national report follows a recognisable structure so that both technical and policy readers can find what they need. An executive summary of no more than two pages states the purpose, the indicators reported, the headline values with their confidence intervals or coverage levels, the comparison with previous values or with targets, and the two or three most policy-relevant conclusions, in plain language and with a single summary table or infographic. A methodology section records, for each indicator, the coverage, scope and period. For sampled indicators it describes the sampling design and the sizes achieved against the targets, and for database indicators the source, the extraction and classification rules, and how the coverage level was determined.

A results section reports every value with its measure of confidence, overall and by stratum, set against the target and the previous rounds. The statistical significance of any difference is stated explicitly rather than left to the reader to infer from overlapping intervals. An interpretation and policy-conclusions section states whether performance is improving, stable or deteriorating, places the result in context (legislation, enforcement intensity, traffic volumes, data systems), and makes specific, actionable recommendations. A closing limitations section gives a transparent account of sample shortfalls, partial coverage, classification ambiguities and protocol departures, with the implication of each for interpretation. Reporting an unfavourable limitation is a condition of credibility, and, where police or other authorities were involved in the data collection, a matter of the reporting principle agreed with them.

Five mistakes recur often and should be avoided:

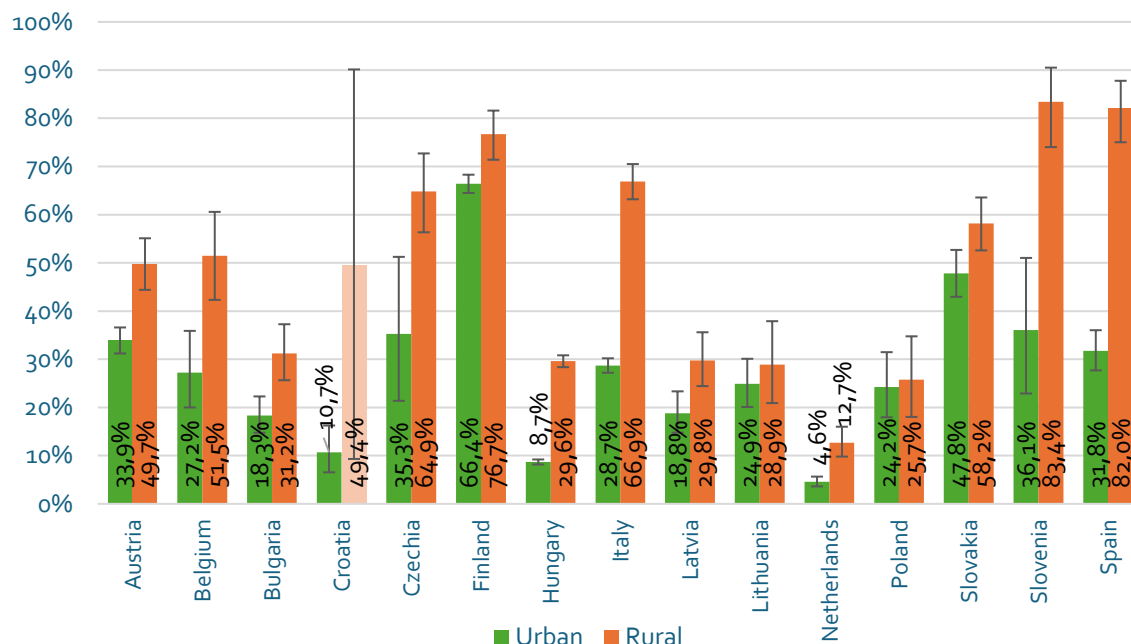
- Reporting point estimates without their confidence intervals, which makes trends look more certain than they are.
- Presenting results without comparison to targets or previous rounds, leaving numbers without context.
- Burying limitations in an appendix, when they should be prominent enough for policymakers to calibrate their confidence in the findings.
- Using the same technical report for every audience, when a ministerial brief and a full technical report serve different purposes.
- Presenting only the results. A good SPI report always ends with a clear answer to the question "what should we do next?".

4.1.6 Visualisation and oral presentation

Visualisation should support the message. Bar charts with error bars showing the 95% confidence interval are the standard for cross-sectional comparisons of sampled indicators, and line graphs with confidence bands for trends; every figure carries a caption stating what is shown, what the error bars or coverage caveats represent, and what the key message is. Scales are not truncated in ways that exaggerate differences, colour coding is consistent, and figures remain legible in black and white. Observed, database and self-reported indicators are visually distinguished so that the reader never mistakes one type for another, and values from different measurement methods are never combined in a single series. Where dashboards are used, they may show several indicators together provided they preserve each indicator's type and uncertainty.

Figure 4-1 provides an example graph from the Trendline final report (Van den Berghe and Stelling, 2025). The bars introduce 95% confidence intervals, and the colours are different for urban and rural roads. Lighter colour indicate a measurement with fewer observations, thus having a wide confidence interval.

Figure 4-1. Percentage of cyclists wearing a helmet by road type (Trendline programme).



For oral presentations good practice and rules of thumb include: an opening statement that gives the single most important finding; one key finding per slide; a "so what?" slide answers what the result means for policy and what should be done next. The presenter should be ready to explain confidence and significance in plain language if challenged.

4.2 Using SPIs in the policy cycle

4.2.1 SPIs in the policy cycle

An effective SPI system depends on coordination between the bodies that hold the data and act on the results: statistical agencies, road authorities, the police and other enforcement bodies, and the health and emergency-medical services. Clear and durable institutional responsibilities for data collection, analysis, policy integration and dissemination improve both the quality of the data and the sustainability of the system, and their absence is one of the institutional risks noted above.

The value of an SPI is realised only when it informs decisions. SPIs are intermediate performance measures that bridge the gap between the adoption of a policy and its eventual effect on deaths and injuries: without them, policymakers would have to wait years for crash data to confirm whether an intervention had worked, by which time the chance to correct course might have passed. An SPI can support several stages of the road safety policy cycle, as set out in Table 4-4.

Table 4-4. The role of SPIs across the policy cycle.

Policy stage	Role of the SPI
Problem identification	Detect high-risk behaviours or gaps in the system
Target setting	Define measurable performance goals
Monitoring	Track progress between strategy periods
Evaluation	Assess the effectiveness of interventions

An SPI system is not a one-off reporting exercise but a permanent part of road safety governance. Over time a country should aim to widen the coverage of its indicators, harmonise its definitions with the common framework, increase the automation of data collection where this improves quality and lowers cost, and embed the indicators ever more firmly in its national road safety strategy. When the set of indicators grows, it may also be useful to consider leaving out indicators where the targets have been largely achieved or which are unlikely to evolve further, to make room for some new indicators that have become more relevant from a policy perspective. In this way the indicators are not merely collected and reported but actually used, round after round, to steer policy toward the long-term goal of moving close to zero deaths and serious injuries. Interpreting and using self-reported indicators

Self-reported indicators tend to underestimate the prevalence of socially disapproved behaviours and to overestimate somewhat that of approved ones. They are nevertheless valuable for relative comparison across countries and groups, for trend monitoring, and for the demographic and attitudinal breakdowns that roadside data cannot provide. Sometimes they are the only representative national source for behaviours that are hard to observe, such as occasional drink-driving, drug driving, drowsy driving, and phone use not visible from outside the vehicle.

Within policy they serve three uses in particular: identifying high-risk groups, informing the design of communication campaigns by revealing where attitudes are out of step with behaviour, and tracking cultural change following a new rule, a campaign or a major enforcement operation. They are interpreted alongside observed indicators rather than as competitors, and the distinction between a self-reported and an observed indicator is always made explicit so that the audience knows what is being measured.

4.2.2 Setting targets

Every SPI value should be reported against a target. Setting an appropriate target for an SPI is a balancing act between ambition and credibility. Several starting points may be available:

- the average or recent past performance of the country itself
- benchmarking with comparable countries or regions
- the best performance observed elsewhere in the world
- feasibility considerations
- ethical considerations.

In practice, most countries combine several of these references. Targets should be linked explicitly to the interventions that are planned to reach them, and they should have a time horizon of at least two to three years so that the impact of those interventions can become visible. A horizon that is too short tends to lead to less ambitious targets, while a horizon that is too long allows for slow starts and loss of momentum. Annual monitoring is recommended in either case. It is also good practice to make explicit what type of target has been chosen.

Conceptually, for most SPIs three types of targets can be considered:

- Idealistic targets (*for example, zero speeding*). Such targets are ethically attractive and may even be a legal obligation, but they can lack credibility if they are perceived as unreachable. A target that is not credible cannot motivate the actors who must work towards it.
- Minimalistic targets (*for example, the current level of performance*). A target that is easy to reach can give a false impression that the system is safe when in fact serious risks remain. In practice, such targets are only useful if the current performance is acceptable, but there is a critical threshold value that must never be crossed.
- Realistic targets (*for example, a seat belt rate of drivers of 98 %*). The most effective targets are ambitious yet credible: they should sit between the minimalistic and the idealistic value, and be supported by all key stakeholders.

An SPI report should state explicitly whether the country is on track, behind schedule or ahead of schedule against its national strategy. For each SPI that shows poor performance or no improvement in relation to the target, the report should identify at least one intervention shown to be effective elsewhere.

4.2.3 An example

Consider a country in which a roadside SPI study establishes that 2 % of car drivers are driving under the influence of alcohol. Policymakers judge this unacceptable and set a target of 0.5 % over a ten-year horizon. The national road safety plan adopted in pursuit of the target combines several measures: the legal blood-alcohol limit is lowered, nationwide awareness campaigns are launched, police enforcement is intensified, penalties are increased, and nighttime access to alcohol is restricted in licensing law.

Every two years the country re-runs the roadside study and checks the share of drivers over the limit against the target trajectory. The results should show progress: the SPI value is roughly 1.7 % at year two, 1.4 % at year four, and so on toward 0.5 % at year ten. If the value tracks the trajectory, the measures continue; but if progress is too slow, more should be done such as more random breath testing, reinforcement of the campaign, tightening of alcohol licensing restrictions. The role of the SPI here is not merely to measure but to trigger: it makes the gap between intent and reality visible within the policy cycle, so that policy can adjust before the ten-year horizon expires and the impact indicators reveal that the target was missed.

The same logic applies at every level of governance: a city tracking seat belt use in its fleet, the EU comparing indicators across member states, the World Health Organization applying target-setting to a worldwide framework.

4.2.4 Follow-up interventions

Because SPIs measure the behaviours and conditions that interventions are meant to change, they are most useful where they are tied to specific measures. Table 4-5 gives examples of possible interventions.

Table 4-5. Examples of links between SPI findings and interventions.

SPI finding	Possible intervention
High drink-driving prevalence	Increased alcohol enforcement; awareness campaigns; lower BAC limit
Low helmet use	Helmet campaigns or legislation
High handheld mobile-phone use	Enforcement and awareness campaigns
Low share of safe roads	Infrastructure investment

The two interventions most closely bound up with SPIs are awareness campaigns and enforcement. A road safety campaign aimed at a particular behaviour finds its natural performance measure in the corresponding SPI: a seat belt campaign is tracked through the restraint indicator, an anti-distraction campaign through the distraction indicator, an anti-drink-driving campaign through the alcohol indicator. The relationship runs in both directions. SPIs help to evaluate campaigns, by comparing the indicator before and after, but they also help to design them: breakdowns can identify the groups and contexts where the problem is concentrated. A national seat belt rate of 94% may conceal 78% among young male drivers on rural roads at night, which points directly at the audience a campaign should address. Attitudinal and self-reported indicators can reveal whether the problem is one of attitude, knowledge or habit, which determines the message.

Three cautions apply when an SPI is used to evaluate a campaign:

- A campaign's effect on behaviour is often small relative to the indicator's confidence interval, so a before-and-after comparison needs an adequate sample size and a design able to detect a small change.
- Second, campaigns rarely run in isolation: they often accompany enforcement or a legal change. The evaluation must therefore allow for confounding and for time lags, and should not attribute to the campaign a change that several measures produced together.
- An SPI measures the behaviour, not the campaign. Where an SPI shows that a campaign has not moved behaviour, focus groups and other qualitative work are the appropriate way to understand why.

Enforcement stands in two distinct relationships to SPIs. First, most enforcement-relevant behaviours have a matching compliance indicator: speed, restraint use, helmet use, driving under the influence and handheld phone use each correspond to one of the behavioural SPIs. Second, enforcement activity is itself the subject of a dedicated indicator: the enforcement indicator cluster discussed in Chapter 22. These SPIs measure the effort, effectiveness or level of enforcement, an input or process measure rather than a behaviour. Reading the two together is informative: rising enforcement effort accompanied by rising compliance suggests the programme is working, whereas rising effort without rising compliance suggests that it is not, or not yet.

SPIs can also be used to support enforcement, but only in a specific and bounded sense. SPI results show where and when compliance is lowest such as the types of roads, time periods and road-user groups. This information can inform the aggregate targeting and deployment of enforcement resources. This population-level use is distinct from any individual-level use since SPI data are collected for research and monitoring and are never used to initiate or support enforcement action against an identified individual. The point is sharpest for driving under the influence, where the police are operationally part of the measurement: the SPI requires the random selection of drivers rather than the selective stopping of normal police work; the anonymised research record

never merges with the police enforcement record. SPIs thus inform enforcement strategy at the level of the system.

4.2.5 Risks of using SPIs and how to manage them

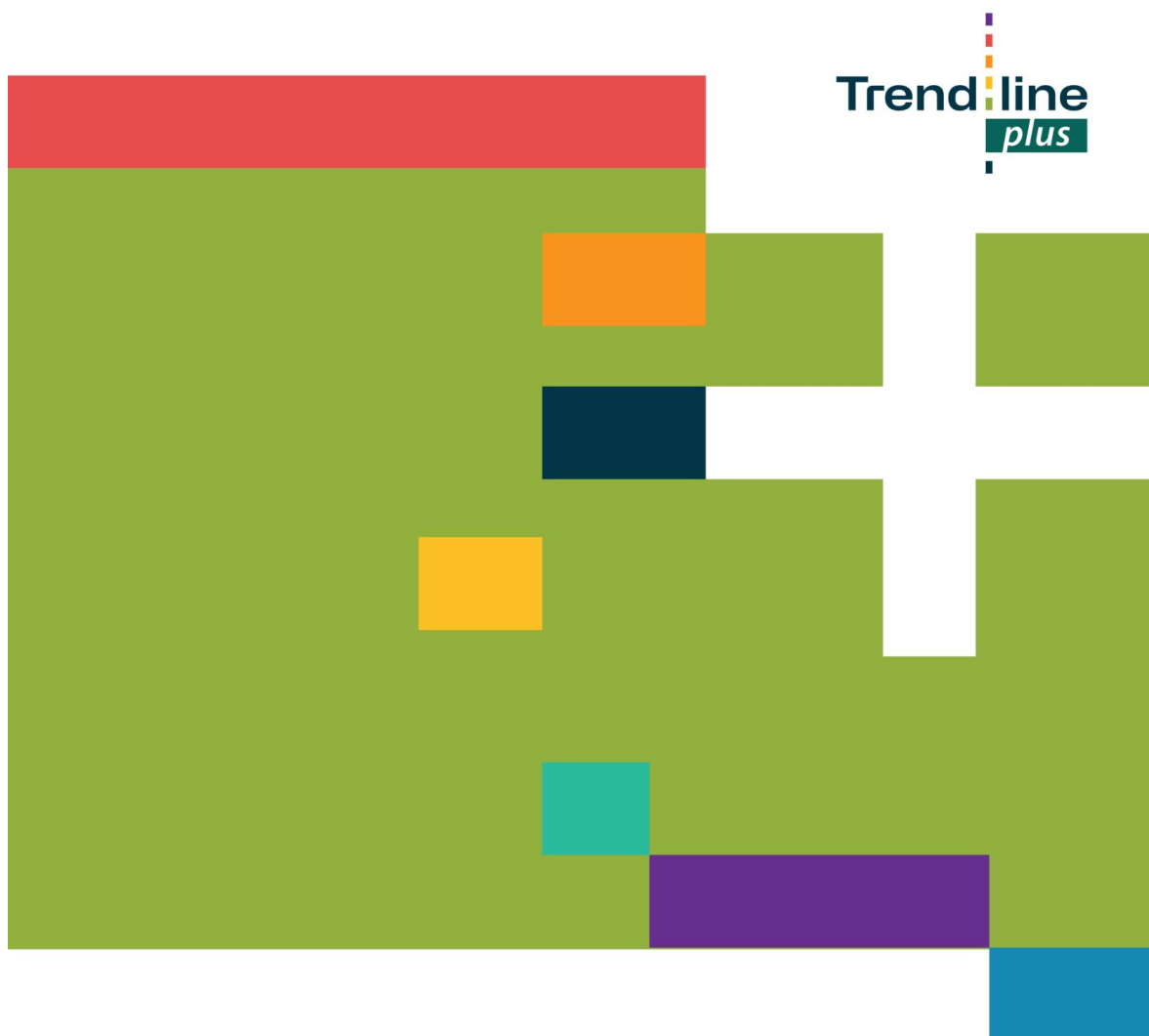
The value of an SPI system is not unconditional, and one should be aware of the risks at the outset such as:

- Data collection and analysis can be costly.
- Cost can drive the choice of indicators, so that what is cheap gets measured rather than what is useful.
- An indicator can be biased because the underlying dataset is not representative.
- The SPI methodology may not support useful subgroup breakdowns.
- An indicator's sensitivity may be too low to detect a real trend in time.
- An indicator may confirm a known problem without pointing to the appropriate countermeasure.
- Indicators may not be integrated into the national strategy.
- The frequency of measurement may not be appropriate (too frequent or not frequent enough).
- Areas with no indicator may be overlooked.
- Authorities may be reluctant to adopt indicators because the indicators would hold them accountable.

None of these risks invalidates the approach; each has a known mitigation (Table 4-6), and acknowledging them at the outset lets them be planned for rather than emerging later as unwelcome surprises.

Table 4-6. Risks of using SPIs and recommended mitigations.

Category	Risk	Recommended mitigation
Data	High cost; sampling bias; poor subgroup breakdowns; cost-driven choice of indicators	Realistic budgeting; weighting; minimum subgroup sample sizes; relevance-first selection
Methodology	Low sensitivity; weak link to countermeasures	Adequate sample sizes; complementary diagnostic studies (focus groups, deeper analyses)
Institutional	Poor integration with policy; wrong measurement frequency; gaps in coverage; political reluctance	Embed in the national plan; align with the policy cycle; commission missing indicators explicitly; build political support up front



PART B – BEHAVIOURAL SPIs

5 Introduction to behavioural SPIs

5.1 Why direct observation or roadside measurement is needed

Many safety-relevant behaviours cannot be reliably measured using administrative data. Alcohol use while driving is rarely captured in routine traffic databases; distraction by mobile phone is not systematically recorded; and helmet use is often not registered in crash databases for non-injured riders. Self-reported surveys may underestimate risky behaviours due to social desirability bias, meaning that respondents tend to report behaviour they believe is more socially acceptable rather than their actual behaviour. Therefore, direct roadside observation or measurement is required to obtain valid prevalence estimates that reflect real-world conditions.

Behaviour-based Safety Performance Indicators (SPIs) measure the behaviour of road users directly in traffic. These indicators describe the prevalence of risk-related or protective behaviours among road users exposed to real traffic conditions. Examples include driving under the influence of alcohol, use of handheld mobile phones while driving, helmet wearing among cyclists and powered two-wheeler riders, and compliance with traffic rules at pedestrian crossings or intersections. Although their operational details differ, these SPIs share a common methodological structure that ensures the resulting indicators are comparable over time and across countries

Depending on the SPI, data collection may involve either visual roadside observations by human observers, instrument-based roadside testing or observations, or self-reported behaviour (see Table 5-1). Although the measurement approaches differ, the sampling logic and statistical principles remain the same.

Table 5-1. Types of measurements for behavioural SPI

Type	Examples	Key feature
Visual roadside observation by human observers	Helmet use, distraction by mobile phone, seat belt use, compliance at pedestrian crossings	Observer classifies visible behaviour
Instrument-based testing or observation	Breath tests (alcohol), radars (speeding), induction loops (speeding)	Objective measurement device
Self-reported behaviour	Fatigue, drug use, alcohol use, inappropriate speed	Road users report past risky behaviour

5.2 Positioning SPI-oriented studies within behavioural research

Behavioural studies in road safety pursue the goal of observing, explaining and predicting road-user actions. They investigate how road users make decisions, consciously or unconsciously, that affect safety outcomes. Some studies focus on the psychological factors that lead to risk-taking; others are designed to evaluate the impact of specific laws, infrastructure changes or safety campaigns.

Within this broader landscape, four main types of behavioural study can be distinguished, each with a different purpose and a different methodological character:

1. **Explanatory and theory-driven studies** aim to build and test theoretical models of human behaviour, exploring constructs such as attitudes, social norms, risk perception and habit.
2. **Intervention evaluation studies** assess the effectiveness of a specific intervention (e.g., a new traffic law, an educational campaign, or an infrastructure change) by examining the causal link between the intervention and a change in road-user behaviour.

3. **Naturalistic driving studies (NDS)** equip vehicles with cameras, sensors and GPS to observe every-day driving behaviour over extended periods, including near-misses and critical incidents, without researcher interference.
4. **Monitoring studies for Safety Performance Indicators** have a non-experimental purpose: they aim to provide reliable and comparable measurements of specific, predefined behavioural metrics that serve as proxies for safety. The focus is on accurately measuring how often a given behaviour occurs in a defined population at a given time.

It is the last of these four types, SPI-oriented monitoring studies, that constitutes the principal subject of the second part of this manual. The methodologies described in the following sections are designed primarily to support this type of study. Several of the principles also apply to intervention evaluation studies, because evaluation studies typically use SPI-style measurements as their before-and-after metrics.

SPI-oriented studies have a different character from other types of behavioural research. Their primary function is monitoring: they answer the question of how safe the current behaviour of a given road-user group is, expressed in measurable terms. They allow trends over time to be tracked and the effectiveness of road-safety measures to be assessed at the population level. A second function is to build evidence for policy by identifying how large a problem is and which road-user groups require priority attention.

In line with this monitoring function, SPI studies have several specific characteristics:

1. They aim to make accurate estimates of the average behaviour of a large population, typically a country as a whole.
2. Measurement must be consistent: definitions, site-selection rules and observation protocols are followed rigorously to ensure standardisation and comparability over time and space, even when this constrains operational flexibility.
3. Each SPI study should produce a single, robust figure which can be used for benchmarking against previous years, against national targets or against values of other countries.

In countries that have already established a tradition with SPIs, the standard set typically covers several behavioural SPIs such as speeding, seat belt use, driving under the influence of alcohol, helmet wearing by motorcyclists, and distraction by mobile phone. Additional indicators added to national SPI frameworks may include indicators on topics such as red-light running, use of child restraint systems, helmet use among cyclists, use of illicit drugs or fatigue – some of these may be based on self-reported data.

SPIs can be calculated at different levels of aggregation. The national level is the core reporting level for tracking a country's overall progress and for international comparisons. The regional level is optional within Trendline but valuable for identifying geographic disparities and for allocating resources to areas with the poorest performance. Breakdowns by road type, vehicle type, time of day, part of the week, month, age group, gender (where appropriate), and vehicle occupant type (where appropriate) are essential for diagnostic analysis. However, all breakdowns must be based on methodologically consistent sampling or data coverage. A breakdown that mixes observations from different sampling frames, or that applies different definitions in different regions, will produce misleading comparisons and may conceal more than it reveals.

5.3 General characteristics of behavioural SPI studies

All behavioural SPI studies follow the same core logic:

1. Define the target population. The population are typically particular road users (e.g., car drivers) exposed to traffic during a given time period and over a defined road network.
2. Select a representative sample of observation moments and locations. Because it is impossible to observe all traffic, a structured sampling design is used.
3. Observe or measure individual road users. Each observed road user provides one data point indicating compliance or non-compliance.

4. Weight the observations. Statistical weights are used so that the final SPI estimate reflects the distribution of traffic exposure.
5. Estimate national SPI values. The weighted data are used to calculate the percentage of traffic complying with the rule or exhibiting the protective behaviour.

Across all behavioural SPIs, the following methodological elements are shared:

- Definition of the **sampling frame** (road network and traffic exposure)
- **Stratification** by road type (urban, rural, motorway)
- Consideration of **time periods** (day/night, weekday/weekend)
- Minimum **sample size** requirements
- Procedures for **traffic counting** to support weighting
- Standardised **data coding**
- Statistical **weighting** and estimation of SPI value.

These elements are described in detail in the following chapters; additional specific considerations per SPI are presented in the SPI-specific chapters.

Every behavioural SPI in this Part B rests on the same backbone: the sampling logic of Chapter 6 (Sampling strategy for behavioural SPIs), the fieldwork organisation described in Chapter 7 (Fieldwork: data collection and data recording), and the two-layer weighting and confidence-interval framework of Chapter 8 (Data analysis for behavioural SPIs). The SPI-specific chapters that follow do not restate this backbone; each opens by noting only the respects in which its indicator genuinely diverges from it (e.g., a different measurement instrument, a different unit of analysis, a partly hidden behaviour, a sparse target group, and so on) and refers back here for everything the indicators share.

A behaviour may be measurable by more than one method, such as direct observation, instrument-based measurement, or self-report. The same indicator measured by different methods yields values that are not interchangeable. Observation gives point prevalence; period-prevalence surveys give a recall-window figure; trip-based surveys give a retrospective point estimate. Each carries different biases, so values from different methods should be reported separately and never combined into a single series.

The behavioural indicators discussed in this manual are at different stages of maturity. Most of the original eight core indicators that were used in Baseline are established, with a settled methodology and a track record of national delivery. However, the infrastructure indicators and several of the experimental indicators that were developed and tested in the first phase of Trendline are still being developed and tested on a limited scale. Their methodology is a working basis rather than a fixed standard, and a country delivering them is contributing to that development.

Traditionally, most behavioural SPIs focus on the behaviour of vehicle drivers. The Compliance and Helmet SPIs broaden this picture in two important ways. First, they extend the observation logic to vulnerable road users (VRUs) whose behaviour cannot be captured adequately by speed radars or by camera-based vehicle classification. Second, they introduce the notion of an interaction as a unit of observation: at unsignalised pedestrian crossings, what is observed is not a single road user but the encounter between a pedestrian and a driver, with compliance assigned to the interaction as a whole. These two features shape the observer training requirements, the location selection logic and the sampling design, which will all be different from approaches used for observing passenger cars and goods vehicles.

Pedestrian and cyclist observations also impose specific operational constraints that other behavioural SPIs do not face. Pedestrian volumes and cyclist volumes are highly seasonal and concentrated at particular times of day, with school zones, commuting peaks and weekend leisure flows producing very different observation conditions. They are also highly weather-sensitive; an observation campaign that ignores adverse-weather

days will systematically over-represent fair-weather behaviour. Observations of VRUs should take place when activity is highest while avoiding adverse weather and conditions of insufficient visibility.

SPIs are often expressed as percentages, but the underlying concept of exposure differs across indicator types. For behavioural SPIs concerning drivers, the implicit exposure is traffic volume or kilometres driven. For helmet SPIs concerning cyclists and powered-two-wheeler riders, the implicit exposure is the number of riders exposed to traffic. For infrastructure SPIs, exposure is the distance driven on the road network. For vehicle safety SPIs, it is the number of vehicles in the fleet. For post-crash care SPIs, it is the number of injury crashes requiring emergency response. Understanding the exposure concept is crucial for correct sampling, proper weighting, and meaningful comparisons between countries or over time. An SPI that is not weighted by exposure may give a misleading picture: a road type that carries 2% of traffic but accounts for 10% of observations will be over-represented unless the weighting reflects its true share of vehicle-kilometres. Please note also that in general the exposure measure for behavioural SPIs in general differ from those used for database-based SPIs (see Table 5-2 – the database rows are developed in part C of the manual).

Table 5-2. Examples of exposure concepts by SPI type

SPI type	Implicit exposure
Behavioural (drivers)	Traffic volume or kilometres driven
Helmet (cyclists/PTWs)	Number of riders exposed to traffic
Infrastructure	Distance driven on road network
Vehicle safety	Number of vehicles in fleet
Post-crash care	Number of injury crashes requiring emergency response

5.4 Advantages and limitations of using human observers

Behavioural studies that rely on human observers have a number of well-documented advantages:

- They are often the easiest type of behavioural study to implement
- The same general approach can be reused for very different types of risky behaviour (helmet use, seat belt use, mobile-phone distraction)
- A substantial body of documented expertise is available
- The equipment requirements are often modest, so costs are relatively low and no highly specialised staff are needed
- The method can be implemented with high reliability and comparability across years and countries
- There is a degree of in-built flexibility, in the sense that observation locations can be replaced on the spot if needed.

At the same time, the method has well-known disadvantages:

- The selection of observation sites is decisive for the validity of the results
- The subjectivity in the observer's judgement can lead to errors or even of cheating
- Sample sizes can be small in low-traffic conditions or for certain road-user groups
- The results require statistical weighting to account for differences in traffic exposure between sites and time periods (and the weighting factors can be inaccurate).

Despite these disadvantages, observer-based measurements are still the most used method for most behavioural SPIs, with the exception of speeding.

5.5 Recurring challenges for SPI-oriented behavioural studies

Several challenges arise repeatedly in the design and execution of behavioural SPIs. Acknowledging them up front is the first step towards mitigating them in any specific study. On the methodological side these challenges include:

- An SPI may be biased because the underlying dataset is not representative (and cannot be corrected by weighting). For example, biases may emerge when too few sessions in rural areas were included, or when all observations on busy roads were conducted on the same day of the week.
- The indicator may be insufficient for policy because it cannot be broken down into useful subgroups, for instance by age group or region.
- Decisions about methodology may have been driven by cost rather than by accuracy, with the result that the sensitivity of the indicator is too low and the confidence intervals too wide to detect real trends over time.
- Data collection and analysis can be time-consuming and costly, particularly for low-prevalence behaviours where very large samples are needed for narrow confidence intervals.
- The design of an appropriate sampling strategy and weighting approach can be difficult in countries where comprehensive traffic data are not available.

Challenges on the policy and institutional side include:

- There may be disagreements between stakeholders about which SPIs to select, or about the value and nature of the targets that are set.
- The SPIs may not be sufficiently integrated into the national road-safety strategy or plan.
- The frequency of SPI measurements may not be aligned with the policy monitoring cycle, so that results arrive too late to inform the next round of decisions.
- SPI values may not, in themselves, point to specific countermeasures, so that the indicator confirms the existence of a problem without indicating what to do about it.
- Topics for which no SPI is available may be overlooked in the policy debate.
- Authorities may be reluctant to adopt SPIs and set targets because they are then held accountable for the results.
- The set of SPIs may not keep pace with changing risks and new road-user groups (for example, riders of personal mobility devices).
- Responsibilities for data collection, data analysis, policy integration and dissemination may be unclear or fragmented.

None of these challenges should be regarded as decisive arguments against using SPIs. They are reasons to design the SPI system carefully, to invest in stable institutional arrangements, and to keep the methodology under continuous review.

5.6 Focus groups as a qualitative complement for behavioural SPIs

Observation studies deliver representative quantitative data on people's behaviour on the road, but they do not deliver the explanatory depth that policy designers and campaign developers sometimes need. The focus group approach is a qualitative research method where a small group of people are brought together to discuss their attitudes, opinions, or reactions to a specific topic. Focus groups can deliver the kind of explanation and contextualisation that quantitative methods cannot reach. Where observations and surveys answer 'what?', focus groups help answer 'why?'. Focus groups may run before a survey (see Chapter 9), in order to identify themes and refine the questions, or after a roadside measurement or survey, in order to explain and contextualise quantitative results.

A focus-group study begins with a clear research question: which knowledge gap are the focus groups intended to fill? Examples of road safety topics that are well suited to focus groups are:

- why drivers do not wear seat belts even when they know the risks
- the social norms and peer pressure surrounding drink-driving
- perceptions of police enforcement and fairness
- the attitudes of young drivers towards speed
- the barriers to helmet use among motorcyclists
- the cultural factors shaping attitudes to traffic rules
- reactions to a specific road safety campaign
- the reasons why people use their phones while driving.

When designing focus groups, the first step is to design target groups, for example young male drivers, professional drivers, motorcyclists, pedestrians, older drivers. Then the number of groups is set; for most road safety topics, four to eight groups are sufficient to reach thematic saturation, with additional groups if several distinct segments are to be studied. The optimal group size is six to ten participants; fewer than five limits discussion dynamics, while more than ten makes the group difficult to manage. Recruitment is purposive and criterion-based, with explicit eligibility criteria (for example, holds a driving licence, drives at least once a week, age 18–35). Self-selection bias is avoided by using structured recruitment channels (driving schools, employers, community organisations, police referrals) rather than relying on volunteers responding to a public call. Each potential participant is screened in advance to confirm eligibility. A 20 to 30% over-recruitment buffer is used to cover no-shows.

The session itself is run by a trained, neutral moderator and a separate note-taker. The moderator uses open-ended questions, ensures that all participants contribute, prevents dominant speakers from monopolising the discussion, and avoids expressing personal opinions or showing approval or disapproval. The discussion guide is structured in three phases: a warm-up with easy, non-threatening questions; a main phase that moves from general to specific and from factual to attitudinal to normative; and a closing phase with a summary and a brief member-check. Sensitive topics are approached gradually and indirectly. Each session lasts 60 to 90 minutes; beyond 90 minutes, participant fatigue degrades quality. Audio recording is used as standard, with a back-up device. Video recording, where used, requires specific explicit consent.

Audio recordings should be transcribed verbatim. Transcripts should be anonymised, and a thematic analysis framework is applied to identify recurring themes and contrasting views across groups. Findings are compared across groups to distinguish consistent themes from group-specific views. Negative cases that contradict the main themes are sought out and reported because they often reveal important nuance.

The analysis acknowledges explicitly the limitations of focus groups: the findings are not statistically representative. Focus group findings are then linked back to the quantitative findings from roadside observations and surveys: they are used to interpret unexpected or unclear results, to identify themes that should be incorporated into the next survey, and to provide the qualitative reasoning that explains the patterns shown in the SPI numbers. The combined report distinguishes clearly between quantitative evidence (what the data show) and qualitative insights (why this may be the case).

6 Sampling strategy for behavioural SPIs

6.1 Stratification and sampling of observations

6.1.1 The need for sampling

An SPI value is intended to represent the average behaviour of a population of road users such as all car drivers in a country, or all motorcyclists using public roads. It is not feasible to measure the behaviour of every road user at every moment on every road. SPI values are therefore estimates based on a representative sample of vehicles and/or road users observed or surveyed during a defined fieldwork period. Determining the appropriate sample size, the selection of locations, and the time periods for data collection is therefore essential to obtain representative and reliable data. Because samples cannot be drawn proportionally from every part of the network, the results of different sites must subsequently be weighted to reflect actual traffic exposure.

Sampling of observed road users within a stratum needs to be carried out at two levels:

1. the random selection of observation locations within each road type considered
2. the random selection of vehicles or road users within each observation session.

In Section 6.2 the sampling of locations is discussed. Section 7.3 explains how to sample vehicles within each observation session.

Most of the methodologies for sampling and weighting that are presented in this manual were originally developed in the SafetyNet project (Hakkert and Gitelman, 2007) and were further refined in the Baseline and Trendline programmes. They have also been applied successfully outside Europe, for example in the Western Balkans (Westbelt project) and in Tajikistan.

6.1.2 Stratification dimensions

In SPI-oriented behavioural studies it is common practice to define a number of strata. A **stratum** is the basic unit of analysis for SPI calculations. For each stratum, an SPI value can be calculated as the weighted average of all the session averages of the observation sessions in that stratum.

One example of a stratum is a combination of a road type and a time period. Other stratification variables such as vehicle or road user category can also be used as dimensions of a stratum. A simple combination of two road types and two time periods produces $2 \times 2 = 4$ strata; three road types and two time periods produce $3 \times 2 = 6$ strata; adding a vehicle-type dimension with two categories raises the figure to $3 \times 2 \times 2 = 12$ strata. Each additional dimension multiplies the number of strata, and each stratum requires sufficient observations of its own.

Stratification choices have to balance analytical richness against the realistic sample size that can be collected. As a general rule, strata with fewer than 500 observations should not be considered for calculating separate SPI values, because the number of distinct observations and locations is too small and the resulting confidence intervals are too wide. Such strata can, however, be combined with other strata to reach the minimum threshold.

Once the SPI values for the different strata are known, a national SPI value can be calculated across all road types, time periods and possibly even vehicle types. This requires a second weighting, this time a weighting of the SPI values of the different strata considered. Hence, each individual observation should be weighted: twice: (1) within-stratum weighting (using the traffic volume passing at the observation site during the session as the weighting factor), and (2) between-stratum weighting (using the relative traffic exposure of each stratum as the weighting factor). This will be discussed in Section 8.2..

To ensure representativeness, the sampling frame is typically stratified by

1. Road type
 - Motorways
 - Rural roads
 - Urban roads
2. Time period
 - Day vs night
 - Weekdays vs weekends
3. Region (optional)
4. Vehicle or road user type (optional).

Each stratum represents a combination of these factors (e.g. urban roads at night). If you have 3 road types, 2 time periods and 3 regions, you need $3 \times 2 \times 3 = 18$ strata. Each stratum should have a sufficiently high sample size (see Section 6.3).

The theoretically optimal strategy for estimating overall prevalence is to sample each stratum in proportion to its share of traffic volume. In practice, strict proportionality allocates too few observations to low-volume strata (minor road types, weekend periods) that are themselves of policy interest, producing confidence intervals too wide for the targeted comparisons that motivated the disaggregation. Trendline therefore deliberately accepts disproportionate sampling: minimum number of observations are imposed on policy-relevant strata even where this overshoots their traffic share. Proportionality is restored at the analysis stage through weighting (see Section 8.2).

When applying regional stratification (in addition to other stratification factors) the national minimum sample sizes should ideally be applied in each region separately, not just at the national level. Otherwise, the regional figures will not be reliable enough to support the comparisons that motivated the regional breakdown in the first place. Countries and regions are also free to add supplementary stratifications (e.g. by vehicle type, by age group) provided that the same minimum-threshold logic is then applied within each added stratum.

6.1.3 Obtaining a representative sample when road users are sparse

The helmet SPIs are an illustration of a context where the target groups are an often small and very unevenly distributed share of traffic. A purely random draw over the whole road network would place most observation points on roads where cyclists, motorcyclists or PMD riders are rare or absent, so that many sessions would yield few or no usable observations and the minimum sample would be reached only at disproportionate cost. The same problem would arise if one would want to measure buses or other particular types of vehicles.

This problem can be addressed by combining a random draw with a relevance filter, so that the sample remains representative of where the target group actually rides rather than of the network as a whole. Three principles need to be respected:

- The sampling frame is conceived as the part of the network on which the subcategory rides. For example, motorways are not part of the cyclist sampling frame at all, and the PMD sampling frame is restricted to the urban hotspots where PMDs concentrate (Section 12.4.3).
- When selecting locations, the relevance of a site to the subcategory is applied only at the last step, when the operational location is chosen in the vicinity of each random point.
- Where a randomly drawn point falls on a road with almost no traffic of the riders targeted, a nearby road of the same type carrying more of the target group may be substituted, on the assumption that most riders on the low-flow road also use the higher-flow road.

In this way, the relevance filter is applied to *where* observation takes place and never to *what* will be recorded there: once a site is selected, every passing member of the target group is observed in turn, so that sites are chosen for the presence of riders but helmet use is never pre-judged by the choice of site. Sites with a known or perceived helmet problem, or an atypical rider mix, are avoided for the same reason.

Because the resulting sample is disproportionate by over-representing the busier roads and, for cyclists, the urban network, the representativeness at national level should be restored at the analysis stage by weighting each stratum by its share of the target group's traffic volume (Chapter 8), not by its share of the road network or of all motor traffic. The substitutions made, the flow threshold applied, and any stratification level that could not reach its minimum are recorded in the metadata, so that the limits of representativeness are transparent. Where even this approach cannot deliver a representative national picture, for instance in countries with little micromobility or cycling, the indicator is reported for the available strata only and no national value is formed, rather than presenting a national figure that rests on too few or too selective locations.

6.2 Organisation and planning of observation sessions

In operational terms, sampling for behavioural SPIs can be organised in a few sequential phases. The phases are interdependent: poor execution of any one phase weakens the entire chain, and there is no shortcut that recovers representativeness once the underlying sampling frame has been compromised.

6.2.1 Defining the sampling frame

First, the study population needs to be specified:

- which road users or vehicles are relevant (passenger cars, powered two-wheelers, all motorised vehicles, cyclists, ...)
- what is the geographical coverage (national, regional, or specific corridors)
- whether all administrative units must be represented.

The road-type categories are defined using the standard broad Trendline categories (urban roads, rural roads, motorways where relevant), and the stratification variables are chosen: at minimum road type, optionally also time period and vehicle type. Each combination of stratification variables forms a stratum, and at least 10 observation locations per stratum should be planned. Minimum sample-size targets (overall and per road type) are set at this stage as lower bounds, not as final ambitions.

Where a country historically uses a different road categorisation from the standard three, an attempt should be made to infer the standard categories from the national typology, using characteristics such as traffic signs that demarcate the inside-versus-outside-built-up-area boundary, the speed regime, physical separation (single- or dual-carriageway), or the number of lanes. Whatever categorisation is adopted, and whatever deviations from the minimum requirements result, must be explained in the methodology section of the report.

A practical question arises in some smaller countries where the national network does not contain motorways, or where motorways only represent a very small portion of the traffic and road network length. In such cases the overall result is calculated using only the remaining road types, and the requirement to provide separate motorway figures is waived.

6.2.2 Selecting observation locations

Within each stratum, candidate locations are drawn initially by simple random sampling or by systematic random sampling (for example, every n^{th} kilometre of road). A Geographical Information System (GIS) is the most convenient tool for this phase. When a GIS is not available a national digital map service or a public mapping service can be used. At least 10 locations per stratum are selected. This approach may however yield locations with very low traffic or have an overrepresentation of locations with certain characteristics.

Although a fully randomised approach to location selection is methodologically ideal, it often leads to many practical problems. Some sites have very limited traffic or a specialised traffic mix (e.g., the entrance to a large factory).

Finally, the time investment and travel cost required to reach particular location also influences the practical feasibility of the sampling plan. Therefore, the random set of locations, especially when the number is low (10 to 20 locations per stratum) is often adapted to take into account:

- geographic spread across all administrative units or regions
- sufficient variation in terrain (plains, hills, valleys, ...)
- a mix of urban agglomerations and smaller towns or villages
- variation in expected traffic volume (and excluding sites where fewer than 20 relevant vehicles pass per hour)
- inclusion of different road-use functions (city-centre links, daily commute routes, commercial routes, tourism routes) are included where possible
- sites in the same stratum should be far enough apart to avoid re-sampling the same micro-context..

Additional considerations for selection and inclusion of locations include:

- The site must genuinely represent the road-type stratum to which it is assigned: an "urban road" site should reflect typical urban driving conditions, not an unusual access road or service lane. The local speed limit should match the road-type definition (urban ≤ 50 km/h; rural typically 70–90 km/h; motorway ≥ 110 km/h).
- Practical logistics should be considered, such as feasibility within the fieldwork plan and budget, mobile-phone coverage for communication with team leaders, parking space for observers if they arrive by car and capacity to accommodate the team size.
- Ensuring the safety of the observers is crucial, especially on rural roads and on motorway approaches. This topic is discussed more fully in Section 7.3.3. When planning observation sites, locations that might imply safety and security risks for observers should be excluded.
- The vehicles should be clearly visible by the observers. They should have a clear, unobstructed sightline to the observation trigger point and be able to see the driver (and for seat belts and CRS also the other vehicle occupants) clearly for three to five seconds as vehicles approach. The observation angle must be appropriate to the behaviour being measured. For seat belt and mobile-phone observations, the driver-side window area must be clearly visible; for helmet observations, the rider's head must be visible. Adequate lighting is required for all planned session times, including evening sessions. If tinted windows, obstructions or vehicle speed mean that more than approximately 15 to 20 % of passing vehicles cannot be observed reliably, the site should be reconsidered. Very high traffic speeds make reliable behaviour observation impossible; in such cases the site should be reconsidered or cameras should be used.
- The site and the presence of observers should not make drivers and riders change their behaviour. Sites should not be in immediate proximity to a speed camera, a police checkpoint, or any other enforcement point that would cause drivers to alter their behaviour for the measurement.
- Sites with persistent obstructions (permanent road works, regular large-vehicle parking) should be avoided. The same holds for sites with atypical traffic for the type of road (e.g. dominated by a single non-representative vehicle type such as a heavy-goods-vehicle depot access road)
- For speed SPIs, sites near speed cameras are excluded because only free-flow behaviour is relevant.

In order to verify whether such criteria are fulfilled, each candidate observation site should be assessed by desk reconnaissance (satellite imagery, street-view tools) or by a physical pre-visit.

Each selected location should be documented with its basic characteristics: coordinates wherever possible, address or other geographical reference, number of lanes, target lane and direction of observation, and visibility of the traffic from the chosen observation point. Photographs of the observation location are also very valuable. These metadata are essential for replication of the study in subsequent observation studies and for the audit of the sampling design.

A reserve list of 20–30 % additional candidate locations should be prepared so that unsuitable sites can be replaced. Indeed, when implementing the fieldwork location substitutions may become necessary because of unanticipated factors such as road works, weather conditions, mass sport events that close roads, or local visibility problems on the day. Such substitutions should always be documented, and the actual locations and dates used should appear in the metadata of the final dataset, even where they differ from those originally planned.

6.2.3 Time-period and session planning

An observation session is the set of observations made at one location during one defined time slot. A single location may host several sessions (e.g. morning and afternoon, or weekday and weekend). Once the number and distribution of observations sessions are chosen the time periods for the observation sessions should be determined.

The preference is for periods of normal, typical traffic, which in most countries means late spring or early autumn. As a default, fieldwork avoids December and January (adverse weather, short daylight) and July and August (holiday traffic), together with national school and bank holidays and any spell of severe weather; in some countries June is also a holiday month and is treated likewise. An SPI chapter departs from this default only for a stated, indicator-specific reason (for example, the cyclist-lighting indicator, which is measured in the dark and therefore favours late autumn and winter). Where a country holds a historical series, it retains the same window for comparability, and where two waves are run, the minimum sample applies to the combined dataset.

Time periods must be linked to locations in a balanced way. A design in which all weekday sessions occur at urban sites and all weekend sessions at rural ones would conflate road type with time of week and would be uninterpretable. For most behavioural SPIs the time-period split is weekday versus weekend. Driving under the influence of alcohol requires four periods (weekday daytime, weekday nighttime, weekend daytime, weekend nighttime) because alcohol risk varies sharply between them. For some SPIs (e.g. distraction) only weekday daytime is considered.

Sessions are planned across multiple time periods: at minimum morning, afternoon and evening, with night sessions added when nighttime measurement is needed (e.g. alcohol or drugs). Atypical timing and sites (e.g. immediately next to schools at opening hours) should be avoided unless specifically required by the SPI definition. Each session should last a minimum of 30 minutes (without the traffic counting), but longer sessions of one hour are preferred. Traffic counting is planned for every session, either by a separate team member or by counting before and after the observation.

6.3 Minimum numbers of observations

For each stratum in a behavioural SPI study the following minima should be set:

- the minimum number of individual observations.
- the number of observation sessions.

As to the number of observations, there is no closed-form formula that yields a single "minimum sample" guaranteeing national representativeness, because the answer depends on between-location variance, which is unknown until data are collected. The 'textbook' approach to determine minimum samples would require doing a pilot a study to learn about variance and then apply power analysis to count sample size properly. But in practice, this approach is not much used in SPI studies.

Therefore, the minima that are set within Trendline are pragmatic thresholds that balance affordability of fieldwork with the need for meaningful international and historical comparisons. The Trendline minima for the minimum number of observations and the minimum number of observation locations were set at a deliberately moderate level so that all participating countries could meet them. These minima are shown in Table

6-1 (for speed, higher numbers are required – see Chapter 10). It is reminded that there can be more than one observation session at an observation location (e.g., one on a weekday morning and one on a weekday evening).

Table 6-1. Typical Trendline minimum sample size requirements

	Trendline minimum
Minimum number of observations (all road categories and all strata together)	2,000
Minimum number of observations per road type	500
Minimum number of observations for each selected time period	500
Minimum number of observations for each stratum selected	500
Minimum number of observation locations	10
Minimum number of observation locations per time period	10
Minimum number of observation locations per combination of road type and time period	2

In practice, these minima are often exceeded, sometimes to a large extent, in particular for speeding. However, for some road-user groups such as motorcyclists, cyclists, and children in restraint systems, meeting the minimum thresholds may be too expensive, and a judgement must be made about whether the resulting indicator can still be reported reliably or whether the sample should be combined with the next measurement round. SPI-specific requirements and allowed reductions of the minima for particular breakdowns are presented in the SPI-specific chapters.

7 Fieldwork: data collection and data recording

This chapter describes the common elements to be considered for the fieldwork and data collection for behavioural SPIs. SPI-specific considerations are dealt with in the SPI-specific chapters.

7.1 Preparation of the fieldwork

Before data collection begins, a detailed fieldwork plan must be prepared. This plan should include:

- the list of selected observation locations per stratum
- the planned observation periods (dates, times of day) for each observation location
- the number of observers or field teams required
- the equipment needed (forms, tablets, breathalysers, cameras where applicable)
- safety arrangements for observers.

An 'observation session' (also referred to simply as a 'session') is the period during which a number of observations are made at a particular location and on a particular date. Sessions are the operational unit of fieldwork and the unit at which traffic counts are recorded; they are also the unit used in the within-stratum weighting described in the chapter on statistical analysis. Fieldwork should be distributed across different days and times to avoid systematic bias (e.g. only measuring during good weather or low-traffic periods).

As stated in the previous chapter, it is generally recommended to organise observation sessions at the end of spring or at the beginning of autumn, when weather conditions are stable and traffic patterns are typical for the year. In principle it is possible that an observation team observes several behaviours at once at the same site (for instance, seat belt use and mobile-phone use), but this is not generally recommended. Indeed, each behaviour requires a different observation angle and a different set of decision rules, and combining them tends to degrade data quality.

One observation session should last at least 30 minutes; an ideal duration is one hour, excluding the time needed for traffic counting and for collecting metadata. The same observation location can be used for several sessions, preferably at different time intervals; this provides additional information at modest extra cost, since the travel time to the site is incurred only once.

Strong winds, heavy rain, very low or very high temperatures all degrade observer endurance and the data quality, and may require a session to be rescheduled. When traffic is insufficient at a site, the recommended on-the-spot decision is to extend the duration of the session rather than to reduce the threshold for inclusion (typically 20 observations).

7.2 Using observers for roadside measurements

7.2.1 Composition of the observation team

Observers play a crucial role in behavioural SPIs. Their task is to apply definitions consistently and record observations accurately. Poorly trained or unsupervised observers are the single largest source of systematic bias in roadside observation data. The recruitment and training of observers therefore deserve at least as much attention as the design of the sampling plan.

An "ideal" team consists of three persons:

- The first person observes the road users and decides on compliance or not; he/she communicates that with the second person
- The second person records the data on the compliance decision and the associated variables (e.g. gender, vehicle category, ...)

- The third person counts the traffic of the observed vehicles.

In practice, the following can happen:

- Often the first and second person change roles halfway the observation session.
- If the team consists of only two persons, traffic counting should be done before and after each observation period.
- In exceptional cases, such as low traffic or the absence of a team member, the observation and data recording can be conducted by one person. However, this should only happen in exceptional situations, since the presence of a second person is highly recommended, both for safety reasons and as a check on data quality.

A team leader should be assigned for each observation team; he/she is responsible for quality control, problem-solving in the field, and communication with the study coordinator. The total number of observers should be matched to the fieldwork plan with at least 10 to 15 % reserve capacity to cover illness and no-shows.

7.2.2 Selection criteria for observers

Observers should be adults of at least 18 years of age; for observations connected to roadside alcohol testing, a minimum age of 21 is recommended. Candidates should demonstrate the ability to concentrate on a repetitive task for 30 to 60 minutes without losing focus. A basic familiarity with road users, vehicle types and traffic behaviour is required, and prior experience in transport research, public-health research or similar fields is an advantage. Observers must be physically able to stand for extended periods outdoors in variable weather conditions, and they must be able to read and complete observation forms accurately in the language used for data recording. Finally, they should have no personal interest in the outcome of the measurements.

Where a country or region has prior experience in measuring SPIs, field researchers with a track record from earlier rounds are a natural starting point. In countries without prior experience Master's students (e.g. in Transport studies) can be considered as team members, but it is recommended to have a team leader that has more seniority. Such an arrangement combines the cost-effectiveness of student labour with the methodological reliability of supervised observation.

It is generally inadvisable to use staff from enforcement agencies for observations because they could be perceived by the public as linked to enforcement. The exception is data collection for SPIs on alcohol or drugs, where involvement of the police is often mandatory.

7.2.3 Training of observers

The training of the observers serves three purposes: it ensures that observers understand why the measurements matter, it gives them a uniform interpretation of the observation protocol, and it provides hands-on experience with the data-recording system. Observer training is typically organised as a half-day to one-day workshop. Full-day training is recommended where child restraint systems are involved, because these have the most complex set of decision rules.

Training examples (videos, photographs or drawings of typical and borderline observation cases) should be assembled in advance for use during the training. Paper forms are pre-printed and numbered, and any digital application should be installed and tested on all devices to be used. Training and support materials should be distributed in advance: all observers should receive the materials, observation forms, the variable coding guide, and the location map at least two days before the workshop.

Typical components of an observer training session are:

- **Introduction.** Observers are introduced to the purpose and context of the study, to the concept of an SPI and the use that will be made of the data. They are made aware that representativeness and data quality are critical: the data they collect will directly inform national road-safety monitoring. Ethical obligations

are made explicit at this stage: observations are confidential, observers may not share data with unauthorised parties, and they may not photograph individual road users.

- **Scope of the observations.** The observers need to know which vehicles are to be included or excluded. Some vehicles or drivers do not have to comply with certain rules (for example, police cars or emergency vehicles) and others may be excluded by design (for example, public transport vehicles or tractors in a study focused on private cars). These exclusions are clearly communicated, and excluded vehicles must be removed from both the observations and the traffic counts.
- **Methodology training (protocol).** Each observer is given a full walkthrough of the observation protocol: what is to be observed, in which direction, for which vehicle types, and in which order. Every variable on the observation form is explained in detail: the precise definition, what counts and what does not, and how to handle ambiguous cases. Specific decision rules for borderline cases are taught and discussed.
- **Traffic-counting.** Explanations are given on how to count, what to count, and how to record the count on the session metadata form. All fields of the metadata form are explained and each observer practises filling it in.
- **Equipment and data-recording.** Observers practise using the paper forms or, where applicable, the smartphone or tablet application, during the training session itself. The correct use of high-visibility safety vests is explained. A back-up procedure is taught for equipment failure (form damage, drained battery, loss of connectivity).
- **Logistics and safety.** This covers roadside positioning, how to enter and leave an observation point safely, the minimum distance from the road edge, what to do if approached by a road user, a police officer or a member of the public, when to abandon a session because of insufficient traffic or weather, and emergency procedures including first-aid kit location and what to do if an observer is injured or witnesses an accident.
- **Hands-on training and calibration.** Observers should undertake short observation sessions. They should independently classify the same road users and compare results; any differences should be discussed to improve consistency. Where instrument-based measurement is used (e.g. breath testing), additional technical training is required.
- **Schedule.** Observers are informed or reminded on when and where they need to observe, who their fellow team members are, when they need to be present at the observation location, etc.

A laminated or printed one-page coding reference card is prepared for each observer to carry in the field. It is also useful to hand them over one or more checklists with to do's, frequently asked questions, contact information and a concise version of the observation protocol.

7.3 Observation procedure

7.3.1 Observation protocols

Each SPI study must have a written field protocol that specifies where observers should stand, which road users should be included or excluded, how to handle cases where behaviour cannot be clearly seen, and how to deal with unusual traffic situations. The detailed observation protocol depends on the behaviour being measured, but the basic operational principles are the same across all behavioural SPI studies. Observation sessions ideally last for one hour (but sometimes 30 minutes-sessions may suffice), with the exact duration noted on the metadata form.

Within each observation session, individual vehicles or road users must be selected for observation. In some cases, all relevant vehicles passing the observation point can be recorded. When traffic is light, it may be

possible to observe all vehicles in both directions; when it is busier, observation is normally restricted to one direction. Whatever the case, the approach taken should be documented in the metadata.

On busy roads it is not feasible to observe all vehicles. In such cases a systematic random selection is applied. This means that the observer starts with the first vehicle that can be observed and then selects the next vehicle that can be clearly observed. Care must be taken to avoid bias related to lane position or vehicle type; the selection protocol should be designed so that all lanes and all vehicle types have an equal chance of being included in the sample. This implementation has the advantage of being unambiguous in operation: the observer never has to decide which of several simultaneous vehicles to record. It also handles the case of bunched arrivals at signalised intersections naturally: the queue is processed in physical order, not by an arbitrary observer choice. As a consequence of this process, some vehicles will pass without being observed. This phenomenon is also the reason why separate traffic counting is needed (see next section).

Observers must not interact with road users (except in controlled roadside testing scenarios such as CRS use or DUI), must not influence traffic behaviour, and must not put themselves or others at risk. Observers must also be positioned so they do not influence the behaviour they are measuring. Two protocol elements support this requirement: concealment (the use of unmarked vehicles or hidden observers for speed and distraction observations), and the choice of viewing angle. For seat belt use and distraction by phone, a 45-degree angle on the oncoming traffic is recommended to see through the windshield and side windows.

Observer recording should be monitored and checked regularly during fieldwork so that data-quality problems are detected and resolved while they are still correctable, rather than after all the observations have been completed.

7.3.2 Traffic volume counting

Traffic counts make it possible to assess the relative ‘importance’ or the “weight” of each session within its stratum, since a session with high traffic represents more exposure than a session with low traffic. For example, for measuring DUI of alcohol or drugs, the count of total passing traffic is needed to interpret the share of drivers who were tested and the share of those who did not comply with the alcohol and/or drug regulations. The counting should include all relevant vehicles passing during a defined time period. Which vehicles are ‘relevant’ to include depends on the SPI. For example, for a study of helmet use, every powered two-wheeler riding in the same direction as those being observed should be counted.

If the observers are able to observe every passing vehicle (for example, in low-traffic conditions where every relevant vehicle can be observed and compliance can be recorded), no separate traffic count is required, because the count is by definition equal to the number of observations. In a team of three persons, the third person can count the traffic in parallel with the observation, so that the count is continuous over the full duration of the session.

It is recommended to count traffic for a minimum of 20 minutes during each observation session. This can be organised either as ten minutes of counting before and ten minutes of counting after the observation, or as a continuous 20-minute counting period in the middle of the session. The two-count design (one before, one after) has two advantages over a single longer count or a continuous count. First, it brackets the observation phase symmetrically and so averages out short-term fluctuations in flow; an unrepresentative ten-minute spike in either direction is partially offset by a more typical second count. Second, it gives the observer a cognitive bookend at each end of the session which can be useful for maintaining concentration during the observation phase itself. The trade-off is that the total session takes 80 minutes of observer time rather than 60.

7.3.3 Observer safety and security

Observer safety is a primary concern. Measures should include high-visibility clothing, safe positioning away from moving traffic, avoiding observation from hazardous locations (e.g. motorway shoulders without protection), and working in pairs where necessary. No observation should be conducted if conditions are unsafe

(e.g. severe weather, road works, or dangerous traffic conditions). Observers should have a safe off-road standing position that does not require them to stand in, or adjacent to, a live traffic lane, such as a verge, footpath, lay-by or area of raised ground. The parking and arrival route should not itself create a road-safety risk, and an identified and practicable emergency-exit route should be available. Each observation site should be formally approved on safety grounds by the team leader or study coordinator before any deployment.

Personal protective equipment, in particular high-visibility vests, should be provided to all field researchers, with the caveat that high-visibility equipment may itself influence the behaviour being observed in some specific cases (for example, drivers slowing down or putting down their phone on seeing a high-visibility vest at the roadside).

Observers should have a written declaration with them which explains what they are doing, confirms that they have all the required permissions to make the observations (where applicable) and states that no personal data will be stored. The declaration should also include a telephone number of the study team leader who can be called to confirm the statements on the declaration. The declaration can be shown to police officers or other authorities who are approaching the observers and asking them what they are doing.

7.4 Road-side data recording

7.4.1 Data recording approaches

There are several ways of recording the data:

- Recording the data on a paper form, with a pen or pencil
- Typing the data directly in an electronic document (e.g., a Word document, a spreadsheet) on a laptop or tablet)
- Using a tablet or mobile phone to record the data in a customised application.

In all three cases the next step is to transfer the recorded data to a database for further processing, and the ease with which this can be done differs markedly between the methods.

Paper recording remains the most robust and accessible option. It requires no power supply, is unaffected by screen glare in bright sunlight, functions in any weather provided the form is protected, and carries almost no risk of technical failure in the field. Pre-printed forms with clearly delimited fields can also be quick to complete once the observer is familiar with the layout. The principal drawback is that the data must subsequently be entered into a digital format by hand, a step that is time-consuming and a recurrent source of transcription error. Paper forms can also be misplaced or damaged, illegible handwriting may render entries unusable, and there is no automatic record of the time at which each observation was made. If paper forms are used, double data entry or systematic checks should be performed, and data entry protocols must ensure consistency.

Typing directly into a generic electronic document removes the separate data-entry step and produces a file that is already machine-readable. However, standard documents and spreadsheets impose no structure on what is entered: there are no controls to prevent invalid codes, out-of-range values or skipped fields, so errors are easily introduced and may go undetected until analysis. Operating a laptop or tablet at the roadside is also cumbersome, the device is exposed to weather and theft, and the observer's attention is divided between the screen and the traffic being observed.

Recording into a customised application on a tablet or mobile phone is generally the recommended approach. A purpose-built application can enforce coding rules at the point of entry, restrict responses to permitted values, require mandatory fields to be completed, and apply automatic time- and location-stamping, thereby reducing both transcription and coding errors. Data can be uploaded to the central database directly, often in near-real time, which shortens the processing chain and allows early checks on data quality. These benefits come at a cost: the application must be specified, developed and tested in advance, observers must be trained

in its use, and there is a dependence on battery life, connectivity and the device itself, all of which can fail in the field.

In choosing between the methods, the decisive factors are usually the scale of the survey, the available budget and technical support, and the environmental conditions at the observation sites. Digital data-collection tools are preferred wherever the necessary preparation and resources can be secured. Where this is not possible, a well-designed paper form, backed by prompt and careful data entry, remains a wholly acceptable alternative. It is also good practice to retain a paper fallback for use when devices fail.

7.4.2 In-field quality control

The quality of the fieldwork de facto starts with a solid preparation and training of the observers. As part of the observer training (see Section 7.2.3) a practice session should be conducted at a real roadside location. The practice session should use the actual data recording forms or apps and follow the full observation protocol. After the practice session, all observers' forms are reviewed: recording errors, omissions and inconsistent coding are identified and discussed as a group. Where two or more observers record the same vehicles independently, the percentage agreement between their records is calculated; a minimum agreement of 90 % on the primary behaviour variable is recommended before observers are deployed. Observers who do not reach this reliability threshold receive additional targeted coaching and must repeat the practice session before being deployed; no observer is allowed to collect study data before passing the practice session.

During the field work, quality control mechanisms should include:

- Observers should adequately use the paper forms or the specific software and correctly record the observations (see previous sections).
- Observers should record also the relevant metadata for each session, including at least: date and time; location (with geographic reference); road type; direction of traffic observed; weather and lighting conditions; names or IDs of observers. Observers should also note unusual events or problems encountered. This documentation supports transparency, quality assurance, and replication in future years.
- Supervisors should conduct occasional site visits to verify that the protocol is being followed. The supervisors could also perform a spot-check by recording a subset of the same vehicles independently and compare the records with the observers' ones.
- Double observation of approximately 5% of the sample is recommended as a routine quality-assurance activity. A Cohen's κ score of less than 0.80 between independent observers indicates that observer agreement is insufficient, and that re-training or replacement of observers is required.
- Data recorded should be checked regularly for inconsistencies or missing values.
- Observers should receive brief feedback, preferably after each day in the field, and persistent errors should be corrected before the next session. A log of quality-control checks is maintained for each observer and each session.
- If systematic problems are detected, additional training or replacement of data collection sessions may be required. Sessions should be flagged for exclusion if more than one third of the records have coding errors or missing data, if the observer did not follow the prescribed observation direction, or if weather or traffic conditions rendered observations unreliable.

After fieldwork is complete, a group debrief can be held with all observers and the team leader. Observers can report the difficulties they encountered during the fieldwork (for example, unclear vehicle types, ambiguous behaviours, access problems, unusual traffic conditions) so that these can be documented as study limitations. Observers should also be reminded of their continuing confidentiality obligations regarding the data and the individual observations.

7.5 Camera-based and technology-based observation

For some SPIs, roadside cameras or automated detection systems may be used to supplement or replace direct observation. In such cases, the detection algorithm must be validated, the definition of the behaviour must match the manual observation definition, data protection and privacy regulations must be strictly followed, and a methodological report describing the system must be provided. Automated tools can improve efficiency and objectivity but must remain methodologically transparent.

7.5.1 Cameras and AI-based image analysis as an alternative to human observation

Roadside cameras can be used to collect SPI data either as a complement or as an alternative to direct observation. Once captured, still images or video footage are analysed off-line in order to encode the relevant variables. The advantages, in principle, are considerable: the reliability of the recording is higher than for human observers, the number of observations per session can be larger, the duration of sessions can be extended (including into the night), and the same recordings can be re-analysed if a new SPI is added later.

The disadvantages, however, must also be evaluated carefully. The cost of cameras and of secure mounting equipment can be substantial. Many candidate locations do not offer a position from which a useful camera angle can be obtained. Cameras are exposed to a non-trivial risk of theft or damage. Drivers may modify their behaviour when they notice a camera, biasing the very observations the camera was deployed to capture. And privacy issues, in particular the identifiability of drivers, must be addressed before any deployment. For all of these reasons, camera-based observation should be tested and validated locally before being adopted as a standard method.

Semi-automatic detection of seat belt use and handheld mobile-phone use from camera images is the more recent development, made feasible by progress in image-recognition based on deep-learning techniques. A trained model can identify, from a high-resolution camera image of an approaching vehicle, whether the driver and the front-seat passenger are wearing seat belts and whether the driver is holding a mobile phone. Detection accuracy is now usable in well-lit, frontal-view conditions, although it deteriorates in poor light, with heavily tinted windows, or with unusual body postures. Validation against manual observation under controlled conditions remains essential before such systems can replace human observers. Moreover, it may be challenging to apply the principles of stratification and random selection of locations as can be applied with human observers. Other recent developments include automatic detection of motorcycle helmet use and have been proven useful in countries with high numbers of mopeds and motorcycles on the road; this is covered in Section 6.

7.5.2 Other technology-based observation approaches for speeding

Several technology-based observation tools have already been applied for speeding SPIs:

- Using roadside **radars**. This has been the standard technology-based observation method for vehicle speed for several decades. The advantage over visual observation is decisive: a radar records every passing vehicle, so that sample sizes are limited only by the duration of the measurement window rather than by the observer's processing capacity. The challenges are the cost of installing and maintaining the equipment, the limited range of variables that can be captured (essentially only speed and rough vehicle classification), and the need to ensure that the chosen sites are not affected by speed-camera presence, by congestion or by other factors that would render the speed observations atypical.
- Using handheld **laser guns**. For the speed of motorcyclists and LGVs, laser-gun measurement is the established method. The challenge is that motorcycle speeds vary much more than passenger-car speeds within a given site, and that motorcycle traffic is much sparser, so that the operational logistics of motorcycle-speed observation differ from those of car-speed observation. The general framework, however, is the same.

- Using data from **induction loops**. The advantage is that a massive amount of data is available whole year round, but the main challenge is that the location of induction loops only partially corresponds with the type of locations needed for SPIs. Only for motorways in some countries it may be possible to calculate a robust SPI for speed.
- Using the **GPS location** of vehicles. This method has been successfully applied in some countries. The GPS data allow estimating speed of specific vehicles (known as floating cars, probe vehicles or connected vehicles). However, these constitute only a sample of traffic flow, and it is not possible to apply any selection rules, such as avoiding congestions, distinguishing vehicle categories, etc. Thus, this approach requires a new interpretation of the concept of free flow and also a stability over time in the underlying fleet to cope with inherent bias (unclear representativeness of the fleet).

7.6 Ethical and legal requirements

7.6.1 Legal framework and obtention of authorisations

Ethical and legal compliance is not optional in behavioural studies: violations can invalidate a study, expose the research team to legal liability, and undermine public trust in road-safety research. The principles set out in these sections apply to every behavioural study of the type covered in this manual, irrespective of whether formal ethics-committee approval is required in the country concerned.

First, the legal framework must be identified before fieldwork begins. European wide principles such as GDPR and national data-protection legislation and any traffic or public-order laws relevant to roadside studies should be checked and reviewed. Where required, written permission should be obtained from the road authority, the municipality or the landowner for each observation site (e.g., to place cameras or radars, or to observe vehicles from the footpath).

Where cameras or video equipment are used, the legal basis for recording in public spaces should have been confirmed; signage informing the public that recording is in progress should be displayed where this is required by law; and recordings should be stored securely, access-restricted, and deleted within the retention period stated in the study protocol.

For studies involving driving under the influence of alcohol, a formal written agreement with the national police, covering scope, protocol and data-sharing arrangements, should be in place before fieldwork starts.

Any required ethics-committee approval should be obtained before the observations start. Whether or not an ethics committee approval is needed varies by country and sometimes even between research institutes within a country. Where the national data-protection authority requires notification or registration of studies involving personal data, the required notification should be filed before data collection begins.

7.6.2 Personal-data principles and GDPR compliance

Behavioural SPI measurements often involve observing or testing individuals in traffic. Therefore, data must be collected anonymously, no personal identification may be recorded, national legal frameworks for roadside testing (e.g. alcohol checks) must be respected, and observations must not interfere with normal traffic flow or safety. Ethical approval may be required in some countries, particularly for roadside testing or survey components.

Only the personal data strictly necessary for the study objectives should be collected: variables like gender, approximate age category and vehicle type should only be recorded if one plans to produce breakdowns based on these. Also, data collected for the study can only be used for the stated research purpose. They cannot be shared with enforcement authorities, insurers or any other party for secondary purposes. No direct identifiers such as licence plates, names, or images that allow identification of individuals can be recorded as part of the SPI observation data.

A named data custodian should be appointed and is responsible for the security of all study data from collection to deletion. Paper observation forms should be stored in a secured location immediately after each session and transferred swiftly to the study coordinator. Electronic data should be stored on password-protected devices or secure cloud storage, with access limited to named study personnel. A data retention schedule should be defined and a data-breach response procedure should take place.

7.6.3 Consent requirements

Consent requirements differ by data-collection method. Roadside observations in public traffic spaces are generally conducted without individual consent, on the legal basis that road users are acting in a public space; this does not, however, remove other ethical obligations. The legal basis for passive observation must be confirmed under national law. Studies should be publicly announced in advance where this is required. Observers should not stop, detain or approach road users during the observation session.

For surveys, a clear informed-consent statement is presented to all respondents before the questionnaire begins, covering purpose, who is conducting the study, how the data will be used, that participation is voluntary, and the right to withdraw. For online surveys, respondents must actively confirm consent. For focus groups, written informed consent is obtained from every participant before the session begins, with separate consent for audio (and where applicable video) recording.

Special categories of participants require additional protection. If a study includes individuals under 18, parental or guardian consent is required for surveys and focus groups. For studies involving professional drivers, care is taken that the data cannot be used by employers for individual performance monitoring or disciplinary action. Roadside alcohol testing may only be conducted by law-enforcement officers with the legal authority to stop vehicles; research staff do not stop vehicles independently. Table 7-1 summarises typical consent requirements and personal-data risk by data-collection method.

Table 7-1. Typical consent requirements and personal-data risk by collection method

Data-collection method	Consent requirement	Personal-data risk
Roadside observation	None required — announce publicly if feasible	Low (no identifiers recorded)
Observations with camera	Public notice recommended, legal basis confirmed	Medium (facial / plate data possible)
Online or telephone survey	Active opt-in consent before questionnaire	Low if anonymised at source
Face-to-face interview	Verbal consent documented before interview	Medium (individual contact)
Focus group	Written consent; separate audio consent	Medium (voice and views recorded)
Police DUI stop-and-test	Police authority replaces consent; data minimised	High (health data; police data)

7.6.4 Observer confidentiality and ethical conduct of analysis

Observers are bound by confidentiality. All observers should sign a confidentiality agreement before deployment and be briefed on its requirements. Observers should not discuss, share or comment on individual observations outside the official data-recording process; they should not use any data collected for personal purposes (for example, reporting a specific vehicle to the police). If an observer witnesses a traffic crash or an immediate threat to life during a session, the session is suspended and the emergency procedure takes precedence over data collection.

Ethical conduct extends to data analysis. The study cannot create enforcement obligations: data collected for research cannot be used to initiate enforcement actions against specific individuals. Results that are unfavourable to the commissioning authority are reported with the same accuracy and completeness as favourable results, and the study team should resist any pressure to suppress or modify findings. Any potential conflict of interest should be disclosed in the study report.

7.7 Collaboration with the police and other authorities

7.7.1 The formal agreement

Several behavioural SPIs can only be measured with the active involvement of the police or other authorities. The clearest case is objective roadside testing for alcohol or drugs, where in nearly all Member States only the police hold the legal power to stop a moving vehicle. But collaboration with a road authority, a municipality or a landowner is also needed wherever equipment must be installed or vehicles observed from private or restricted ground.

Before any joint fieldwork begins, a written agreement should be concluded between the research team and the participating authority. A verbal understanding is not sufficient, because the responsibilities, the data flows and the liabilities involved are too consequential to rest on memory or goodwill. The agreement need not be long, but it should state, as a minimum:

- the parties to the agreement and a single named contact on each side;
- the scope and purpose of the collaboration — which SPI is being measured, over which period, and at how many sessions — together with an explicit statement that the data are collected for research and monitoring and not for enforcement targeting;
- the division of operational roles: who selects the locations and time periods, who stops the vehicles, who administers the test, who records the result, and who carries out the traffic count;

- the data-sharing arrangement: which variables are recorded, in what anonymised form, who holds the data, and on what legal basis they are processed
- liability and insurance, covering the safety of the research staff working alongside an enforcement operation and the handling of any incident during a session;
- the publication and confidentiality terms, including the principle that unfavourable results are reported with the same completeness as favourable ones.

A short standard template carrying these headings should be prepared once and reused, so that each new collaboration starts from an agreed basis rather than being negotiated from scratch. The signed agreement is referred to in the study metadata.

7.7.2 Random versus selective enforcement, and its consequence for the SPI value

The single most important methodological point in any police-supported SPI measurement is the distinction between random and selective enforcement, because the two produce fundamentally different quantities. In their normal duties the police practise selective enforcement: they stop the vehicles they have reason to suspect. An SPI, by contrast, requires random selection: every driver in the traffic stream must have the same chance of being included, irrespective of any suspicion of impairment. A measurement built on selectively stopped drivers does not estimate the prevalence of the behaviour in normal traffic; it estimates the prevalence among drivers the police considered suspicious, which is a different and much higher figure, and it cannot be compared over time or between countries.

Collaboration therefore requires the police to depart from their normal practice for the duration of a research session and to stop drivers at random — in the ideal case every passing driver, and where volume or capacity prevents this, the next driver to arrive after each completed test. Because this departs from routine practice, it is strongly recommended that a researcher attend each session to confirm that selection is genuinely random, that the correct location and time period are used, and that no targeting of “suspicious” drivers occurs. Where a researcher cannot attend, the session data should be screened afterwards for signs of selectivity, such as an implausibly high proportion of drivers over the limit. The agreement should record explicitly that selection during research sessions is random, so that the operational officers understand that this is a condition of the measurement and not a discretion left to them.

7.7.3 Confidentiality protocol for enforcement outcomes

A research-supported operation will, by its nature, generate enforcement outcomes: some tested drivers will be over the limit and will be sanctioned by the police under their normal powers. This creates a tension that must be managed explicitly, because the research data and the enforcement action concern the same individuals at the same moment. The governing principles are the following.

- The sanctioning of an individual driver is a matter for the police under their own legal powers and procedures; it follows a positive test exactly as it would outside a research operation. The research team neither directs nor obstructs it.
- The research dataset, however, is anonymous. No direct identifier — name, licence plate, or any value that would allow an individual driver to be identified — is recorded in the research data, even though the police record holds such information for their own enforcement purposes. The two records are kept separate, and only the anonymised research record passes to the analysis.
- Research data cannot create or feed an enforcement obligation. Data collected for the SPI cannot be used to initiate, support or revisit enforcement action against any individual, and cannot be shared with enforcement authorities, insurers or any other party for a secondary purpose (see §6.6.2 and §6.6.4).

- Where a researcher is present and becomes aware of an individual enforcement outcome, that knowledge is treated as confidential and is not recorded, discussed or transmitted outside the anonymised research process.

These arrangements should be written into the agreement so that both parties understand, before fieldwork begins, that the enforcement and research streams run in parallel but never merge at the level of the individual driver.

8 Data analysis for behavioural SPIs

8.1 Creation of the database with observation data

8.1.1 What an SPI value is, statistically speaking

Behavioural SPI values are estimates. The values that the methodology aims to estimate are the proportions of vehicles or road users on the entire road network of a country over a defined reference period (typically one year) that comply with a given safe behaviour such as wearing a seat belt or travelling within the speed limit. The statistical population to be considered is, formally, the total traffic volume of moving vehicles over the geographical area concerned during the reference period; this is most naturally expressed in vehicle-kilometres driven. The estimator is constructed by sampling individual vehicles or road users at particular locations and at particular moments in time, and then weighting those samples to reflect the underlying traffic exposure.

Two consequences follow from this framing, and they shape the entire statistical apparatus:

- First, the SPI is not a count of compliant vehicles divided by the total number of vehicles observed; it is an estimate of the proportion of kilometres driven under compliant conditions. The two are equal only when the sample is exactly proportional to traffic exposure across all relevant strata, which it almost never is.
- Second, the precision of the estimate depends to a large extent on the number of locations sampled and on the variation between those locations. Confidence intervals must therefore reflect the variation between sites and sessions, not the variation between individual vehicles within a session.

8.1.2 Types of data required

Three types of data are combined in behavioural SPI studies:

- individual-level records (one observation for each road user or vehicle)
- session-level metadata (information about the observation session itself)
- study metadata (information about the whole dataset, legal context, timing and methodology).

At the individual level, each record contains a number of pre-defined variables, which typically include the following categories:

- identification variables such as location ID, session ID, date and observation time
- stratification variables such as road type (urban / rural / motorway), time category (day/night, week-day/weekend), region (if applicable), or vehicle category
- behaviour variables, i.e. the observed behaviour (typically binary, with 1 denoting compliance and 0 non-compliance), and, where applicable, the behaviour category (e.g. type of distraction)
- road-user variables such road-user category (pedestrian, cyclist, ...), vehicle occupant category (driver, rear passenger, ...) and optional demographic characteristics (gender, age group) if visible and can reliably be classified.

Session-level metadata typically include:

- session ID
- start and end time
- location information (address, GPS coordinates, location ID)
- observer IDs
- weather conditions
- traffic count and the timing and duration of the traffic counting.

Study metadata typically include:

- year and months of data collection
- data sources (if applicable)
- sampling design
- definitions used
- known limitations.

For most SPIs there are also additional contextual variables, such as BAC limits for the SPI driving under the influence of alcohol, speed limits for the SPI speed, and national regulations for the use of child restraint systems.

The number of variables recorded for a study on behavioural SPIs can vary. In a “minimalist approach” only the variables are stored that are absolutely necessary for being able to calculate the prime SPI values and interpret these correctly. When using a “maximalist approach” several other variables are also recorded, allowing a range of breakdowns and combinations of breakdown to be produced and deeper insight into the risk factor being analysed. In Trendline, these two options are explicitly foreseen: the “minimalist approach” corresponds with meeting the “minimum requirements”; this approach should be feasible in all countries. The optional additions that are listed for every SPI are part of a maximalist approach; but in general, only a limited number of these extra variables can actually be observed and recorded.

8.1.3 From field observations to database

The transformation of field observation data into an analysis-ready dataset proceeds in a number of steps. Each of these has its own quality controls, and each contributes to the integrity of the final SPI value.

1. **Data entry and database construction.** All observed data are collected, checked for completeness, and stored securely before data entry begins. Where forms are paper-based, the data are transferred to a structured electronic database by trained personnel. All column headers, data types and coding schemes should have been pre-defined. When data were recorded using software, the data need to be transferred from one database to another with a similar structure. All records should be assigned unique identifiers that enable linkage between observation data and session data.
2. **Data cleaning.** Cleaned data refers to a dataset in which improperly formatted or incorrectly recorded data have been corrected or removed. Typical cleaning activities are:
 - identification and removal of test or training records (observations made during observer training rather than during real fieldwork)
 - removal of records with missing values in key variables (such as the behaviour variable, the vehicle type, the location and the session ID)
 - range checks, which identify values falling outside the physically or operationally possible range for a variable (e.g., a blood-alcohol concentration of 5.0 g/l, for example, is physically impossible and must be flagged, as is an invalid road-type code)
 - consistency checks, which identify combinations of values that are individually plausible but jointly impossible or implausible (e.g., a child observed in a rear-facing child seat should not be coded as 'unbelted' where the seat is itself the protective device, and a 'helmet use' record should not appear for a car driver)
 - temporal plausibility checks, which identify combinations that are unlikely given the time and location of the observation (e.g., an average speed of 130 km/h recorded in an urban 30 km/h zone, for example, suggests either equipment malfunction or a misclassification of the site);
 - removal of duplicate records (the same observation entered twice).

Records that fail a check should be flagged and investigated, not silently accepted or removed. Where values are missing, the pattern of missingness should be examined and 'not observed' should be distinguished from 'not recorded', since the two have different implications for analysis.

If more than one third of the records in a particular session are identified as poor quality, it is good practice to exclude the entire session from further analysis. The presence of a corresponding traffic count should also be verified for every session; any session with missing traffic data should be flagged for investigation. All cleaning decisions, including any corrections and exclusions, should be recorded in a written cleaning log.

3. **Traffic-volume normalisation.** Before the count is normalised, the effective session length should be established. There may be a time gap between the nominal start of the session and the arrival of the first vehicle, and again between the last vehicle and the nominal end; these gaps do not form part of the session, which begins with the first vehicle coded and ends with the last. Session length should therefore be corrected by taking the span between the first and last coded vehicle and adding half of the average inter-vehicle time.
For each observation session, the total traffic count of relevant vehicles during this corrected session period is retrieved. The counts are then normalised to a standard time unit, typically the number of vehicles per 60 minutes, so that sessions of different durations can be compared and appropriately aggregated. It should be verified that the traffic count covers the same direction, lane(s) and vehicle types as the behaviour observations. If a traffic count is missing for a session, an assessment should be made of whether the session can still be included using an estimated count derived from comparable sessions. Such decisions should be documented.
4. **Stratification.** Every observation session is assigned to its correct stratum, defined by road type, time period and (where applicable) vehicle type and/or region. The total number of observations per stratum is counted. Strata with fewer than 500 observations should not be used to calculate separate SPI values; merging with adjacent strata is the standard remedy (for example, 'weekday daytime' and 'weekday nighttime' can be combined into 'weekday'). Strata with fewer than 10 observation locations should be documented as a limitation, and supplementary data collection should be considered. A stratum-summary table is prepared showing, per stratum, the number of sessions, the number of observations, the number of unique locations and the traffic count.

After cleaning, individual-level data are merged with session-level data, stratification and weighting variables are attached, and a final analysis dataset is produced. This dataset forms the basis for the statistical estimation set out in the following sections.

8.1.4 Coding of behavioural variables

Coding of data must follow the standardised definitions provided in the SPI-specific chapters. General principles apply across all SPIs:

- binary behaviours should be coded consistently (e.g. 1 = compliant, 0 = non-compliant, 9 = not observable)
- 'not observable' must be coded separately from missing data
- ambiguous cases should follow predefined decision rules.

Where behaviours have multiple categories (e.g. types of distraction), coding schemes must be mutually exclusive and must cover all expected observable cases.

Some behaviours cannot always be clearly observed. Rules should specify when an observation is classified as 'not observable', and whether such cases are excluded from analysis or treated as a separate category. The treatment of unobservable cases must be consistent across all observation sessions.

8.2 Weighting process

8.2.1 General principles and the two layers of weighting

Weights adjust the contribution of each observation so that the sample reflects the true distribution of traffic. For SPIs, two layers of weighting are required:

1. a **within-stratum** layer, which combines the different sessions in the same stratum
2. a **between-strata** layer, which combines the strata themselves into a national figure.

At the within-stratum level, observation sessions differ in duration and traffic volume. A session with high traffic represents more exposure than a low-traffic session, and the weighting should reflect this. At the between-strata level, some strata (e.g. urban daytime) represent a larger share of national traffic than others (e.g. rural nighttime), and stratum weights correct for unequal sampling across strata. The combined weight ensures that each observation represents the appropriate share of national traffic exposure.

Weighting factors for the observation sessions within each stratum are based on the traffic counts per session. In other words, sessions with higher (normalised) traffic counts have a higher weight within the stratum than sessions with a smaller (normalised) traffic count. For the overall national value of the SPI combining all strata, the strata values should be weighted on the basis of estimated traffic volume (vehicle-kilometres) per stratum. The ideal situation is using national traffic data per stratum. Sometimes official data are not available, but an estimate of the distribution of traffic volume is possible. Where vehicle-kilometre data are not available, road-network length can be used as a proxy, but this choice should be documented. If regional stratification is used, the analysis must weigh the regional results back together according to the regional traffic volumes when calculating any aggregated national figure.

8.2.2 Variables collected for weighting

Two sets of variables are collected for weighting purposes. The first set is recorded at the level of each individual session; the second is recorded at the level of each stratum. The session-level variables describe what was observed during the session itself; the stratum-level variables describe the structural characteristics of the road type and time period to which the session belongs. The notation given below follows that used in the original Trendline statistical reference document (Elgner et al., 2024), with each variable identified by a single symbol so that the weighting formulas in the next sub-section remain readable.

At the session level, four values need to be recorded for every session k (by convention, capital letters in this notation refer to the observation session and lowercase letters refer to the counting period):

- The first value is $t(k)$, the duration in seconds of the period used to count passing vehicles. This may be a single counting period or, more commonly, the sum of two counting periods, one before and one after the observation.
- The second value is $n_j(k)$, the number of relevant passing vehicles of vehicle type j counted during the counting period; if there are several counting periods within the session, $n_j(k)$ is the sum across all of them.
- The third value is $T(k)$, the duration in seconds of the observation session itself (exclusive of the time spent on counting).
- The fourth value is $N_j(k)$, the number of relevant vehicles of vehicle type j actually surveyed (that is, observed and coded) during the observation session. In cases where traffic is low and all vehicles passing by are observed, $N_j(k)$ is identical to $n_j(k)$.

From these four values a fifth value can be derived for each session: $ETN_j(k)$, the estimated total number of relevant passing vehicles of vehicle type j during the observation session: see Equation 1.

Equation 1. Estimated total number of relevant vehicles during the observation session

$$ETN_j(k) = n_j(k) \cdot \frac{T(k)}{t(k)}$$

Having an accurate session-level estimate of the total number of vehicles that passed the survey location, notated as $ETN_j(k)$, is essential, because it is this estimate that determines the share of the stratum that the individual session represents, and hence the weight assigned to it. For example, if the duration of the observation session – $T(k)$ – is exactly 60 minutes and the duration of the counting session – $t(k)$ – is exactly 20 minutes and the number of vehicles observed during the counting session – $n_j(k)$ – is 40, then $ETN_j(k)$ is $40 \times (60/20) = 120$. This will be the weighting factor for the within-stratum weighting.

In exceptional cases the counting period may be unrepresentative. For example, if a sudden traffic jam during the observation reduced the flow well below the count, or a one-off event such as a bridge opening occurred during the count itself. In such cases a more direct expert estimate of the session's passing traffic should be substituted.

At the stratum level, five values are needed for each stratum i :

- **Sl_{ij}** is the stratum-level weight characteristic for vehicle type j . It has to be estimated from the values below.
- **TR_{ij}** is the relative traffic volume of vehicle type j in stratum i , expressed as a share of the total of traffic volume (this requires to have a national estimate, for example based on traffic surveys). The TR_{ij} values should sum to 100 % over all included strata and vehicle types. If only one vehicle type is observed, the sum over all included strata for this vehicle type should be 100 %. If only a subset of strata is considered, e.g. daytime only, the TR_{ij} values should be normalised accordingly.
- **RL_i** is the total road length, in kilometres, of the road type in stratum i .
- **NS_{ij}** is the average number of vehicles of type j per hour for stratum i , derived from the count-period data of all sessions in that stratum. In other words, **NS_{ij}** is the average value of $ETN_j(k)$ within the sessions in stratum i .
- **TP_i** is the relative length of the considered time periods of stratum i — for example, if weekday daytime is treated as 48% of the year and weekend daytime as 19%, those proportions go in here.

For the speed SPI, **FSL_i** , the fixed speed limit in kilometres per hour for stratum i , is also recorded, since speed compliance is defined relative to it.

The different values and their meaning are given in

Table 8-1.

Table 8-1. Values needed for the weighting process

Symbol	Level	Meaning
$t(k)$	Session	Duration (s) of counting period(s) within session k
$n_j(k)$	Session	Number of vehicles of type j counted during the counting period(s)
$T(k)$	Session	Duration (s) of the observation session itself
$N_j(k)$	Session	Number of vehicles of type j actually surveyed during the session
$ETN_j(k)$	Session	Estimated total passing vehicles of type j during the session = $n_j(k) \times T(k) / t(k)$
Sl_{ij}	Stratum	Stratum-level weight characteristic for vehicle type j of stratum i
TR_{ij}	Stratum	Relative traffic volume of vehicle type j in stratum i
RL_i	Stratum	Total road length (km) of the road type in stratum i
NS_{ij}	Stratum	Average vehicles of type j per hour in stratum i (from count-period data)
TP_i	Stratum	Relative length of the considered time periods of stratum i
FSL_i	Stratum	Fixed speed limit (km/h) of stratum i (used for speed SPI)

8.2.3 Step 1: within-stratum weighting, session by session

Within a stratum, the analysis works session by session. For each observation session, two quantities are determined:

- the **SPI value** of that session (for example, 90 % of the powered two-wheeler riders observed during the session were wearing a helmet)
- the **traffic volume of the observed vehicle type** passing the location during the observation period (for example, 12 powered two-wheelers counted during 20 minutes).

The traffic volume of the different sessions is then normalised to a standard duration, denoted as $ETN_j(k)$. For example, 12 vehicles in 20 minutes becomes 36 per 60 minutes. This allows sessions of different lengths to be compared on the same scale.

The normalised traffic volume of each session is then used to construct a within-stratum weighting factor. Sessions with higher traffic volumes carry more weight than sessions with lower traffic volumes, reflecting the larger amount of exposure they represent.

Formally, the aim of the weighting step is to ensure that the overall weighted average of compliant observations reflects the overall percentage of compliant vehicles in the underlying population. Within a stratum, it is assumed that what is observed among the surveyed vehicles is representative of all vehicles that passed during the session. Each surveyed vehicle therefore represents a number of passing vehicles equal to $ETN_j(k) / N_j(k)$. Within-stratum weight assigned to one surveyed vehicle of type j within session k of stratum i , for vehicle type j is expressed as

Equation 2. Within-stratum observation weight

$$A_{ij}(k) = \begin{cases} \frac{ETN_j(k)}{N_j(k)}; & \text{when } N_j(k) > 0 \\ 0; & \text{when } N_j(k) = 0 \end{cases}$$

Index j runs over vehicle types ($j = 1$ for passenger cars, $j = 2$ for goods vehicles, and so on)

Index i runs over the basic strata (combinations of region, road type and time period)

Index k runs over the K sessions in stratum i .

There are situations in which count-period data are not collected, particularly when the entire traffic flow is captured by the observers. In these cases, the variables $n_j(k)$ and $t(k)$ are unavailable, and $ETN_j(k)$ is approximated by the total number of relevant vehicles observed during the standard observation session, i.e., $ETN_j(k) = N_j(k)$. Consequently, Equation 2 is simplified, i.e. $A_{ij}(k) = 1$, when $N_j(k) > 0$. In this case, all individual observations within the same stratum and vehicle type receive equal within-stratum session factors. The stratum-level importance is then introduced later through the combined observation weight.

The same procedure is repeated for every stratum and for every breakdown of interest (see Section 8.3.3) The same calculation can be performed for any breakdown variable provided that the resulting sub-stratum still meets the minimum sample-size criterion. In general, the SPI value will differ from one stratum to the next: helmet wearing on urban roads, for example, may differ markedly from helmet wearing on rural roads.

8.2.4 Trimming of extreme weights

Weighted averages may be sensitive to outliers in the weights themselves. For some combinations of stratum and vehicle type, particularly where the sample size is small, an extreme session factor may strongly influence the calculated SPI value. The recommended remedy is to trim the weights at the basic stratum level to the standard boxplot whiskers and leave weights between the two bounds unchanged. The lower and upper bounds are calculated separately for each stratum i and vehicle type j , using the distribution of over all sessions k :

- Lower bound: $L_{ij} = Q1_{ij} - 1.5 \cdot IQR_{ij}$
- Upper bound: $U_{ij} = Q3_{ij} + 1.5 \cdot IQR_{ij}$
- for each $A_{ij}(k)$: $A_{ij}^{trim}(k) = \max(L_{ij}; \min(U_{ij}, A_{ij}(k)))$

where $Q1$, $Q3$ and $IQR = Q3 - Q1$ are the first, third quartiles and interquartile range of $A_{ij}(k)$ over all sessions k within the same stratum i and vehicle type j .

8.2.5 Step 2: between-strata aggregation

From a policy perspective it is useful to aggregate the SPI data across strata, in order to arrive at a national indicator value that takes into account all road types, all time periods and (where applicable) all vehicle types. Obtaining such national indicators within Europe was actually the starting point of the Baseline and Trendline programmes.

When two or more strata are aggregated, the relative importance assigned to each stratum should reflect the (estimated) volume of traffic in that stratum. The ideal basis for stratum weighting is national mobility data expressed in vehicle-kilometres travelled (VKT), broken down by the relevant vehicle type and road category. For powered two-wheelers, for example, the weight given to rural roads in the national helmet-use SPI should be proportional to the rural share of total powered-two-wheeler vehicle-kilometres.

Where official VKT data are not available, the recommended fallback is to combine road-length data with traffic counts collected during the observation sessions: road length provides a stable structural denominator, while session traffic counts provide a contextual correction. If road-length data are also missing, expert estimates from the relevant national road administration should be requested. If neither vehicle-count, traffic-

volume estimates nor road-length information can be obtained, the sensible course is to keep the strata separate and to refrain from aggregation, since any aggregation under such circumstances would be more misleading than informative.

The stratum-level weight characteristic Sl_{ij} can be calculated in either of two versions, depending on what national data are available:

- Version 1 (recommended) uses the relative traffic-volume estimates TR_{ij} — see Equation 3.
- Version 2 (fallback) builds the stratum weight from road length, average hourly vehicle count and the relative length of the time period — see Equation 4.

Equation 3. Stratum-level weight. Version 1 (recommended)

$$Sl_{ij} = \frac{TR_{ij}}{\sum_{i=1}^M TR_{ij}}$$

The summation index i runs over all M strata.

Equation 4. Stratum-level weight. Version 2 (when no vehicle kilometre data are available)

$$Sl_{ij} = \frac{Rl_i \cdot NS_{ij} \cdot TP_i}{\sum_{i=1}^M (Rl_i \cdot NS_{ij} \cdot TP_i)}$$

The summation index i runs over all M strata.

Where national vehicle-kilometre data are available (by vehicle type, road type and (where applicable) time period) Version 1 is the natural choice. Version 2 is more vulnerable to bias than Version 1, because it relies on session-derived hourly vehicle counts as proxies for stratum-level traffic volume. Version 2's validity is highest when the locations sampled in the stratum are themselves randomly selected (including low-volume locations) so that the average hourly count NS_{ij} genuinely approximates the stratum mean, and when sessions are scheduled at mixed times during the considered time-period strata so that peak and off-peak hours are both represented.

The final observation weight combines the within-session factor with the stratum-level weight characteristic at the vehicle type level (see Equation 5)

Equation 5. Combined observation weight for final observation

$$W_{ij}(k) = \begin{cases} \frac{A_{ij}^{trim}(k)}{\sum_{k=1}^K N_j(k) \cdot A_{ij}^{trim}(k)} \cdot Sl_{ij}; & \text{when } N_j(k) > 0 \\ 0; & \text{when } N_j(k) = 0 \end{cases}$$

The denominator normalises the session-level factors within each stratum and vehicle type. This ensures that the total weight of all individual observations within stratum and vehicle type equals the stratum-level weight Sl_{ij} . The weight $W_{ij}(k)$ is assigned to each individual survey observations x belonging to vehicle type j , session k , and stratum i as W_x , where x runs from 1 to n and n is the total number of observations in the dataset. In other words, all individual observations of vehicle type within the same session receive the same combined observation weight.

If a stratum contains no valid observations for a given vehicle type, no observation weight can be assigned to that stratum. Such strata should be treated as missing and excluded from the valid reporting domain, unless they can be merged with a methodologically appropriate neighbouring stratum. The stratum-level weights Sl_{ij} are then renormalised over the remaining valid strata before calculating the combined observation weights. This ensures that the total weight of all individual observations sums to 1 and that each valid stratum contributes according to its adjusted stratum-level importance.

8.2.6 Dealing with breakdowns: subpopulations within a stratum

Several SPIs strongly benefit from a distinction between vehicle types or road-user groups within the same stratum. The SPI on distraction, for instance, requires a minimum of three vehicle types; the SPI on seat belts in goods vehicles, where it is reported, requires the goods-vehicle subset to be tracked separately from passenger cars. The principle in such cases is that the logic developed in the previous sections is applied to each considered subpopulation in its own right, with separate values of n_j , N_j and ETN_j recorded for each value of the subpopulation index j .

This requires, however, that the different subpopulations can also be counted during the traffic count. If they cannot, because for example counters at a particular site cannot reliably distinguish between vans and small lorries at speed, the practical workaround is to assume that the distribution of subpopulations in the passing traffic stream is the same as in the observed and surveyed sample. This assumption is generally acceptable because the standard fieldwork rule is to observe the next arriving vehicle after coding the current one. In other words, using simple random sampling among the passing flow with no deliberate over-sampling or under-sampling of any specific vehicle type. The estimated value $ETN_j(k)$ is then adapted accordingly using the observed proportion of the subpopulation.

Variables such as age category and gender are treated differently. They are not stratification variables in the sampling sense because locations and sessions are not chosen so as to ensure a particular gender mix. Rather they are variables of the surveyed subpopulation. If, for instance, separate SPI values are needed for male and female drivers, the relative number of female drivers in the set of observed vehicles is treated as representative of the female share in all vehicles passing by. In this respect, roadside SPI surveys differ from questionnaire surveys, where age and gender are themselves sampling strata and where weighting against population statistics for gender and age is appropriate.

8.2.7 Small strata, missing strata and reporting limits

In some strata, the number of observations may be small, and this can lead to unstable estimates. Possible strategies are: combining strata where methodologically justified; reporting wider confidence intervals; and clearly flagging estimates with low precision. Minimum sample-size requirements in each SPI-specific chapter aim to reduce this problem.

Calculating SPI values for the crossed strata of road type by time period is not in itself a minimum requirement of Trendline, but it is recommended where these categories were part of the sampling strategy. As noted earlier, it is recommended to not report a separate SPI value for strata with fewer than 500 observations, because both the number of distinct observations and the number of distinct locations are too small for reliable estimation. In such cases, adjacent strata can be combined, for example, 'weekday daytime' and 'weekday nighttime' merging into 'weekday'.

Weighting schemes that amplify the influence of very small strata should also be avoided. Where strata have fewer than 10 locations or fewer than 500 observations, the recommendation is to merge them with adjacent strata before calculating SPI values. The full sampling design, including stratification, weighting approach, and any deviations from the planned protocol, should be documented in writing.

In roadside measurements, non-response at the individual stratum is generally negligible because observations are collected directly by trained observers. Nevertheless, due to operational limitations or the use of automated data collection systems (e.g., cameras, inductive loops, radar guns), incomplete observation sessions or missing strata may occasionally occur due to operational constraints. In such situations, the use of non-response correction factors or weight adjustments may be considered to compensate for the missing observations. However, these methods are not developed in the present manual. In general, it is recommended to follow the rule that if the sample size in a given subset does not reach at least 30 observations (Parker et al., 2017), the resulting values are treated as not available (NA).

8.3 Estimation of SPI values

8.3.1 The SPI value as a weighted Bernoulli proportion

The dependent variable in a behavioural SPI study is binary at the level of each individual observation: a given vehicle or road user either complies with the rule ($Compliant_x = \text{TRUE}$) or does not ($Compliant_x = \text{FALSE}$). Statistically, this is a Bernoulli variable, and the expected value of positive counts, i.e. the SPI value is calculated as a weighted average over all n observations in the dataset, using the weights W_x produced by the procedures described in Section 8.2 — see Equation 6.

Equation 6. SPI value as a weighted proportion (general form)

$$\hat{p} = \frac{\sum_{x=1}^n (W_x \cdot I(Compliant_x))}{\sum_{x=1}^n W_x}$$

where: $I(Compliant_x) = 1$ if observation x is compliant, 0 otherwise

W_x = weight assigned to observation x by Section 8.2

x = 1, ..., n (n = total observations in the dataset)

The same formula is applied at every level of stratification: it gives the overall national SPI when applied to the full dataset, and it gives the stratum-specific SPI when applied to a subset corresponding to a single stratum. For multi-category SPIs, weighted category proportions are calculated by the same formula applied per category.

8.3.2 Special case: when all passing vehicles can be observed

Some observation methods make it possible to observe and effectively count every passing vehicle. This is typical of speed measurements made with induction loops or radar, and it is also realistic for observations of helmet use by motorcyclists when motorcycle traffic is light enough that no vehicle passes unobserved. In such cases, the formulation of the weighting and confidence-interval calculation may in principle be simplified, because there is only one stage of sampling: the stratified sampling of observation locations. In this situation, the within-session sampling component is absent and the estimate may be interpreted as a descriptive estimate for the observed traffic flow.

The simplification, however, comes with a caveat. The user must be entirely sure that all relevant traffic has indeed been counted; under-counting at high-flow moments would re-introduce a within-session sampling element without the methodology being aware of it. And even where the simplification is genuinely available, **it remains strongly preferable to use a complex-survey package** rather than to implement closed-form formulas by hand: the gain in transparency and the avoidance of typing or programming errors in equations involving Beta-distribution quantiles outweigh the conceptual elegance of the simpler closed form.

8.3.3 Calculation of breakdowns

SPI values can be reported for subgroups such as road type, vehicle type, time of day, gender, age groups or any combination thereof. Breakdowns must always use the same weighting principles as national totals and must be based on sufficient sample sizes.

As noted in Section 8.2.6, breakdowns by subpopulation variables that are not themselves sampling strata (notably gender and age category) rely on the assumption that the distribution of the subpopulation in the surveyed sample is the same as in the passing traffic stream. In such cases the existing session weights can be used directly. Where this assumption is doubtful, the limitation should be documented in the metadata.

8.4 Estimation of confidence intervals

8.4.1 Principles and the precision of SPI estimates

Because behavioural SPIs are based on samples, estimates are subject to sampling uncertainty. Typical measures of uncertainty are standard errors and 95% confidence intervals. Variance estimation must account for the stratified sampling design and the unequal weights produced by the procedure in Section 8.2.

The precision of an SPI estimate depends on three factors:

- the size of the sample
- the number of observation locations
- the variation between observations.

A larger sample, drawn from more locations and showing less variability, produces a smaller confidence interval and a more precise estimate, and increases the likelihood that an observed difference between SPI values is statistically significant rather than the product of random fluctuation.

An illustration from the long-running Belgian national alcohol-observation series shows what these uncertainties look like in practice. Between 2003 and 2019, between eight and ten thousand drivers were tested in each annual round, yet the resulting confidence intervals on the percentage of drivers exceeding the legal limit (0.5 g/l) remain wide enough that most year-to-year differences were not statistically significant. On the basis of those SPI values alone, one cannot claim a significant downward trend over the period. Such results underline two practical points: first, even relatively large samples can produce uncertain estimates when the prevalence of the behaviour being measured is low; second, increasing sample sizes is not always feasible for logistic and financial reasons, so the interpretation of SPI changes must always take confidence intervals into account.

Under the simplifying assumption of simple random sampling, the 95% confidence interval for an SPI estimate depends on the sample size and on the prevalence level. The Trendline minimum of 2,000 observations produces a half-width of approximately ± 2 percentage points around an estimate near 50%, narrowing to approximately ± 0.4 percentage points near 1 % or 99 %. For a sample of 500 observations the half-widths are roughly twice as large.

For planning purposes, it is useful to have a quick reference for the 95% confidence interval that a given sample size will produce at different prevalence levels. The table below assumes simple random sampling, and so will tend to be optimistic relative to the actual confidence intervals produced by the complex-survey calculation discussed later (Section 8.4.3): actual intervals will typically be wider by a factor that depends on the design effect.

Table 8-2. 95 % confidence intervals at fixed sample sizes (simple random sampling assumption).

Prevalence	n = 2,000	n = 500	n = 200	n = 50
50%	47.8% – 52.2%	45.5% – 54.5%	42.9% – 57.1%	35.5% – 64.5%
75%	73.0% – 76.9%	71.0% – 78.7%	68.4% – 80.8%	—
90%	88.6% – 91.3%	87.0% – 92.5%	85.0% – 93.8%	78.2% – 96.7%

At a Trendline-minimum sample of $n = 2,000$, the half-width is approximately ± 2.2 percentage points around a prevalence of 50%, narrowing to ± 1.4 percentage points around 90 %. At $n = 500$ — the minimum per road type — the half-widths are roughly 4.5 and 2.5 percentage points respectively. At $n = 200$, which is below the recommended threshold but is sometimes the best that can be achieved for low-traffic strata, the half-widths are 7 and 4 percentage points. At $n = 50$, the figures become so wide that the resulting estimate is rarely useful

for policy comparison. The table makes visible something that is sometimes overlooked at the planning stage: precision improves disproportionately at the high-prevalence end. A behaviour that is observed at 90% prevalence requires a substantially smaller sample to estimate to a given precision than a behaviour at 50%. For seat belt use in countries where the headline figure is already above 90%, sample sizes that would be inadequate at 50% prevalence are perfectly serviceable. For low-prevalence behaviours such as drink-driving, where the headline figure is in the low single percentages, the same sample size yields very narrow confidence intervals at the headline level but unusable ones for many subgroup breakdowns.

8.4.2 Why simple-random-sampling formulas are insufficient: clustering and design effect

Confidence intervals such as displayed in Table 8-2 are, however, insufficient for clustered data, where observations are nested within specific sessions. The most important methodological consideration is that the confidence interval must reflect the variation between locations and sessions, not the variation between individual vehicles within a session. Table 8-2 is therefore best read as a lower bound on the achievable precision, useful for setting realistic expectations at the planning stage but not as a substitute for the proper calculation at the analysis stage.

Approximations that assume simple random sampling of individual vehicles produce confidence intervals that are unrealistically narrow, because they ignore the clustering of observations within sessions and the consequent within-session correlation. Such approximations are not acceptable for Trendline reporting, regardless of how convenient they may be. Trendline beneficiaries are required to use a confidence-interval method that takes the sampling design into account, and to indicate in the metadata which method has been used.

Two practical consequences follow. First, the Design Effect (*Deff*) must be calculated to account for the loss of efficiency due to clustering. Second, for complex indicators such as 95th-percentile speed or post-crash response time, the *BC_a* (bias-corrected accelerated) bootstrap interval is the recommended standard, since it ensures that the 95% confidence intervals remain robust even when the underlying distribution is non-normal.

8.4.3 Calculating confidence intervals — the Korn–Graubard approach

The recommended method for calculating Trendline confidence intervals is the one set out by Korn and Graubard (1998) and reaffirmed as the presentation standard for proportions by Parker et al. (2017) at the United States National Center for Health Statistics. The method constructs the confidence interval using an incomplete beta function with an effective sample size derived from the estimated variance of the proportion, and is therefore well suited to complex survey designs.

The construction proceeds in three steps. First, an effective sample size *n.eff* is defined as the value that would, under simple random sampling of vehicles, produce the same variance for $\hat{p}$ as the actual complex design produced (see Equation 7). Second, the effective sample size is adjusted for degrees of freedom by multiplying by the squared ratio of the relevant Student-t quantiles (Equation 8). Third, the lower and upper bounds of the $(1 - \alpha)$ confidence interval are obtained as quantiles of a Beta distribution whose two shape parameters are functions of *n.dof* and $\hat{p}$ (Equation 9).

Equation 7. Step 1 - effective sample size.

$$n.eff = \frac{\hat{p} \cdot (1 - \hat{p})}{\widehat{var}(\hat{p})},$$

where $\widehat{var}(\hat{p})$ is an estimator of the variance of p that respects the sampling design (in practice, the variance estimate produced by the survey package).

Equation 8. Step 2 - degrees-of-freedom adjustment:

$$n.df = n.eff \cdot \left(\frac{t_{n-1} \left(1 - \frac{\alpha}{2}\right)}{t_d \left(1 - \frac{\alpha}{2}\right)} \right)^2,$$

where $t_d(\alpha)$ is the α -quantile of Student's t with d degrees of freedom and significance level α . For large $n.eff$, the standard normal quantile may replace the t -quantile in the denominator.

Equation 9. Step 3 - Beta-quantile interval:

$$CI = \left(Q\left(\frac{\alpha}{2}, n.df \cdot \hat{p}, n.df \cdot (1 - \hat{p}) + 1\right), Q\left(1 - \frac{\alpha}{2}, n.df \cdot \hat{p}, n.df \cdot (1 - \hat{p}) + 1\right) \right),$$

where $Q(p, sh1, sh2)$ is the p -th quantile of the Beta distribution with shape parameters $sh1$ and $sh2$.

The justification for the Beta distribution is that the underlying proportion is bounded between 0 and 1; the Beta-based interval respects those bounds and remains usable even when $\hat{p}$ is very close to 0 or 1, where normal approximations are particularly unreliable.

In applications with a known sampling design, modifications of this construction may be appropriate, for instance, truncating the degrees-of-freedom-adjusted effective sample size when the estimated design effect falls below one because of variance-estimator instability. Korn and Graubard (1998) discuss these modifications in detail.

Manual implementation of the formulas in Sections 8.2 to 8.4 is error-prone, particularly the variance estimation step. The strong recommendation is therefore to use a dedicated complex-survey package. This manual recommends four such packages, all of which can accommodate the design considered here: the survey package in R, the complex-survey commands in STATA, IBM SPSS Complex Samples, and the SURVEYFREQ / SURVEYMEANS set of procedures in SAS. An overview is provided in Table 8-3.

Table 8-3. Recommended complex-survey software for SPI calculations.

Software	Key procedure	Notes
R (survey package)	svydesign() / svyciprop(method="beta") / svyby()	Open source; documentation extensive
STATA	svyset / svy: proportion / svy: tab	Standard tool in many national statistical offices
SPSS Complex Samples	CSPLAN / CSDSCRIPTIVES / CSTABULATE	Graphical interface; widely available
SAS	PROC SURVEYFREQ / PROC SURVEYMEANS	Strong support for large datasets

All four packages allow the user to specify the sampling design (stratification by road type and time period, clustering by session, observation weights), and produce point estimates, standard errors and confidence intervals consistent with the methodology above.

Appendix B includes a worked out Example of the application of the procedures, formulas, and equations in this Chapter.

8.5 Using the data templates for data reporting

Final SPI datasets should include the weighted national SPI value, its confidence interval, the sample size, and a description of the sampling and weighting methods used. These results feed into the reporting and interpretation steps covered in Chapter 4.

Trendline beneficiaries have the option of submitting raw observation data to the coordination team using a standardised set of templates in Excel. When raw data are submitted, the coordination team performs the weighting and the confidence-interval calculation centrally and returns the consolidated results, ensuring methodological consistency across countries. Beneficiaries who choose to submit aggregated results instead must implement the algorithms described in this manual in their own statistical software, and must complete the metadata sheet with particular care so that the methodological choices made are transparent. The structure and use of the data templates are further described in Appendix A. The current templates (Excel sheets with multiple tabs) are available at url: <https://trendlineproject.eu/templates>.

A typical template for a behavioural SPI has seven sheets.

- Sheet 1, 'Roadside_RawData-7level', holds the source data, with one row per observed vehicle occupant; the typical columns are Region ID (optional), Road type, Time period, Session ID, Location ID, Vehicle ID, Vehicle type, Seat position, Gender, Age group, and the binary outcome variable Used.
- Sheets 2, 'Roadside_RawData-6level_1' holds the count-period data per session: vehicle counts by vehicle type in sheet 2.
- Sheet 3, 'Roadside_RawData-6level_2' contains the location information for each session, including the counting durations $T(k)$ and $t(k)$, as well as the speed limit at the location (for speed SPI).
- Sheet 4, 'Roadside_RawData-3level', holds the relative traffic-volume estimates TR_{ij} used in Version 1 of the stratum weighting.
- Sheet 5, 'Roadside_RawData-1level', holds the structural data used in Version 2 — road length Rl_i , the time-period proportions TP_i , the average hourly vehicle counts NS_{ij} derived from the count-period data, and (for the speed SPI) the fixed speed limit FSL_i . This is the speed limit defined for each vehicle type and road type combination, e.g., 90 km/h for trucks on a motorway with 110km/h speed limit.
- Sheets 6 and 7, 'Roadside_Results-aggr' and 'Roadside_Results-semiaggr', are optional: they need only be filled in if the beneficiary has chosen to deliver aggregated results rather than raw data.

Source records can be partially anonymised through the use of opaque codes for Region ID, Session ID, Location ID and Vehicle ID; a unique code linked across the source tables is sufficient. The exact time at which individual vehicles passed need not be recorded.

In addition, there is a sheet for metadata. That includes in particular, which version of the stratum-level weight component (Version 1 or Version 2) has been used, and any SPI-specific choices — for example, the age-group convention adopted for the seat belt and child restraint systems SPI.

9 Using self-reported data for SPIs

9.1 Conceptual framework

9.1.1 Self-reported data in road safety research

This chapter sets out the methodological framework for measuring road safety performance indicators through self-reported data. It introduces the conceptual position of self-reported indicators within the Trendline framework, explains how they complement the observed indicators that dominate the rest of this Manual, and specifies the sampling, questionnaire, fieldwork, analysis and reporting requirements that should be followed in order to produce nationally representative and internationally comparable results.

Collecting and analysing data provided by interviewees or survey respondents is a method used in many research areas and for collecting information within organisations and jurisdictions. Examples of applications are attitude surveys, censuses, market research surveys, psychological assessments, reporting on health behaviour, and polls on political preferences. Within road safety research, the so-called “self-report” or “self-declared” approaches involve collecting information directly from individuals about, among other things, their experiences, characteristics, mobility, behaviour in traffic, attitudes and perceptions. These methods can be based on various survey techniques, including face-to-face, telephone or online video interviews, written surveys (handed out to groups of people or sent by post) and online surveys. Each technique offers specific advantages and disadvantages, which influence the quality and reliability of the data collected (Lajunen & Özkan, 2011).

In the Trendline framework, self-reported behaviour and attitudes are part of the pillar of behavioural SPIs. While observed SPIs measure what road users actually do at specific moments in real traffic and database-based SPIs measure structural conditions of the road safety system, self-reported SPIs measure how often road users report engaging in given behaviours, and how they think and feel about those behaviours, over a defined reference period. Self-reported indicators therefore operate at the level of operational safety culture and behavioural norms, which influence real-world behaviour but cannot be captured directly through roadside observation.

Within Trendline, use of self-reported data is explicitly recognised as a viable (and sometimes the only feasible method) for the SPIs on driving under the influence or alcohol and drugs. But SPIs based on self-reported data can also be used as a complementary SPI for speeding, seat belt use, helmet use, distraction and red-light negation. It can also be used for some risk behaviours which are not covered by Trendline at all, such as tired driving and some risky behaviours of cyclists.

9.1.2 Period prevalence and trip-based prevalence

Data on self-reported behaviour can be collected through two distinct types of survey, referred to as the period-prevalence and the trip-based prevalence survey:

- **Period-prevalence surveys** measure the prevalence of a risky behaviour over a specified period of time, typically the last 30 days. Respondents are asked how often, over that reference period, they performed a given behaviour. The period-prevalence approach is currently the most common method in road safety research. It has been used in the SARTRE series (Social Attitudes to Road Traffic Risk in Europe) and is at the heart of the ESRA series (E-Survey of Road users' Attitudes, <https://www.esranet.eu/>), as well as Traffic Safety Culture Index (TSCI) published by AAA Foundation for Traffic Safety in the US (<https://aaaafoundation.org/research/traffic-safety-culture-index/>). Its principal advantage is that the values obtained are generally higher and show more variation than in trip-based surveys, so that a much smaller sample is sufficient to produce meaningful results.

- **Trip-based prevalence surveys** measure the prevalence of a risky behaviour during a specific, randomly chosen recent trip, typically the day before the survey is taken. This method is most often used in connection with the analysis of mobility patterns, but has also been used to measure driving under the influence of alcohol and driving while drowsy (Diependaele, 2015). One of its advantages is that the values obtained are in principle comparable to those obtained through observational studies, because both methods aim at estimating a kind of point prevalence. Three types of data collection can be used for trip-based prevalence surveys: roadside interviews with traffic participants approached immediately after a trip; telephone interviews in which the respondent is asked about their last trip; and online surveys in which the respondent is asked about all trips made in the last 24 hours (or the last 7 days) and then about their behaviour during one randomly selected trip.

Whereas roadside observation measures 'point prevalence' (the proportion of vehicles or road users in violation at a particular point in time), period-prevalence surveys measure the proportion of road users who have engaged in a given behaviour at least once over a longer reference window. The trip-based approach sits in between: it asks about behaviour during a specific recent trip and so produces values that are conceptually close to a point estimate, but it does so retrospectively, with the limitations that recall introduces.

9.1.3 Link to observed behaviour

Roadside observation captures momentary, externally visible behaviour at a defined location at a defined time. It is objective in the sense that the observer records what is in front of them rather than what the road user reports. A major limitation is that it can only capture variables that are externally visible – it cannot record an attitude, an opinion or a recollection of past behaviour. Because the road users observed remain anonymous, demographic and contextual information is largely unavailable. It also requires a fairly sophisticated study design to remain representative (see Chapter 6).

Self-reported behaviour, by contrast, captures what the respondent recalls about their own past behaviour and what they believe and feel about traffic safety. It allows access to dimensions that observation cannot reach: attitudes (e.g., whether the respondent thinks speeding is dangerous), perceived social norms (e.g., whether the respondent thinks their friends would disapprove of drink-driving), risk perception (e.g., whether the respondent considers the probability of being caught driving over the alcohol limit to be high or low), and existing habits (e.g., whether the respondent says they always wear a seat belt on short urban trips). It also allows much wider population coverage than observation, since an internet panel can reach demographic groups that would never pass any specific roadside observation point.

Despite the inherent subjectivity of self-reports, period-prevalence values from large international surveys often appear to correlate significantly with observed prevalence values from roadside surveys. This has been demonstrated repeatedly within the ESRA project by comparing self-reported alcohol and mobile phone use indicators against observed values from the DRUID, Baseline and Trendline roadside studies. For example, Goldenbeld & De Craen (2013) compared an online and a face-to-face sample of car drivers to study differences on a number of selected questions from the SARTRE-4 road safety survey. Contrary to expectations, there was no indication that online respondents were more likely to come from higher educated or more privileged social groups; in addition, face-to-face respondents did not tend to give more positive answers in judgment of road safety measures. Also, Holló and colleagues (2018) provide an example of correlation between self-reported and observed data. The implication is that the two methods track the same underlying reality in a comparable way, even if the numbers produced are quite different (in general the figures of for the level of non-compliance are much higher in self-reported period-prevalence data than in point-prevalence observed data). A country can therefore use self-reported SPIs as a useful proxy where observation is unavailable or as a complementary signal where both are available, without expecting numerical agreement between the two.

9.1.4 Measurement of behaviour and attitudes

The Trendline SPI framework is built on the policy logic established by the European Commission, in which actions to improve road safety produce intermediate outcomes that in turn affect the impact in terms of crashes, injuries and fatalities. Most of the original eight Baseline SPIs are anchored in objective measurements or roadside observations, which is the methodological gold standard for assessing compliance with traffic rules but is also relatively expensive. Self-report surveys, by contrast, allow a fast and relatively inexpensive assessment of self-reported traffic behaviour. They make a different kind of cross-country comparison possible, namely the prevalence of risky behaviour among the population over a period of weeks, months or years. Two large internationally comparative survey series illustrate the approach: the SARTRE series, with four editions between 1991 and 2010, and the ESRA series, with three editions between 2015 and 2023 (and the fourth edition starting in 2026), the most recent of which sampled 26 European countries.

In addition to the prevalence of self-reported behaviours, self-report surveys also typically measure attitudes towards road safety. The term 'attitudes' is used in the Trendline framework as an umbrella for several socio-cognitive concepts that include attitudes in the strict psychological sense, perceived behavioural control or self-efficacy, social norms (both perceived social acceptability and personal acceptability), subjective risk perception, habits and behavioural intentions.

The conceptual relationship between these constructs and behaviour is itself an active area of research. The classical theory of planned behaviour treats attitudes, subjective norms and perceived behavioural control as causes of intentions, which in turn cause behaviour. Studies (e.g., Kroesen et al., 2017) have shown that the causal chain may also run in the reverse direction, with intentions and behaviour shaping the attitudes that follow them, in line with self-perception theory. In the field of mobility, dissonant travellers have been shown to adjust their attitudes to align with their behaviour rather than the other way round.

These theoretical positions matter because they shape the policy interpretation of attitudinal indicators. Where attitudes primarily influence behaviour, attitudinal change campaigns and information interventions are the natural levers; where behaviour primarily influences attitudes, the priority shifts to changing behaviour directly through regulation, enforcement or pricing, with attitudinal change following. The manual takes the pragmatic position that attitudes can be regarded as performance indicators to the extent that they are correlated with crash occurrence, while remaining agnostic about the direction of causation in any given case.

9.1.5 Limitations and biases

Self-reported data are subject to several well-documented sources of bias and response tendencies, which need to be acknowledged when results are interpreted.

- **Social desirability bias.** Respondents tend to provide answers that present a favourable image of themselves, deflating the reported share of socially undesirable behaviours (drink-driving, speeding, mobile phone use) and inflating the reported share of socially approved behaviours (seat belt and helmet wearing). This bias is stronger in interviewer-administered modes than in self-administered online surveys.
- **Recall bias.** Respondents may unintentionally provide inaccurate answers due to memory error. The longer the period between the behaviour and the survey, the higher the risk of recall bias.
- **Unconscious behaviour bias.** When driving, road users may be unaware of certain behaviours, for example unintended speeding, momentary distraction or early drowsiness. Behaviour that the respondent was not aware of at the moment itself cannot be reported afterwards. This bias can be at least as important as recall bias.
- **Reduced representativeness.** For surveys based on non-probability sampling, the representativeness of the sample is limited. Even with probability sampling, non-response of certain groups (those without

internet access, those uninterested in participating, those with privacy concerns) introduces bias if the non-responders differ systematically from responders.

- **Careless responses.** The anonymity of questionnaire surveys can encourage inattentive responding, particularly when participants are motivated by incentives rather than by the survey topic. Careless respondents fail to read or attend to item content, producing data that do not reflect their true level on the construct being measured.
- **Cultural and international differences.** In cross-national studies, differences in social values, traffic culture, infrastructure, traffic composition, enforcement intensity and culture-related response tendencies (for example, the meaning of intermediate scale values) can influence answers in ways that complicate the interpretation of differences between countries.

Despite these biases, self-reported data have repeatedly been shown to track observed values closely in international comparisons. This manual takes a pragmatic position: self-reported and attitudinal indicators are accepted as performance indicators to the extent that they correlate with crash occurrence, while remaining agnostic about the direction of causation in any given case. They are presented as complementary indicators that enrich the observed and database-based sets rather than replacing them.

9.1.6 Ethical considerations

Ethical considerations in self-reported research are core conditions of credible data collection. Participants must be properly informed about the purpose of the study, their role in it, and their rights, including the right to withdraw without penalty. Written informed consent should be obtained from every participant, covering the purpose of the study, the handling of recordings or transcripts where these are used, and the intended use of the data. Data privacy is crucial, especially where sensitive information is involved (e.g., about alcohol use, drug use or risky driving). Within the European Union, all data collection must comply with the General Data Protection Regulation (GDPR). Researchers must adhere to protocols that anonymise data and secure both storage and transmission to protect participant confidentiality.

Where online panels provided by market research agencies are used, the primary responsibility for participant privacy lies with the agency. When selecting an agency for a survey, researchers should verify that the agency operates under documented privacy procedures and complies with the relevant national and European legislation.

9.2 Survey methodology

9.2.1 Representative online panels as the preferred survey mode

Self-report data can be collected by self-administration on paper or online, by face-to-face or telephone interview, or by combinations of these. Because survey results are sensitive to even minor methodological differences, standardisation is essential. Self-reported SPIs must be measured using nationally representative anonymous surveys, ideally following the ESRA methodology to ensure cross-national comparability. Using a common design, questionnaire structure and analysis protocol is the most effective way to minimise methodological variance between countries.

This manual treats representative (panel-based) anonymous online surveys as the recommended method, on the grounds that online surveys are cost-efficient, allow large samples, preserve anonymity, eliminate the interviewer effect and reduce socially desirable answers. Alternative modes (telephone, face-to-face, roadside) can be considered but must be clearly documented in the metadata, since mode effects can substantially influence responses. This does not mean that such values are inferior; it means simply that the user of the data needs to be aware of the methodological difference and the limited comparability with results of surveys undertaken in other countries.

An online-access panel is a pre-recruited group of individuals who have agreed to participate in market research, surveys, and feedback sessions. In addition to the general advantages of questionnaire surveys, online access panels offer specific operational benefits:

- **Greater reach.** Online panels provide access to a wide range of demographics, geographic locations (even across countries) and specialised groups (for example specific driver categories).
- **Cost-effectiveness.** Compared with face-to-face interviews or postal surveys, online panel surveys are generally less expensive, with almost no costs for physical materials, travel or personnel time. This makes it feasible to collect data from a large number of participants across diverse geographic locations.
- **Speed and efficiency.** Online panels allow rapid data collection, often within a week or so, because surveys can be distributed to thousands of participants simultaneously. A well-designed online survey with built-in error correction also reduces the need for data cleaning.
- **International comparisons.** Online panels allow systematic and uniform data collection across countries: the survey can be programmed once and subsequently translated. Comparability across countries still depends on the quality of translation and on cultural and structural differences between countries.

9.2.2 Specific limitations of online panels

In addition to the generic limitations of questionnaire surveys discussed earlier, the use of online panels comes with some additional concerns:

- **Composition of the panel.** Despite the efforts of panel providers, the distribution of panel members is never fully representative of the total population. Variations in internet penetration and use, lack of interest in panel membership among certain demographic groups, and economic incentives for joining a panel all introduce systematic distortions.
- **Panel conditioning.** Regular participation in surveys may lead respondents to become overly familiar with survey techniques, which can bias their responses.
- **Careless respondents.** Some panel members are mainly motivated by collecting incentives and tend to complete surveys as fast as possible. Experience from ESRA suggests that careless responding is increasing in the panels of commercial market research agencies, and dedicated quality control items in the questionnaire are needed to identify and remove such respondents.
- **Low panel size in some countries.** In small countries and in low- and middle-income countries, available panels may be too small for large-scale sampling, leading to higher risks of non-representativeness, panel conditioning and higher unit costs.
- **Manipulation.** Two types of manipulation can occur: respondents answering several times to maximise incentives, and polling agencies adding duplicate entries to reach a target sample. Data should always be checked for duplicates.

9.3 Sampling design for surveys

9.3.1 Target population and sampling method

The typical theoretical population for self-reported behaviour and attitude indicators is the national general population aged 18 and over, with relevant subpopulations (such as car drivers, motorcyclists, cyclists, pedestrians) defined for each indicator. For certain types of surveys, in particular those focusing on vulnerable road users (cyclists, pedestrians, moped riders, PMD riders) it can be considered to lower the age threshold to 16 years or even lower. On the other hand, it is often needed to put an age limit (e.g. 75) on the population to be studied, because of the low and unrepresentative prevalence of 75+ year olds in online panels.

Random sampling is required; random sampling from available online research panels is allowed. Convenience samples, open invitations on social media and recruitment limited to specific target groups are not acceptable, since they are too biased to generate a representative estimate. Sampling strategies should ideally be in line with the ESRA method, which uses interlaced (crossed) quota for gender (male, female) by six age groups (18–24, 25–34, 35–44, 45–54, 55–64, 65–74), together with a non-interlaced regional quota based on national population statistics (proportionate stratified sampling). In ESRA, any disproportionality is corrected through post-stratification weighting so that the sample is representative for the national population 18–74 in terms of age, gender and region.

For countries not (or not yet) participating in ESRA, the general recommendation is to conduct an online survey using the ESRA sampling, question formulation and analysis methods. When using other modes (telephone, face-to-face interviews, ...) the recommendation is to sample as randomly as possible and to describe the sampling procedure explicitly in the report. Quota weights may be needed where random sampling has produced over- or under-representation of specific socio-demographic groups.

In order to monitor change, it is in general not needed to repeat a survey every year. Depending on budget and expected rate of change, a repeat interval of between 3 and 5 years should be considered.

9.3.2 Minimum sample size

Sample size requirements depend on the indicator and the type of survey. For the Trendline SPI on self-reported drink driving, the minimum total sample for period-prevalence surveys is 750 car drivers, with a recommended target of 1,000. When using trip-based surveys, the minimum number of car drivers in the sample should be 2,500.

For countries that do not follow the ESRA method in all respects, the recommended minimum is 1,000 randomly sampled respondents from the general adult population. The figure of 750 is in line with the ESRA approach, in which a country sample of around 1,000 road users typically contains between 61.9% and 90% of car drivers, with a European average of 76.8%. This percentage reflects car driving at least a few days per month and is in line with the reference period used in the questions on self-reported behaviour.

This sample size of 750 for period-based surveys is sufficient to keep confidence intervals for population and subpopulation estimates within acceptable bounds. Larger samples are needed for fine-grained subgroup analyses; in particular, for any SPI that is intended to be disaggregated by week period, road type or vehicle type, the minimum sample requirements need to be considered for each stratum separately, and over-sampling of low-frequency strata is recommended. For each indicator, subsamples of fewer than 250 respondents should be flagged in the data delivery with the marker '<250'. Such estimates can still be reported, but the wider confidence intervals around them mean they should be interpreted with care.

9.4 Question phrasing and indicator content

9.4.1 Practical principles for survey question design

Where designing a survey in view of obtaining SPIs, a number of practical principles apply:

- Questions should be **short** and should be easy to understand.
- They should about **one issue** at a time; double-barrelled items such as 'Do you think speed limits are enforced and fair?' should be split into two separate questions.
- The wording should be **concrete**, avoiding legal or technical jargon unless it is defined.
- **No negative questions** should be included (constructions such as 'do not fail to ...' are particularly problematic).
- The phrasing of questions must remain **neutral**: words that imply a correct answer or a value judgement ('irresponsible', 'dangerous', 'should') are avoided.

- For sensitive behaviours such as drink driving, a normalising **preamble** that acknowledges the behaviour occurs without endorsing it can reduce social desirability bias.

Response scales should be matched to the construct: frequency scales for behaviour questions, agreement scales for attitudes, likelihood scales for risk perception. Four to seven scale points are sufficient; more than seven adds no precision and confuses respondents. A middle option ('sometimes', 'neutral') is included only when genuine ambivalence is meaningful, and 'Don't know' or 'Not applicable' options are offered only where they are real possibilities. All scale categories should be labelled, not only the endpoints, because fully labelled scales produce more reliable responses. Scale direction should be kept consistent throughout the questionnaire.

Self-reported behaviour items always specify a reference time period – 'in the last 30 days', 'in the last 12 months', 'on your last trip' – because questions without a recall window cannot be interpreted reliably. Frequency ('how often do you use your phone while driving?') is distinguished from intensity ('on a typical occasion, for how long?'). Sensitive items are placed in the middle of the questionnaire, after rapport has been established but before end-fatigue sets in, never at the very start. Trip-based anchors ('on your last trip as a driver ...') are used where episode-specific recall is more reliable than a general frequency estimate.

Questionnaire length should be matched to the mode: online surveys should not exceed 20 minutes, telephone surveys 15 minutes, and face-to-face roadside interviews should be kept under 10 minutes. Open-ended questions are to be minimised or even completely left out in online surveys, because they raise the drop-out rate and are difficult to analyse. Matrix questions should be limited to six to eight items per matrix to avoid straight-lining (the tendency to select the same response for every row without reading).

All questions should be piloted with a small test group before the main fieldwork. If a survey is made available in several languages, each of these language versions should be piloted. For international surveys, translation should follow a forward-backward procedure to ensure that the meaning is preserved across languages. Quality control can also be built into the questionnaire itself in different ways:

- inclusion of at least one instructed-response item to identify inattentive respondents ('for quality control purposes, please select option 3')
- inclusion of a consistency-check pair to identify inconsistent answering
- for online surveys, response-time thresholds to flag implausibly fast completions.

9.4.2 Behavioural questions in ESRA

Even subtle differences in question wording, item order, questionnaire length and the inclusion or omission of neighbouring items can affect responses in a self-report survey. When designing surveys in view of self-reported SPIs and when seeking international comparability, it is strongly recommended to use questions that are identical to those from the ESRA survey for both behavioural and attitudinal indicators. This minimises methodological variance between countries and ensures that Trendline indicators can be compared with the broader ESRA results.

The behavioural questions in ESRA₃ are organised by road user role and use a five-point frequency scale running from 'never' to '(almost) always', applied to a 30-day reference period. The introductory wording is:

"Over the last 30 days, how often did you as a [car driver / car passenger / moped rider or motorcyclist / cyclist / pedestrian / e-scooter rider] ...? You can indicate your answer on a scale from 1 to 5, where 1 is 'never' and 5 is '(almost) always'."

For car drivers, the items cover drink driving (driving when possibly over the legal limit and driving after drinking alcohol), drug and medication driving (within one hour of taking drugs other than prescribed or over-the-counter medication, and within two hours of taking medication that may affect driving ability), speeding inside built-up areas, outside built-up areas and on motorways, seat belt non-use, child restraint non-use for children

under 150 cm and seat belt non-use for children above 150 cm, and three forms of mobile phone use (handheld conversation, handsfree conversation, reading messages or social media).

Equivalent batteries of items are defined for car passengers (seat belt non-use in front and rear seats), for moped riders and motorcyclists (drink riding, speeding outside built-up areas, helmet non-use, reading messages while riding, riding within one hour of taking drugs), for cyclists (drink cycling, helmet non-use, listening to music through headphones, reading messages while cycling, drug cycling, crossing on red), for pedestrians (listening to music while walking, walking after drinking, reading and texting while walking, crossing on red, jaywalking) and for e-scooter riders (riding with more than one person, drink riding, crossing on red, riding on the pedestrian pavement, helmet non-use).

Indicators are typically derived as the percentage of respondents in the relevant subpopulation who report 'never' having engaged in the risky behaviour over the reference period. In principle, it is also possible to use a less strict approach (e.g. combining two answer categories) but within ESRA this is seldom practiced.

9.4.3 Attitudinal questions in ESRA

There are different socio-cognitive attitudinal concepts that can be investigated through surveys. The concepts measured in ESRA are:

- **Attitudes in the strict sense**, for example the degree of agreement with the statement that for short trips one can risk driving under the influence of alcohol.
- **Perceived behavioural control or self-efficacy**, for example trusting oneself to drive after one glass of alcohol.
- **Perceived social acceptability**, the extent to which the respondents believe that most other people in their environment would consider a given behaviour acceptable for car drivers.
- **Personal acceptability**, the extent to which the respondent personally considers a given behaviour acceptable.
- **Habits**, for example agreement with the statement that the respondent often drives faster than the speed limit.
- **Behavioural intentions**, for example agreement with the statement that the respondent will do their best not to drive after drinking alcohol in the next 30 days.

In ESRA, such attitudinal items are typically presented with an agreement scale of five points. The complete ESRA questionnaire is available from the ESRA methodological reports – at the moment of writing this manual the latest version is ESRA₃ (Meesmann and Wardenier, 2024) – and forms the recommended reference for Trendline self-reported indicators.

9.4.4 Compensating for bias through question design

Recall bias can be reduced by keeping the reference period as recent as possible (the last 30 days rather than a particular calendar month) and, in trip-based designs, by conducting the interview as soon as possible after the trip. Roadside interviews can be carried out at parking areas or where vehicles are stopped; pedestrians and cyclists can be approached at pedestrian crossings or in pedestrian zones. In online trip-based surveys and telephone interviews, respondents are explicitly asked to recall their last trip as vividly as possible.

Social desirability bias can be reduced by guaranteeing anonymity. Respondents are reassured in advance that all answers are confidential, and no personal data that could link directly to the individual (name, address) are collected. In interviewer-administered modes, interviewers should be trained to avoid any negative comment on reported behaviour and to encourage respondents to recall whatever they can. In online surveys, predefined wording can encourage honest answering, and a standard social desirability scale can be included so that any residual effect can be controlled for in subsequent analyses, as is done in ESRA.

9.5 Data analysis

9.5.1 Data cleaning

Analysing data from a questionnaire survey is a multi-step process that starts with thorough data cleaning. The depth of cleaning depends on the research objective and the chosen study design, but the typical steps are well established:

- **Removing duplicates**, so that each respondent's data appears only once. This is a recurrent concern with online panels, where some respondents may attempt to complete the survey more than once for incentives.
- **Handling missing data**, deciding how to deal with incomplete responses. Options include imputation, mean or mode substitution, or excluding the missing data point or the whole record.
- **Filtering outliers**, identifying data points that fall outside the expected range and deciding case by case how to handle them.
- **Correcting errors**, addressing inconsistencies or obvious data entry errors.
- **Removing careless respondents**, using instructed-response items, checks for straight-lining on matrix questions and response-time thresholds.

For data collected within the ESRA3 framework, the cleaning procedure documented in the ESRA3 methodological report should be followed. This procedure works systematically through speeders, straight-liners, missing-value patterns and other quality checks. For offline survey modes, the cleaning steps should be adapted to the interview technique, but the underlying logic remains the same. Data preparation typically includes checking the dataset received from the market research agency against the survey codebook, merging datasets where multiple agencies are involved, checking programming consistency by comparing the predefined questionnaire filters with the expected number of missing values for filtered variables, and checking the requested quotas at country level with a tolerance of 5%.

9.5.2 Weighting

It is recommended that the representativeness of the final cleaned survey sample be assessed against age, gender and regional population parameters. Any disproportionalities should be corrected through post-stratification weighting before indicator values are computed. The use of the same data source for weighting in different countries is important for international comparability.

Weighting in ESRA is based on the so-called 'Individual Country Weight', a weighting factor derived from the gender by six-age-group distribution in each country, taken from United Nations population statistics. For Trendline, equivalent population data from Eurostat can also be used, which has the advantage of providing detailed information on regional breakdowns.

9.5.3 Calculation of indicators

Behavioural SPIs are typically calculated as the percentage of respondents in the relevant subpopulation who report having engaged in the behaviour at least once over the reference period,

The main self-reported behavioural indicators are expressed as the percentage of particular road users (car drivers, motorcyclists, ...) who report never having engaged in the relevant behaviour over the reference period, or equivalently, one minus the percentage reporting 'never'. Disaggregation by age group, by gender and by vehicle category or road user type is recommended where the sample allows; further disaggregation by week period, by road type, by type of driver (professional or private), by novice driver status (less than two years of licence-holding) and by alternative reference periods is optional. Crossed point estimates for combinations of disaggregation levels are also encouraged, since these are particularly informative for targeted policy.

For attitudinal indicators, results can be reported both at item level and at the level of dimensions or scales constructed by combining items. Attitudinal SPIs can be reported as the percentage of respondents agreeing with a given statement, as the mean score on a Likert-type scale, or as composite indices constructed by combining items into dimensions. The construction method must be documented. Methods that depart from the ESRA mode (telephone, face-to-face, paper) can be used, but the resulting indicator values are flagged with an asterisk and accompanied by a brief discussion of the possible methodological impact, since survey mode is known to influence results in ways that can be substantial

For aggregated questionnaire-based indicators expressed as proportions, confidence intervals may be calculated from the aggregated numerator and denominator using the exact binomial Clopper–Pearson method, provided that the data can be treated as unweighted binomial counts. If x is the number of respondents in the category of interest and n is the total number of respondents, the 95% confidence interval is calculated from beta distribution quantiles.

Equation 10. Binomial Clopper-Pearson interval:

$$CI = (B^{-1}(0.025; x, n - x + 1), B^{-1}(0.975; x + 1, n - x))$$

where B^{-1} denotes the inverse beta distribution function. If the data are weighted or come from a complex survey design, the confidence interval should instead account for the survey design or use an appropriate effective sample-size approach. The method used for confidence-interval calculation should be documented in the metadata.

The Trendline self-reported guidelines specify a delivery format that mirrors the three-tier logic adopted for other Trendline indicators. For each indicator, a point estimate and a 95% confidence interval are reported. Subsamples of fewer than 250 respondents are flagged with the marker '<250', so that users of the data can immediately see where wider uncertainty applies.

9.5.4 Metadata

Each Member State should provide the metadata of its data collection together with the analysis results. The metadata should include:

- The type of study, including the survey method (online, telephone, face-to-face, paper, mixed) and whether the data were collected within the scope of ESRA3 (or a later wave of ESRA)
- The sampling stratifications used and the rationale for them.
- The sampling procedure, including the sampling frame, sampling quotas (strata), recruitment channels and any inclusion or exclusion criteria.
- The vehicle types and operational definitions considered.
- A description of the fieldwork procedure, including the exact wording of questions and any filters applied.
- The total respondent sample and breakdown by main strata.
- A description of the weighting procedure and the representativeness of the final sample.
- The population data source used as a weighting base (UN, Eurostat, national statistical office).
- The statistical techniques used to weight the data and to calculate confidence intervals.
- The period(s) of data collection and any contextual circumstances that may have affected responses.

Where the methodology departs from ESRA3 in any respect, the deviation must be described and its possible effect on the indicator value briefly discussed. If the data are collected within the framework of the ESRA project, a summary of the main elements of the ESRA methodology is sufficient. If not, enough detail on the survey methodology should be provided to allow the survey to be replicated and the results reproduced.

10 SPIs for speeding

10.1 Introduction

Speeding is one of the most frequently cited contributory factors in road crashes and is closely related both to the probability of a crash occurring and to its severity. Expert reviews (OECD/ITF, 2018) broadly agree that an estimated 10 to 15% of all road crashes, and around 30% of fatal injury crashes, are the direct result of excessive or inappropriate speed; in the many further cases where speed is not the main cause, it frequently acts as a contributing or aggravating factor. The Speed SPI therefore captures a behaviour that is directly linked to the kinetic energy released in a collision, and to the survival chances of the road users involved.

The mechanisms behind this relationship are well documented. A driver travelling at high speed covers a greater distance before reacting to an unexpected event, even though the reaction time itself is unchanged; the braking distance increases with the square of the speed; other road users are left with less time to avoid a collision and tend to overestimate the time available; and the driver's field of vision narrows as speed increases. The statistical relationship between speed and both crash frequency and crash severity has been confirmed by multiple studies and is increasingly expressed through the exponential models of Elvik et al. (2019). Given that the relationship is non-linear, small reductions in operating speed can yield disproportionately large reductions in casualties, which is precisely why monitoring the level and dispersion of speeds is of such policy value.

Two forms of speeding are conventionally distinguished. **Excessive speed** means driving faster than the maximum allowed, while **inappropriate speed** means driving too fast for the prevailing traffic, infrastructure or weather conditions, even where the legal limit is respected. Inappropriate speed is to some degree subjective and cannot be observed directly at the roadside; the speeding SPIs therefore concentrates on excessive speed, that is, compliance with the posted limit, while recognising that within a Safe System the reference point of interest is ultimately the *safe* speed for the road, which may or may not coincide with the posted limit.

10.2 Indicator definition and theoretical population

10.2.1 The principal indicator

The principal indicator is defined as:

Percentage of vehicles travelling within the speed limit
on road sections that allow free-flowing traffic.

This indicator was also specified in the initial Commission document on performance indicators (European Commission, 2020). The SPI measures excessive speed, which is directly linked to the number of crashes and fatal injury crashes, and every participating country is required to deliver it. The SPI should be calculated and reported separately for the three mandatory road types (motorways, rural non-motorway roads and urban roads) and further broken down to roads with particular speed limits. Since there are often several speed limits per road type in a country, for each road type, one should either select locations with the most common speed limit (a single speed limit) or, alternatively, present the results separately by speed limit.

The theoretical population for this indicator is the totality of vehicles travelling on the national road network during the reference period, restricted to road sections that genuinely allow a free choice of speed. Where good national data on traffic volumes (vehicle-kilometres travelled) by road type and vehicle type are available, the indicator can equivalently be expressed as the *percentage of kilometres driven* within the speed limit, which gives a more exposure-weighted picture of compliance.

10.2.2 Complementary speed indicators

Because the headline percentage offers only a partial picture of the speeding problem, it is complemented by some additional indicators which are also required to be delivered by Trendline participants. These are:

- the **average speed**,
- the associated **standard deviation and standard error**,
- the **V85** (85th-percentile speed): the speed below which 85% of drivers are found to be travelling.

Together, these statistics describe both the level and the dispersion of operating speeds, and allow meaningful comparison between countries. The average speed and the V85 are indicators of absolute speed and therefore act as proxies for crash rate and crash severity; the standard deviation captures speed dispersion, which is itself associated with crash risk.

Where national resources allow, the cluster is further enriched by the following SPIs that were called **Alternative Speeding Indicators (ASIs)** in the first phase of Trendline:

- The **percentage of vehicles travelling more than 10 km/h, more than 20 km/h or more than 30 km/h above the speed limit**. The proposed intervals of 10, 20 and 30 km/h above the limit may or may not match the thresholds used in national legislation; for the sake of international comparison it is a minimum requirement to calculate the ASI using these common intervals, but each Member State may additionally calculate the indicator using whatever intervals are more meaningful under its own law. Speeding more than 20km/h can be considered as the baseline complementary ASI.
- The speed dispersion expressed as **the difference between the 90th and the 10th speed percentile**, computed per road type. The 10th-percentile speed is the speed below which 10% of measured speeds fall and the 90th-percentile speed the speed below which 90% fall; their difference characterises the spread of operating speeds. These complementary measures are particularly useful for monitoring the most dangerous portion of the speed distribution, since the relationship between excess speed and crash risk is strongly non-linear.

In most countries there are two or more frequently used speed limits per road type. For example: 30 km/h and 50 km/h on urban roads; 70 km/h and 90 km/h on rural roads; and 100 km/h and 120 km/h on motorways. Measurements and analyses should be stratified alongside these frequently used speed limits. This implies that in most countries, many speeding SPI values can be calculated. For example, when considering 2 vehicle types, 6 speed limits and 4 SPI definitions, there will be $2 \times 6 \times 4 = 48$ different SPI values. For urban roads, it is recommended to report the indicator value separately for at least roads with 30 km/h and roads with 50 km/h speed limits.

10.2.3 Summary of the speeding indicator cluster

Table 9.1 summarises the Trendline speeding indicator cluster. A word of caution applies to all members of the cluster. Differences between countries may arise that are not visible from the indicator definition itself but are embedded in the metadata and the practical implementation. Examples are the inclusion or exclusion of particular vehicle types (vans, taxis, buses), particular road types (urban thoroughfares, gravel roads), non-free-flow observations, or particular time periods. Such implicit differences can materially affect comparability, and the consistent documentation of these choices in the metadata is therefore important.

Table 10-1. The Trendline speeding indicator cluster.

Indicator	Definition
% within speed limit	Share of free-flow vehicles travelling at or below the posted limit

Average speed (+ SD, SE)	Mean operating speed of free-flow vehicles, with its standard deviation and standard error
V85	Speed below which 85% of free-flow vehicles travel (85th percentile)
% exceeding by >10 / >20 / >30 km/h	Share of vehicles travelling more than 10, 20 or 30 km/h above the limit
Speed dispersion (V90–V10)	Difference between the 90 th and 10 th percentile speeds, by road type / area / limit / vehicle type

10.3 Position within the behavioural framework

As set out in the introduction to behavioural SPIs, the speeding indicators rest on the common backbone and are not re-derived here. Two divergences shape this chapter and are developed below.

One divergence is fundamental: unlike the other behavioural SPIs, most speeding SPIs are not based on visual observation by human observers but on instrument-based measurement (see Table 5-1). Vehicle speeds are recorded automatically by a device such as a roadside radar, rather than being judged by an observer. This has two practical consequences that recur throughout this chapter: (1) the measuring device usually counts every passing vehicle, so the separate manual traffic count of Chapter 7 is generally unnecessary; and (2) the equipment must meet specific technical minimum requirements.

Other difference with the standard approach for behavioural indicators is the requirement of free-flowing traffic (see next section) which considerably affects the methodology, and the fact that a whole range of speed and speeding indicators can be produced from one and the same dataset. These SPIs do not require separate studies but only need different analyses of the same set of individual spot-speed measurements.

Because they share a single dataset, almost everything in this chapter (equipment, free-flow definition, vehicle scope, site selection, sampling, stratification, fieldwork and weighting, ...) applies identically to every member of the cluster. In Trendline, the percentage of vehicles travelling within the speed limit is the minimum required headline indicator, but the average speed, the V85 and the dispersion measures for alternative thresholds each describe a different and equally legitimate facet of the same underlying speed distribution.

10.4 Scope

10.4.1 Free-flow traffic as the measurement frame

A central methodological choice for speeding indicator is that measurements should be limited to **free-flowing traffic**. This means traffic conditions in which drivers can freely choose the speed they drive and are not restricted by other vehicles, traffic jams, infrastructure (for example speed bumps) or road works. The reasoning for this restriction is that in congested conditions, at traffic-calming features, in road works or on the immediate approach to intersections, drivers cannot freely choose their speed. Any compliance figure obtained in such situations would reflect the infrastructure rather than the driver. To isolate the behavioural element of the indicator, the methodology requires that observation locations are selected so that drivers can travel at the speed limit if they so wish, and that individual vehicles included in the analysis have a sufficient headway to the vehicle in front.

The operational rule is that the headway between an observed vehicle and the vehicle preceding it should be at least equivalent to the distance covered in five seconds at the prevailing speed limit. This would be, for instance, approximately 42, 69 and 167 metres at 30 km/h, 50 km/h and 120 km/h respectively. Vehicles in shorter following distances are assumed to be car-following rather than freely choosing their speed, and are excluded from the analysis. International experience confirms that this five-second headway is sufficient to

guarantee free-flow conditions for the purposes of speed monitoring. An operational requirement is also that the road sections where measurements are taken should allow for free-flow traffic.

This requirement also explains why the concept of *spot speed* (the instantaneous speed at a defined point) rather than *section speed* (the average speed over a length of road) underlies the indicator: section-based measurement averages over a stretch of road and cannot guarantee free-flow conditions for each individual vehicle. Of course, non-free-flow traffic can also be analysed separately, for instance to study the effect of congestion on average speed, but such analyses should be presented alongside the standard free-flow speed indicators.

10.4.2 Vehicle types covered

The minimum scope of the speeding indicators is the speed of **passenger cars** (see the definition in Chapter 3). Countries are encouraged, but not obliged, to extend the indicators to other vehicle types: motorcycles, light goods vehicles, heavy goods vehicles (including road tractors), and buses or coaches. It should be noted, however, that the speed of some vehicles such as motorcyclists and LGVs cannot be measured correctly by most radars, so other measurements methods are needed (e.g., roadside observers with laser guns, video-based recording).

When more than one vehicle type is observed, results should be reported separately for each of them. Aggregating speeds across vehicle types is often not meaningful, particularly because heavy goods vehicles (HGV) and buses are subject to different speed limits in many countries. A device that classifies vehicles purely by length may place vehicles with different speed limits in the same category, which creates a combined indicator for 'long vehicles' that may be of limited value (if buses and HGVs have different speed limits on motorways, it may be possible to distinguish them in most of the cases). Small vans can be difficult to distinguish from passenger cars even with automatic equipment. In the Trendline programme it is accepted that a small proportion of small vans may be misclassified as cars, which is tolerable since small vans differ little from cars in size and driving characteristics, and recommends that the classification logic of the measuring device be documented in the metadata. Also, if it is well known that on certain road sections there are hardly any buses, it can be acceptable to classify all observed "long vehicles" as HGVs.

10.5 Sampling and stratification

10.5.1 General principles

The general logic of stratifying, selecting and sampling observation locations is set out in Chapter 6 and applies in full to the speeding SPIs, including the three-step random procedure, the preparation of a reserve list of additional candidate sites, and the documentation of any on-the-day substitutions in the metadata.

Before sites are selected, the dimensions along which the sample is to be stratified must be fixed, because they determine how many locations are needed and how they should be distributed. For the speeding SPI the primary stratification dimension is the road type; the time period is a further required dimension; and region is an additional dimension that a country may choose to add.

10.5.2 Road types

The speeding indicators must, as a minimum, cover motorways, rural non-motorway roads and urban roads, and results must be presented separately for each. Within a road type, more than one speed limit may apply. For example, several countries operate rural roads at 70, 80 and 90 km/h. Aggregating data across speed limits is not meaningful, because the very definition of compliance depends on the limit. Countries should therefore report the indicators separately by speed limit or, where this is not feasible, restrict the analysis to the most prevalent speed limit on each road type. The same logic applies to the choice between single- and dual-

carriageway sections of the same speed limit: the two distributions of operating speeds should not be mixed, and where they cannot be reported separately the most prevalent configuration should be chosen as more representative of the network. On multi-lane roads, simultaneous measurement of all lanes is desirable but not required; one radar per lane (typically from a bridge) or alternation of measurement between lanes across the sample of locations are both acceptable, provided the choice is documented.

10.5.3 Time periods

The minimum requirement is to measure speeds during weekday daytime hours. Weekend and nighttime measurements are optional but strongly recommended, because the composition of traffic and driver behaviour differs significantly between these periods; comparisons between day and night are especially valuable. Where additional time periods are reported, the harmonised European Road Safety Observatory definitions should be used (see Chapter 3).

Measurements should be carried out in months that are representative of normal traffic conditions, which excludes school and bank-holiday periods (in particular the long summer break) as well as hard winter, when adverse weather distorts speed distributions. Late spring and early autumn are the recommended windows. Countries with a historical time series should use the same months as in earlier waves wherever possible, in order to preserve trend comparability.

10.5.4 Sampling procedure

To produce a representative national picture, locations are selected in three steps.

- First, the required number of locations per stratum is determined for the country or for each region.
- Second, the corresponding number of points is randomly drawn on a national or regional map, ensuring a reasonable geographical spread. The specific requirements of each location (road type, speed limit) need not be considered at this point, which is intended only to secure a reasonable spread. A Geographical Information System (for example ArcGIS) is a convenient tool for this step. Where no GIS is available, general-purpose mapping tools such as Google Maps or Google Earth can be used.
- Third, the operational measurement site is chosen in the immediate vicinity of each random point, taking into account the location requirements (see Section 10.5.6). Tools such as Google Street View are well suited to the final selection step, possibly complemented by a site visit. When selecting the location, it should be verified that traffic is not congested too often so that you can expect free-flow most of the time.

When speed measurements are repeated regularly, for example every two years, it is recommended to use the same site locations as in the previous measurement, except for sites where the context has changed so that they no longer meet the location criteria. Such sites should be replaced by new sites that meet the criteria.

Wherever possible, the sampling frame should consist of a database of uniform road segments with attributes such as road type, posted speed limit, annual average daily traffic, number of lanes and length; additional attributes such as slope, curvature, type of median and surrounding land use further improve the quality of stratification. For each selected location, the basic characteristics should be recorded at the start of each observation: GPS coordinates, address or other geographical reference, target lane and direction of observation.

In several countries, permanent traffic counters on major roads are already used to derive speed statistics. Such infrastructure-based data are not the result of random sampling and are therefore not, by themselves, representative of the national network. A hybrid approach can be acceptable: a smaller random sample of additional locations on non-main roads is added to the permanent-counter data, and a specific weighting procedure then corrects for the respective shares of main and non-main roads in the calculation of the indicators. Another approach, suitable for road networks with a high number of traffic counters, is to select and use only those counters where free flow is typically possible and where no systematic acceleration or deceleration occurs (such as at the exits and entries of motorways).

10.5.5 Minimum sample size

The Trendline minimum sample requirements for the speed indicator cluster mirror those used for the other behavioural SPIs (Section 6.3). At national level, the total minimum is 2,000 observed vehicles across the three mandatory road types. Each road type must contribute at least 500 observations and at least 10 different measurement locations, and the share of observations on each road type must amount to at least 20% of the total. If more than one speed limit is considered, e.g. rural roads with speed limits of 70 and 90 km/h, then these minimum numbers of observation apply to each of these strata.

Exceptions are possible in countries where a particular road type does not exist or carries only a very small fraction of the traffic (for example motorways in some smaller Member States).

In practice, however, the number of observations often greatly exceed the minima, because automatic equipment records every passing vehicle. Since separate samples are taken for each road type and only straight segments meeting strict requirements are considered, the variance between locations should be quite small. Where a location nevertheless shows large variance, it should be checked against the inclusion criteria and either replaced or reinforced with additional observations.

Disaggregation by region (for example NUTS 1 areas) is an optional further stratification dimension. Countries that wish to publish regional speed indicators should apply the national minimum requirements within each region and weight the regional results according to traffic volumes when aggregating to the national level. Genuine exceptions, such as islands, should be precisely defined and communicated.

10.5.6 Location requirements

Ideally the locations selected to obtain speed data should be representative of the whole road network of a country. Road design characteristics and the surrounding environment influence the speeds at which drivers operate, so not every location is suitable for free-flow speed measurement. The original document of the European Commission (2020) specified three minimum requirements for the observation locations:

- the selection should be as random as possible with the objective of a representative national sample, while roads with known or perceived speed problems are best omitted as unrepresentative;
- measurements should not take place near speed cameras, whether fixed or mobile;
- a minimum traffic flow of at least ten vehicles per hour is required.

Beyond these, all places where vehicles are likely to stop, accelerate or brake should be avoided, since free-flow traffic cannot be guaranteed there. Each location should ideally meet the following criteria:

- a straight and uniform section of road, ideally with no nearby curves that might influence the speed at the measurement point
- a section where it is physically possible to drive faster than the speed limit
- a small longitudinal gradient (less than 5% over at least the 500 metres preceding the point)
- away from intersections (more than 500 metres)
- away from any traffic-calming device, such as speed bumps or lane narrowings (more than 500 metres);
- away from road works (more than 500 metres)
- away from pedestrian crossings (more than 500 metres)
- away from any speed-limit change or sign (more than 500 metres)
- away from sections where speed is enforced (for example traffic-enforcement cameras).

Other considerations for the choice of locations are:

- the speed limit is undeniable and clearly posted
- the speed limit has not been changed for at least six months before the measurements
- the measurement equipment can be safely installed at the location.

It is acknowledged that meeting all of these criteria simultaneously can be difficult, especially in built-up areas, where the criteria may be relaxed; even then the location should come as close to the ideal as possible, and any departure should be recorded in the metadata.

10.6 Data collection

10.6.1 Specific features

When it comes to data collection, the approach used for speeding SPIs differs somewhat from the procedure for observer-based (behavioural) SPIs (see Chapter 7) in a number of operationally important respects:

- First, the unit of data collection remains the *observation session* at a particular location, but the session is typically run by an unattended automatic device rather than by an observation team. A radar may record continuously for a period of days, with each defined time slot at a location treated as a session for the purposes of weighting.
- Second, because the device counts every passing vehicle, the separate manual traffic count required for observer-based SPIs is generally unnecessary: the count is produced by the same instrument that measures the speeds. The number of vehicles passing during the session is therefore known directly, which simplifies the within-stratum weighting described in Section 8.2.3).
- Third, the safety of the personnel installing and recovering the equipment is the dominant field-safety consideration, and installation should be carried out by trained and experienced personnel in accordance with national safety regulations and traffic law.

The posted speed limit in force at the location must be confirmed at the time of measurement, since it is the reference against which compliance is defined and it enters the analysis as the stratum-level variable $FSL(i)$ (

Table 8-1). Any location where the limit is temporarily altered (e.g., by variable message signs or temporary road-works limits) should be excluded. The basic location characteristics (coordinates, road type, target lane and direction, confirmation of free-flow conditions) and the equipment metadata (device type, calibration, error treatment) are to be recorded at the start of each session.

Measurements must not be carried out in adverse weather. Heavy rain, snow, ice, fog and strong winds all distort operating speeds and can also affect the performance of the measuring equipment. Countries can define their own exclusion thresholds and should report them with the data; the personnel installing the equipment are usually best placed to advise on the amount of rain that would compromise data quality.

10.6.2 Measurement equipment

Speed must be measured as **instantaneous (spot) speed** at a defined location. The chosen method must allow the observation of instantaneous speed in free-flow situations. Section-based measurements such as those produced by Automatic Number Plate Recognition (ANPR) systems are excluded, because they average over a length of road and cannot guarantee free-flow conditions for each individual vehicle. Floating-car data and speed traces extracted from smartphones might be used when the underlying sample of drivers can be shown to be representative of the wider population, which in practice is rarely the case. The wider possibilities offered by such data sources are discussed separately in Section 10.9.

In practice, for most cases, three families of devices are used:

- **Roadside devices** such as Doppler microwave radars, frequency-modulated continuous-wave radars, LIDARs, active infrared sensors and roadside cameras
- **In-road devices**, in particular inductive loops and axle detectors using pneumatic tubes, piezo-electric or quartz-electric sensors
- **Handheld devices** such as radar guns and laser guns.

Because speed measurement is usually carried out over a short period and ideally at a large number of locations, equipment that can be installed quickly and flexibly is preferred; portable radars or cameras are therefore the natural choice, although permanent or semi-permanent systems such as induction loops are included here for completeness. Doppler microwave radars are the most used and probably the most recommended option. Handheld devices are flexible but conspicuous, and their use is not recommended: drivers who notice the device adjust their behaviour, and modern alerting applications spread the position of mobile checks rapidly to other road users.

A fuller description of the individual device types, the minimum functional capabilities that the equipment should meet, and guidance on installation and calibration are given in Appendix E (Equipment specifications).

10.7 Variables, analysis and reporting

10.7.1 Overall approach

The statistical methodology used to turn raw speed measurements into national indicators is the common two-layer weighting and complex-survey framework of Chapter 8, and is not reproduced here. This section explains how that framework is applied to speed, and then sets out the calculation rule specific to each member of the indicator cluster. The essential point is that all members of the cluster are computed from the same weighted dataset: the weighting is performed once, and the various statistics are then be read off the weighted distribution of individual speeds.

The values recorded should be those measured by the instrument. Countries should not subtract the device error margin from measured speeds before applying the speed limit. Taking the measurement error into account, as is often done when radar measurements are used by the police for enforcement, would amount to

redefining a 100 km/h zone as a 103 km/h zone, and would introduce a systematic downward bias in the share of vehicles recorded as exceeding the limit. A typical 3% radar error margin goes both ways, positive and negative; hence using the raw measured speeds preserves the symmetry of the error and yields the best estimate of true compliance. The treatment of measurement error should nevertheless be recorded in the metadata so that it can be taken into account when describing international differences.

For speed, the stratum is a combination of road type + speed limit, time period and (where relevant) vehicle type and/or region. The posted speed limit $FSL(i)$ is recorded as a stratum-level variable because compliance is defined relative to it. As mentioned before, because in most cases all vehicles are counted, the within-session resampling is often not necessary. For a given road type, between-strata aggregation follows the weighting ladder of the data-analysis chapter (Section 8.2), the exposure measure being vehicle-kilometres travelled. No national indicator combining all the road types and speed limits is required.

10.7.2 Calculating each member of the cluster

Once the weights are determined, the individual indicators are obtained as follows. Each rule is applied within each stratum and then aggregated to the national level using the between-strata weights; the same calculation gives the stratum-specific value when applied to a single stratum and the national value when applied to the full dataset.

- **Percentage within the limit.** Each observation is a binary (compliant / not compliant) outcome, and the indicator is the weighted proportion of compliant vehicles. A vehicle is compliant when its measured speed does not exceed the posted limit $FSL(i)$ for its stratum.
- **Percentages not respecting the speed limit with a certain threshold.** Each threshold is a further weighted proportion, computed exactly like the headline indicator but with the compliance test replaced by the relevant threshold (e.g. 10 km/h).
- **Average speed, standard deviation and standard error.** The average speed is the weighted mean of the individual measured speeds; the standard deviation is the weighted standard deviation of the same distribution, and the standard error follows from the complex-survey variance estimate that respects the clustering of observations within sessions.
- **V85.** The V85 is the 85th percentile of the weighted speed distribution. Percentiles are read directly from the weighted distribution of individual speeds.
- **Dispersion.** The speed-dispersion is the difference between the 90th and the 10th percentiles of the weighted distribution of individual speeds.

10.7.3 Confidence intervals

Every reported value carries a 95 % confidence interval computed by a method that respects the clustered design, as explained in Chapter 8.

For the proportion-type indicators (percentage within the limit and the ASI thresholds) the Korn–Graubard construction recommended in Chapter 8 applies directly. For the percentile-based indicators (V85 and the V90–V10 dispersion), where the sampling distribution is non-normal, the bias-corrected and accelerated (BCa) bootstrap is the recommended method. The V85 has historically often been reported without a confidence interval because of the difficulty of computing one; the bootstrap now makes a defensible interval feasible, and the technique used should always be documented in the metadata. Dedicated complex-survey software is recommended, as listed in Chapter 8.

10.7.4 Reporting and data delivery

Countries can report one or more speeding SPIs of the indicator cluster. At least the principal SPI defined by the percentage of vehicles travelling within the speed limit should be reported, but it is recommended to

report all of them, including the dispersion indicators that were still labelled as “experimental” in the first phase of Trendline. National indicators must be reported separately by the minimum required parameters: vehicle type, road type (each type ideally broken down by speed limit), time period and region (if applicable). For the experimental indicators, variations in the methodology and thresholds can be considered when duly justified for a specific country.

Results can be delivered at the three levels of detail, as defined in Part D (see Table 23-6). The minimum level for the speeding cluster is the headline percentage within the limit with its confidence interval and the minimum-required disaggregations; the ideal level is valuable here because the cleaned individual-vehicle dataset allows the coordination team to recompute every member of the cluster on a consistent basis.

Each delivery should be accompanied by metadata. For speed, the metadata are particularly important because much of the cross-country comparability of the indicators depends on choices that are invisible in the headline figures. The metadata should describe at least (see also the data submission template):

- the measurement equipment used and its accuracy
- the treatment of measurement error
- the location-selection procedure
- the field-work period and months
- the legal speed limits in force (the *FSL(i)* values)
- the weighting version and the confidence-interval method used
- the applied national regulations on speeding
- any deviation from the minimum requirements.

10.8 Minimum sampling and reporting requirements

Table 10-2 consolidates the minimum and optional requirements for the whole speed indicator cluster. It mirrors the format used for the other behavioural SPIs and can be read as the speed-specific specialisation of the common requirements set out in Chapter 6.

Table 10-2. Trendline minimum requirements and optional additions for the speeding indicator cluster.

Aspect	Minimum requirement	Optional
Indicator	% of vehicles within the speed limit (free-flow) Average speed, standard deviation and standard error V85	% exceeding speed limit by >10 / >20 / >30 km/h (20 Km/h being the baseline) Additional speeding thresholds matching national law Speed dispersion (V90–V10) Non-free-flow data % of kilometres driven within the limit
Vehicle types	Passenger cars (M1)	Motorcycles Light goods vehicles Heavy goods vehicles Buses and coaches
Road types	Motorways Rural roads Urban roads (reported separately)	Single vs dual carriageway separately Separate speed-limit classes (for urban roads: at least 30 km/h and 50 km/h) Expressways reported separately from motorways
Time period	Weekday daytime Late spring or early autumn	Weekend Nighttime hours

Sample size	≥ 2,000 observations in total ≥ 500 per road type ≥ 10 locations per road type Each road type ≥ 20% of the sample	Larger samples for regional indicators and for crossed strata
Conditions	Free-flow traffic Good weather No enforcement equipment nearby ≥ 10 vehicles/hour	Stratification by region (NUTS 1)

10.9 Alternative measurement methods

The methodology above assumes that spot speeds are measured with traditional roadside equipment, principally radar. A number of alternative methods exist for measuring or estimating vehicle speeds, ranging from a simple stopwatch to satellite imagery. None of these is part of the standard Trendline methodology, and most do not currently satisfy the free-flow and representativeness requirements that the standard methodology imposes. They are nevertheless described briefly here, both because they may be the most practical option in some settings, particularly in low- and middle-income countries or for very large networks, and because they are developing rapidly and may become more relevant to SPI work in future. Where any of these methods is used, the limitations noted below should be made explicit in the metadata, and the values produced should be presented as complementary to, not as a substitute for, the standard free-flow indicators.

10.9.1 Low-technology measurement: the stopwatch method

Speed can be measured with almost no technology, using only a stopwatch. The *stopwatch method* retains real value as a low-cost approach, especially in low- and middle-income countries (Fondzenyuy, 2024). Two observers are positioned a fixed distance apart (e.g. 60 metres) along a road section. The first observer signals when a vehicle crosses the start line; the second observer, acting as timekeeper, starts the stopwatch on that signal and stops it as the vehicle crosses the finish line. The recorded time and the known distance give the vehicle's average space speed over the section. The method is inexpensive and requires no specialised equipment, but it measures average speed over a short section rather than true spot speed, is labour-intensive, and its accuracy depends heavily on the observers' reaction times.

10.9.2 GPS and floating car data

GPS data from vehicles or mobile phones offer anonymised information about how road networks are used, and a substantial commercial market supplies them. At the simpler end, journey-time services (such as the routing APIs of Google, Mapbox, TomTom and HERE) provide estimated travel times between two points, from which an average speed can be derived using the known road length. These sources carry no information on vehicle mode, so journeys by bicycle or public transport may be included, and they are more suited to journey planning than to road-safety measurement.

A higher-quality source is in-vehicle GPS from satellite-navigation systems, fleet trackers and insurance telematics devices. Such data are usually delivered already aggregated and map-matched to the road network, and they yield link-based average and 85th-percentile speeds together with a probe count that indicates data quality. The principal attraction of GPS data is temporal depth: whereas a roadside survey may cover only a few hours, GPS data can describe a road link over a whole year, with the ability to exclude atypical periods such as national holidays, which makes the data well suited to before-and-after studies of speed-limit changes, enforcement or layout changes. The principal limitations for SPI use are a bias towards newer vehicles, the need for trimming of journey starts and ends (which reduces coverage on low-flow residential roads), the production of link averages rather than point spot speeds, and the difficulty of isolating free-flow

conditions. The last difficulty can be partly addressed by using 85th-percentile speeds, which are largely unaffected by congestion, or by emerging techniques that separate free-flow from congested traffic by fitting a mixed (two-component) speed distribution.

10.9.3 Camera-based and computer-vision systems

The proliferation of roadside and traffic CCTV has made video a significant potential data source. Vision-based systems detect, classify and track vehicles across successive video frames and estimate speed from the tracked movement, combining speed with vehicle class and volume to build a detailed picture of traffic. The technology is mature, with several commercial and open-source solutions available, and camera positioning and resolution are the main determinants of detection quality. Some products are well suited to lower-resource settings: plug-and-play sensors that attach to a window facing the road, process the video on the device for privacy reasons (so that no video is transmitted), and report classified counts and speed bands (including V85) at regular intervals. Camera-based methods can in principle approximate the spot-speed and the free-flow logic of the standard methodology more closely than GPS, cell-tower or satellite data. As the technology matures they may become a practical complement to radar for some SPI applications. For now, careful attention to camera siting, calibration and the representativeness of the locations covered remains essential.

10.9.4 Earth-observation (satellite) data

Vehicle speeds can also be estimated from high-resolution satellite imagery. This approach was piloted, alongside others, in a World Bank project in 2021 that mapped travel and trialled the detection of speed, flow and physical road features in Kenya and Ethiopia, and the speed-detection methodology was reviewed and accredited under the AiRAP programme of iRAP. Very-high-resolution imagery (for example from the WorldView-2 satellite, at around 0.46 m for panchromatic and pansharpened imagery) is procured for defined road corridors. Vehicles are then located using computer-vision techniques—most robustly by semantic segmentation, which classifies each pixel and so allows vehicle size and bearing, and hence road-matching, to be determined. Speed is estimated by exploiting the fact that the satellite's several multispectral bands are recorded a fraction of a second apart (a total span of about 0.35 s): a bright, light-coloured vehicle appears in a slightly different position in the earliest and latest bands, and the displacement of its centroid over the known time interval gives an approximate speed. Detected movements are matched to the nearest plausible road by proximity and bearing, and aggregated into link-level average and V85 speeds.

Validation against telematics data on a sample of the United Kingdom network showed strong correlation for both mean and 85th-percentile speeds. The method nevertheless has clear limitations for SPI use: in its current form it detects only larger vehicles and largely misses motorcycles, which matters where two-wheelers dominate the fleet; it relies disproportionately on light-coloured vehicles (though these make up a large share of the fleet); speeds below roughly 20 km/h cannot be detected; and where imagery coverage or vehicle counts are insufficient, values must be interpolated from adjoining roads. It captures a single instant rather than a sustained free-flow measurement. The approach is therefore promising for broad network-level monitoring, particularly where ground-based surveys are infeasible, but it is not at present a substitute for the standard methodology, and its outputs should be used with a clear understanding of these limitations.

11 SPIs for restraint systems (seat belts and CRS)

11.1 Introduction

The use of the seat belt and of child restraint systems (CRS) is one of the most effective elements of passive vehicle safety. Like helmet use, it does not prevent a crash from occurring but drastically alters the energy transferred to the human body in a collision, shifting an outcome from “fatal” to “serious” or from “serious” to “minor”. Crash analysis shows that a significant proportion of the fatally or seriously injured vehicle occupants were not using the seat belt or child restraint system, or were not using it correctly, at the moment of the crash. The causal mechanism here is biomechanical rather than behavioural, and it applies to every occupant of the vehicle rather than to the driver alone.

Because the protective effect depends on whether each individual occupant is correctly restrained, the restraint indicators differ from most other behavioural SPIs in a way that shapes the whole of this chapter: the relevant behaviour has to be observed not for the vehicle as a whole, nor for the driver alone, but for several occupants in different seating positions, each of whom may behave differently. Drivers, front passengers and rear occupants show systematically different wearing rates, and children require a different protective device altogether. The restraint SPIs therefore measure a cluster of occupant-specific behaviours that happen to be observed at the same time and place.

11.2 Indicator definition and theoretical population

11.2.1 The restraint indicators cluster

The principal indicator is defined as:

Percentage of vehicle occupants using the seat belt or using the child restraint system (CRS) correctly.

The objective is to estimate this percentage for the population of vehicle occupants over the entire road network of a country or region during a defined reference period. In line with the general framing of behavioural SPIs, the theoretical population is the totality of vehicle movements over the national territory, so that the percentage of occupants correctly restrained is interpreted as approximating the percentage of occupant-kilometres travelled with correct use of restraints.

Because the indicator is occupant-specific, the single headline figure is complemented by a set of occupant-position breakdowns and by the distinct CRS indicator. The members of the cluster are:

- the **belt-wearing rate of drivers** of passenger cars;
- the **belt-wearing rate of front passengers** of passenger cars;
- the **belt-wearing rate of rear occupants** of passenger cars (including children);
- a **nationally weighted aggregate belt-wearing rate** for all car occupants combined; and
- the **correct-use rate of child restraint systems**, collected through the supplementary in-vehicle inspection of Part B (Section 11.7).

All of the belt-wearing indicators are obtained from a single roadside observation dataset, in the same way that the speeding cluster is read from a single set of spot-speed measurements: the observation records the restraint status of every occupant of each observed vehicle, and the various occupant-position rates are then read from the same weighted dataset. Where the sample permits, each indicator may additionally be reported for goods vehicles and broken down by the optional variables listed in Table 11-2..

A word of caution applies to all members of the cluster. As for the speeding indicators, differences between countries may arise that are not visible from the indicator definition itself but are embedded in the metadata and the practical implementation. This concerns in particular the national legislation on belt and CRS use, the categories of occupant exempted from wearing a belt, and the visibility conditions under which rear occupants and children can be observed. Consistent documentation of these choices in the metadata is therefore essential for comparability.

Table 11-1. The Trendline restraint indicator cluster.

Indicator	Definition	Method
% drivers belted	Share of observed passenger-car drivers using the seat belt	A – roadside
% front passengers belted	Share of observed front-seat passengers using the seat belt	A – roadside
% rear occupants belted	Share of observed rear-seat occupants using the seat belt	A – roadside
Aggregate belt-wearing rate	Nationally weighted share of all car occupants using the seat belt	A – roadside
% children correctly restrained	Share of children using a child restraint system correctly (no use and misuse excluded)	B – in-vehicle

11.2.2 The two-part structure of the indicator

Given the complexity of determining the correct use of a CRS, two types of observation are foreseen:

- **Roadside observations (Method A):** these allow the use of seat belts to be recorded for all occupants, and the presence of children and of a CRS to be noted for quantitative purposes. This is the standard observer-based behavioural study described throughout Part B of the manual.
- **In-vehicle inspections (Method B):** these allow the *correct* use of a CRS to be assessed. The method requires the cooperation of the driver and is only possible during dedicated sessions in accessible locations, typically parking sites.

It is recommended to conduct both types of observation. However, since Part B can be difficult to carry out in some countries — for example because of legislation, or data-protection constraints on approaching drivers and inspecting vehicle interiors — only Part A is required, and Part B may be omitted. Where Part B is omitted, the reasons must be reported in the metadata. The supplementary in-vehicle procedure is treated separately in Section 10.6, and the remainder of the chapter concerns Part A unless stated otherwise.

11.2.3 Reference for “correct use”

The reference for “correct use” has two meanings:

- The first meaning refers to the national traffic legislation: a behaviour that is legal in the country counts as compliant with the SPI, even where another country’s legislation would treat it differently. Differences exist, for example, on the driver categories that are exempted from wearing a seat belt, or the age and size thresholds for CRS.
- The second meaning is applicable specifically to child restraint systems. It refers to the conformity and instruction label of the CRS itself: the CRS must be used in a way consistent with its own labelled specifications. It is not required to take into account additional national recommendations for the optimal use of a CRS (for example, the Swedish recommendation to use a rearward-facing CRS up to and including four years of age). Compliance with such recommendations may be reported as optional information.

The metadata of the dataset should indicate, in abbreviated form, the specific national rules that apply, so that international comparisons can be calibrated for legislative differences.

A consequence of relying on roadside observation for Method A is that only “use” or “non-use” can be established for the seat belt. In other words, the observation of “use” may therefore also include invisible incorrect use, for instance a belt worn under the arm or behind the back, an incorrect height adjustment of the upper belt guide, or the use of “foreign objects” such as clothes-pegs to deflect the belt or reduce its tension. The use of a CRS, by contrast, can be expressed as “correct use”, “misuse” or “non-use”, but can be established only through the in-vehicle inspection of Method B. There, “correct use” is the complement of “non-use” and “misuse”. If there is no indication of no use or misuse, the use is considered correct, which means that both non-use and misuse have to be actively detected.

11.3 Position within the behavioural framework

As set out in the introduction to behavioural SPIs (Chapter 5), the restraint indicators rest on the common backbone and follow the standard observer-based procedure, including the manual traffic count. Three features set them apart and are the subject of this chapter:

- The behaviour is recorded separately for **several occupants per vehicle**: the driver, the front passenger(s) and the rear occupants. Thus, the unit of observation is the vehicle but the unit of analysis is the individual occupant.
- There are **distinct protective devices, the child restraint system (CRS)**, which are used for young and small children. Its correct use cannot be judged in the same way as belt use.
- For that reason the restraint SPIs are collected through **two complementary methods**: roadside observation (Method A) and a supplementary in-vehicle inspection (Method B).

11.4 Scope

11.4.1 Vehicle types

The minimum scope of the restraint indicators is the occupants of **passenger cars**. The full definition of the passenger-car category is given in Chapter 3 and is not repeated here. Member States are encouraged, but not obliged, to extend the indicators to goods vehicles, ideally separating light goods vehicles (LGV, sometimes called “van”; maximum gross weight under 3.5 tonnes) from heavy goods vehicles (HGV, sometimes called “truck” or “lorry”; maximum gross weight over 3.5 tonnes). Where other vehicle categories are included, their results must be reported separately and never aggregated with passenger cars.

The different vehicle types and their specific categorisation should be clearly defined and illustrated for the observers during training and briefing, because some models exist in both passenger-car and LGV versions with only limited external differences: a passenger car typically has rear side windows and rear passenger seats, whereas the corresponding LGV has neither. Misclassification of these borderline vehicles is the most common vehicle-typing error in restraint observation and should be addressed explicitly in observer training.

11.4.2 Occupant positions

Within each observed vehicle the restraint status is recorded separately by seating position. The minimum reporting categories are **drivers, front-seat passengers and rear occupants**, and results must be presented separately for at least these three. This per-position recording is the defining feature of the restraint observation: a single passing vehicle yields one record for the driver, a record for each front passenger and a record for each rear occupant, so that the number of occupant observations substantially exceeds the number of observed vehicles.

In modern passenger cars there is normally a single front passenger, but provision should be made for a second front passenger, which still occurs in older designs and in some commercial vehicles, and for up to three or four rear occupants. The driver is the person controlling the vehicle; passengers are those merely riding.

11.4.3 Children and CRS age groups

Children are treated as a distinct occupant category because the protective device that applies to them depends on their age, height and/or weight. For reporting purposes children may be recorded as a single category or split into “very young children” — typically those who, under national law, must travel in a child restraint system — and “young children”, who are old or large enough to use the standard adult belt. The precise age, height or weight thresholds are set by national legislation, and the boundary used must be recorded in the metadata so that the CRS indicator can be compared across countries. In roadside observation the presence of a child and of a CRS can be noted, but, as explained in Section 10.1, whether the CRS is used *correctly* can be established only through the in-vehicle inspection of Part B.

11.4.4 Exempted occupants

Vehicle occupants who are legally exempted from wearing a seat belt should be excluded. Typical examples are, depending on the applicable legislation of the respective country: postal-delivery personnel, taxi drivers and the occupants of emergency vehicles. The most practical solution is to exclude the whole vehicle from the data collection. Because the legislation on belt and CRS use, and the exemptions from it, vary between countries, every country must document its national rules and consequently record which vehicles were excluded from the observations.

11.5 Sampling and stratification

11.5.1 General principles

The general logic of stratifying, selecting and sampling observation locations is set out in Chapter 7 and applies in full to the restraint SPIs, including the three-step random selection procedure, the preparation of a reserve list of candidate sites, and the documentation of any on-the-day substitutions in the metadata. As for the other behavioural SPIs, the primary stratification dimension is the road type, the time period (week period) is a further required dimension, and region is an additional dimension that a country may choose to add.

11.5.2 Road types

The restraint indicators must, as a minimum, cover motorways (including expressways if these are considered), rural non-motorway roads and urban roads, and results must be presented separately for each road type and aggregated, after weighting, for the whole network. The full definitions of these road types are given in Chapter 3 and are not repeated here. A deviation from this requirement is only possible in the exceptional case that a particular road type does not exist in a country (for example, no motorways in some smaller Member States). Where the network contains all required road types but the survey covers only some of them, the results for the covered types remain disaggregated and are not aggregated into a national figure.

Roadside observations on motorways may, for feasibility and safety reasons, be organised not on the motorway itself but at its entrances or exits; this should be indicated in the metadata. The main characteristics of the included road types, such as road signs, speed regimes, number of lanes, lane separation, permitted vehicles, should also be described in the metadata, as contextual information that allows the general correspondence of road types between countries to be assessed.

11.5.3 Time periods

Observations are required for both **weekdays and weekend days**, observed and presented separately, because the composition of traffic and the mix of journey purposes differ between them. This difference matters more for restraint use than for some other behaviours, since family and leisure trips at the weekend carry a different occupant profile, including more children, than the commuting traffic of a weekday morning. Observations should cover the whole of the daytime and be carried out in reasonably good weather.

As for the other behavioural SPIs, measurements should be timed for late spring or early autumn. All months are allowed; but December, January, July and August; bank-holiday and school-holiday periods, and hard winter conditions, should be avoided because they disturb normal traffic and occupant patterns. Where a Member State holds a historical series, it should use the same period of the year as in earlier waves in order to preserve trend comparability. There should be a balance between all combinations of road type and time period, to avoid systematic sample bias.

11.5.4 Location selection and minimum number of locations

Locations are selected by the standard three-step random procedure of Chapter 6, which is not fully repeated here. Unlike speed measurement, restraint observation does not require free-flow conditions; on the contrary, sites where vehicles slow down or stop are preferred, because they make the occupants easier to see. Observations on urban and rural roads should therefore be carried out from a safe place along the road, preferably where driving speed is reduced relative to the limit, such as the approaches to towns and villages or intersections where vehicles slow or stop at traffic lights. Observations associated with motorways are possible at the last intersection before an on-ramp, the first intersection after an off-ramp, service stations, rest areas and toll stations.

A minimum traffic flow of at least ten relevant vehicles per hour is required. If a randomly selected road carries less traffic than this, a nearby higher-volume road may be substituted on the assumption that most of its drivers also use the lower-volume road. Because the three road types must all be covered, the proportion of observations at each road type should be at least 20% of the total, even if this implies disproportionate sampling relative to traffic volume. Where national traffic-volume data by road type are available, the sample may instead be allocated in proportion to traffic volume, provided each road type retains its 20% minimum. In any case, a minimum of ten different locations per road type should be selected for the national indicator; ideally more than ten locations per stratum, and more than two locations per crossed road-type × week-period combination, are used, in order to allow more accurate estimation of the disaggregated indicators.

Where cameras are used for Method A, they must be installed on all road types to avoid selection bias. Faces and number plates must not be captured for privacy reasons, and the method must be tested and validated before deployment. The approach must comply with national and international requirements on ethics, privacy and data protection.

11.6 Data-collection specifics

Restraint observation follows the standard observer-based procedure of Chapter 7. The points below concern only the respects in which it differs from, or specialises, that common procedure.

11.6.1 Random selection of vehicles within a session

Vehicles to be observed are selected at random from all the vehicles passing the location. The practical rule is that, after one vehicle has been fully coded, the next passing target vehicle is observed. An observation is only completed if all vehicle occupants have been observed and their compliance has been recorded. At signalised or queuing locations the first vehicle in the queue is observed and the observer then proceeds to the following

vehicle until the queue begins to move. This avoids the bias that would arise if observers preferentially selected vehicles that were easy to read.

11.6.2 Coding the restraint status of each occupant

For each observed vehicle, separate values are coded for the driver, the front passenger(s) and the rear occupant(s). Only “use” or “non-use” is recorded for the seat belt, and it is recalled that “use” may also include incorrect use. As with all observer-based SPIs, the borderline cases are taught and discussed during training so that observers apply the same decision rules. For restraints, the most important borderline rule concerns what to record when the belt is only partially visible, for instance when the diagonal strap can be seen but the lap section cannot, or when a dark belt is hard to distinguish against dark clothing. The agreed convention must be applied consistently and documented.

11.6.3 Observation challenges specific to restraints

Several practical challenges affect restraint observation in particular and influence both site selection and the realistic level of detail that can be recorded:

- **Rear-seat occupants** are much harder to observe from the street than front-seat occupants, because the rear seats sit behind a smaller window and are seen at a less favourable angle; rear belt-wearing rates are therefore the most demanding part of Method A and benefit most from sites where vehicles slow or stop.
- **Correct CRS use** can almost never be verified from outside the vehicle, since the fit of the harness and the orientation and installation of the seat are not externally visible. In roadside observation, only the presence of a child and of a CRS can be noted; correct use is reserved for the in-vehicle inspection of Method B.
- **Tinted or darkened rear windows** can make rear occupants impossible to see at all; where this prevents observation, the rear positions are left uncoded for that vehicle rather than guessed.
- **Trucks and other high-cab vehicles** raise the occupants above the eye-line of an observer standing at street level, which limits the viewing angle and may require a raised observation point or an alternative method.
- **Nighttime, poor weather and fast-moving traffic** all reduce the reliability of observation. Restraint observation is therefore confined to daytime and to reasonably good weather, and fast sites are avoided in favour of locations where vehicles slow down.

As a general rule, contextual or demographic variables (Section 11.8.5) should be added to the protocol only where they can be observed reliably from the roadside. Variables that depend on the driver’s self-report or on close inspection of the vehicle interior do not belong in a passive observation protocol and, for CRS, are collected only in Method B.

11.6.4 Traffic counting

As indicated in Section 7.3.2, a traffic count should accompany each session as defined in the fieldwork chapter; the count’s unit is the vehicle, in the same categories, location and direction as the observation. Where observation is at a service station or rest area, the relevant count is the number of vehicles entering the site. The traffic count is **not** required for the qualitative in-vehicle inspection of Part B (see Section 11.7).

11.7 Supplementary procedure: in-vehicle inspection of CRS (Method B)

11.7.1 Specific challenges of assessing correct CRS use

The correct use of a child restraint system is in most cases impossible to verify from a roadside observation. Whether the harness is correctly tightened, whether the seat is appropriate for the child's weight and height, and whether a rearward-facing seat has been installed in the correct orientation are all questions that require an inspection of the vehicle interior, normally with the cooperation of the driver. For this reason, correct CRS use is collected in a dedicated in-vehicle inspection sub-study (Method B), organised separately from the main roadside programme. Selection bias is inevitable in a survey that depends on voluntary participation, but this is the only way to detect correct CRS use reliably, and the bias is documented and borne in mind when the results are interpreted.

Each child inspected should be classified as "correct use", "misuse" or "non-use". Four specific challenges affect any study of correct CRS use:

- securing the cooperation of drivers, since the inspection requires their consent
- identifying misuse, which takes many forms ranging from a loose harness to an inappropriate or wrongly installed seat
- observing CRS use at night
- coping with poor weather on both the approach to drivers and the inspection itself.

Possible misuses of a CRS can be grouped broadly into:

- inappropriate use, such as a child not in a CRS when one is required, or in the wrong CRS group
- faulty fixation of the CRS to the vehicle, such as incorrect belt routing around the seat, a missing top tether or support leg, a wrongly oriented seat, or a frontal airbag left active in front of a rearward-facing seat
- faulty fixation of the child within the CRS, such as a harness that is too loose or wrongly routed.

11.7.2 Sites and approach

In-vehicle inspections take place at accessible locations such as car parks, rest areas and similar sites, where drivers can be approached safely and at leisure. The inspection sites should be chosen by random selection of the surrounding area followed by manual selection of suitable parking sites, and the allocation of week periods to sites is balanced across road types; The distribution of inspection sites should also take into account the different types of parking and its associated activity (school, shopping, leisure).

The minimum number of inspection sites is two for each combination of week period and road type (six combinations where all road types are covered). Member States seeking more accurate estimates of correct CRS use are advised to boost the sample to the same number of sites as for the roadside observations (a minimum of ten per stratum). The basic characteristics of each site should be documented, including its coordinates or other geographical reference, the location type and the related activity or service (car park of a school, shop, day-care centre, and so on). Because the inspection is qualitative, no traffic count is required; instead, the number of inspected vehicles with children among the occupants serves as the quantitative basis for weighting the Method B results.

At the inspection site, observers approach either drivers who have just parked, before any child is unbuckled, or drivers who are about to leave a parking space, when children are expected to have just been buckled in. On urban and rural roads, suitable sites are the car parks of places where children are expected, such as schools, day-care centres, shops and leisure facilities, chosen with attention to the appropriate time slots. For motorways, the car parks of service stations and rest areas are used. Since very few children travel in goods vehicles, it is recommended to include only passenger cars in the CRS inspections. Adults' belt use may also be inspected at the same time if desired, recording the number of seats and the correct use, misuse or non-use of the belt for each.

11.7.3 Ethical and legal considerations

Because Method B involves approaching members of the public and inspecting the interior of their vehicles, it is subject to national and international requirements on ethics, privacy and data protection. Participation must be voluntary and informed, and in some countries the procedure may be constrained or impossible for legal or data-protection reasons.

11.8 Variables, analysis and reporting

11.8.1 Compliance

Restraint use is coded as a binary outcome at occupant level: a value of one denotes compliant behaviour (the occupant is using the seat belt, or the child is correctly using a CRS) and a value of zero denotes non-compliant behaviour.

The variables below are organised by data-collection level. Users of the manual can adopt the full list or trim it to the minimum, depending on what their fieldwork can sustain reliably.

11.8.2 Location and session variables

At the location level, each observation site is described by a unique location identifier, an optional region code, the road type, the road number or address, the GPS coordinates of the exact spot, the number of lanes, and the target lane and direction. For Method B sites, the location type and related activity are added in place of the lane information. At the session level, each session is described by a unique session identifier, the date, the start and end times of coding, the total session duration with the inter-vehicle correction applied, and a short weather description. For Part A sessions the duration of the traffic count and the count results by relevant vehicle type are added, whereas for Part B these traffic-count fields are not collected.

11.8.3 Roadside observation variables (Method A)

At the observation level the variables are:

- the vehicle type (passenger car or goods vehicle, with an optional split into LGV and HGV)
- the driver's belt use (use / no use)
- the belt use of front passenger 1 and, where present, front passenger 2
- the belt use of rear passengers 1 to 4.

For young children at the front and rear, one needs to record:

- the belt use
- (where a very young child is present at the front or rear) whether a CRS is used.

Estimated age group and gender may optionally be recorded for each occupant, supporting the demographic breakdowns of the headline indicator.

11.8.4 In-vehicle inspection variables (Method B)

The in-vehicle inspection records, for each child observed and depending on national regulation:

- the child's seating position
- the presence and status of a frontal airbag at that position (not present / activated / deactivated)
- the belt type available there (none / two-point / three-point)
- the presence and correctness of an Isofix anchorage
- the child's age, height and weight.

For the seat itself, it records

- whether the child is fastened in the belt or seat
- whether the seat is fixed to the car
- the correctness of use relative to the child's age and size (correct / misuse)
- the orientation of the seat (forward / rearward / sideways)
- the CRS group and the homologation label.

For the quality of use, it records the routing and tension of the belt and of the CRS harness (correct / false / too tight / too loose / not applicable).

Where adults are also inspected, the number of seats and the correct use, misuse or non-use of the belt are added. Trip characteristics such as purpose and length, and the child's gender and age, may be asked of the driver where the country wishes to analyse them.

11.8.5 Minimalist and maximalist variable sets

The set of variables can be designed at two levels of ambition. A *minimalistic* approach records only what is strictly required to compute the wearing rate: whether each occupant of each observed seating position is using a belt or, for a child, a CRS. A *maximalist* approach adds contextual and demographic information — vehicle type, professional use (for example taxi versus non-taxi), driver gender, occupant age category, the time period, and the presence of police or cameras at the site — that supports more detailed breakdowns. The minimalist approach is sufficient to meet the Trendline minimum requirements; the maximalist approach makes it possible to identify which subgroups are most non-compliant and therefore to target subsequent interventions. Additional variables should be added only where they can be observed reliably from the road-side.

11.8.6 Calculation of the restraint SPIs and confidence intervals

The statistical machinery used to turn raw observations into national indicators is the common two-layer weighting and complex-survey framework of Chapter 8 and is not reproduced here. This section notes only how that framework is applied to the restraint indicators.

The defining feature for analysis is that, although the unit of observation is the vehicle, the unit of analysis is the occupant. The wearing rate is computed across occupants of a given seating position: the percentage of drivers belted, of front passengers belted, and of rear occupants belted. Finally, an aggregate rate is calculated across all occupants. Each occupant record is a binary compliant / non-compliant outcome, and each indicator is the weighted proportion of compliant occupants in the relevant group.

Because several occupants are observed within one vehicle, and several vehicles within one session, the confidence intervals must reflect this nested design exactly as set out in Chapter 8. Approximations that assume simple random sampling of occupants understate the true uncertainty and are not acceptable. The Korn–Graubard construction recommended in Chapter 8 applies directly to these proportion-type indicators, and dedicated complex-survey software should be used.

For the stratum weighting, each stratum is a combination of road type, week period and, where relevant, vehicle type and region. Strata are combined into national figures by weighting each stratum by its share of national traffic exposure, ideally expressed in vehicle-kilometres travelled and, where these are unavailable, approximated from road-network length by road type together with the session traffic counts. Where no traffic-volume or road-length information of any kind can be obtained, the strata are kept separate and not aggregated, exactly as for the other SPIs.

The in-vehicle inspection (Method B) is weighted on a different basis, because it has no traffic count. The number of inspected vehicles with children among the occupants serves as the quantitative basis for weighting the CRS data: the correct-use rate observed in each stratum is weighted by the share of child-carrying traffic that

the stratum represents, using the observations of children in Method A as the exposure measure where possible. The treatment used should be recorded in the metadata.

11.8.7 Reporting and data delivery

The indicator values to be reported are the percentages of vehicle occupants using restraints (Method A) and children correctly restraint in a CRS (Method B). At a minimum, countries should report the belt-wearing rate of drivers, of front passengers and of rear occupants in passenger cars, the nationally weighted aggregate rate, and, if Method B is carried out, the correct-use rate of child restraint systems. The equivalent rates for goods vehicles are optional. Each reported value is accompanied by its unweighted base (the number of occupants on which it is based) and by a 95% confidence interval, and is broken down by the minimum required parameters: occupant position, road type (three levels) and week period (weekday / weekend), for passenger cars.

Results can be delivered at the three levels of detail, as described in Part D (see Table 23-6); the minimum level for the restraint indicators is the belt-wearing rates by occupant position, road type and week period for passenger cars, each with its confidence interval, and the CRS correct-use rate if Method B has also been applied.

Each delivery is accompanied by metadata. For the restraint indicators the metadata are particularly important because much of the cross-country comparability depends on choices that are invisible in the headline figures. The metadata should describe at least:

- the national legislation on belt and CRS use and the categories of occupant exempted from it
- which vehicles were excluded
- the observation method used for Method A (roadside observers, observers in moving vehicles, or cameras)
- when cameras were used, how privacy was protected and the method validated
- whether part B was carried out
- the location-selection procedure
- the fieldwork period
- the visibility conditions affecting rear-occupant and CRS observation
- the weighting version and the confidence-interval method used
- any deviation from the minimum requirements.

11.9 Minimum sampling and reporting requirements

Table 11-2 consolidates the minimum and optional requirements for the restraint indicators. It mirrors the format used for the other behavioural SPIs and can be read as the restraint-specific specialisation of the common requirements set out in Chapters 6 and 7. The occupant-level thresholds reflect the fact that, for the restraint indicators, the unit of analysis is the occupant rather than the vehicle: to support occupant-level breakdowns, minimum numbers of non-driver occupants and of vehicles carrying children are specified in addition to the overall vehicle minima.

Table 11-2. Trendline minimum requirements and optional additions for the restraint indicators.

Aspect	Minimum requirement	Optional
Indicator	% belted: drivers, front passengers, rear occupants Nationally weighted aggregate belt-wearing rate % children correctly restrained (Method B, if feasible)	Breakdowns by age group and gender Correct belt use of adults (in-vehicle)
Method	Method A – roadside observation by trained observers or using cameras	Method B – in-vehicle inspection of CRS Observers in moving vehicles
Vehicle types	Passenger cars	Light goods vehicles Heavy goods vehicles (reported separately)
Occupant positions	Drivers; Front-seat passengers Rear occupants (reported separately)	Per-seat detail; second front passenger; rear positions 1–4
Road types	Motorways; Rural roads; Urban roads (reported separately)	Expressways reported separately; regional breakdown
Time period	Weekday and weekend (reported separately) Daytime Late spring or early autumn, good weather	Additional seasons or periods
Sample size – Method A	≥ 2,000 observed vehicles overall ≥ 500 per road type and per week period ≥ 10 locations per road type Each road type ≥ 20 % of the sample	≥ 650 non-driver occupants overall; ≥ 150 per road type ≥ 100 vehicles with children overall; ≥ 20 per road type Larger samples for crossed strata
Sample size – Method B	≥ 200 vehicles with children inspected (Method B) ≥ 50 per road type and per week period ≥ 2 inspection sites per road-type × week-period combination	Boost to ≥ 10 sites per stratum for more accurate CRS estimates
Conditions	Good weather, daytime ≥ 10 relevant vehicles/hour Random location selection Voluntary participation in Method B	Stratification by region

There can be situations where a particular minimum cannot be met, for example because a road type does not exist, because too few children are encountered for the CRS indicator, or because Method B is legally constrained. The shortfall and its reason must be reported in the metadata, and the affected indicators are reported in disaggregated form rather than aggregated into a national figure.

12 SPIs for helmet use

12.1 Introduction

12.1.1 Rationale

The use of a protective helmet is one of the most effective elements of passive safety for the riders of two-wheeled and similar vehicles. As for the seat belt, a helmet does not prevent a crash from occurring but alters the energy transmitted to the head in a collision, and the head is the body region whose injury is most often decisive for survival. A protective helmet is mandatory for the riders of powered two-wheelers in all Member States; for cyclists and for the riders of personal mobility devices the legal position varies between countries and sometimes by age group, which is itself one of the reasons why the helmet indicator is of policy interest for these groups.

The Helmet SPI measures the percentage of riders of two-wheeled and related vehicles who wear a protective helmet while in traffic. Following the SafetyNet project (Hakkert and Gitelman, 2007), a helmet is understood as a crash or safety helmet specifically designed for two-wheeler use, whether motorised or non-motorised. The indicator measures the use of a helmet, not the correctness of its use, since it is often difficult to observe from the roadside the legal definition of correct use varies by country and would limit international comparability. The one exception, for powered two-wheelers, is the treatment of the chin-strap, discussed in Section 12.3.1.

12.1.2 Indicator definition and theoretical population

Three types of riders are considered: cyclists, motorcyclists and PMD-riders (see Section 12.1.4). For each subcategory, the indicator is defined as:

Percentage of riders wearing a protective helmet.

The theoretical population is the total number of kilometres ridden over the national territory by the relevant vehicle group during the reference period. By weighting the sample by traffic volume, the percentage of riders observed wearing a helmet may equivalently be interpreted as the percentage of kilometres ridden with a helmet, which is the policy-relevant exposure measure. Where a country wishes to monitor passengers, the same definition applies to passengers, but passenger monitoring is optional and is often constrained by very small numbers.

Because helmet legislation differs between Member States, particularly for cyclists and PMD riders and at times by age group, every country must document the applicable rules and any age-specific provisions in the metadata accompanying the data delivery. A note should accompany the results for cyclists stating the existing legal requirements, or the absence of requirements, concerning helmet use.

12.1.3 Position within the behavioural framework

As set out in Chapter 5, the introduction to behavioural SPIs, the helmet indicators rest on the common backbone and follow the standard observer-based procedure, including the manual traffic count. What distinguishes them is the three-way subcategory structure (cyclists, powered two-wheelers, PMDs), developed below. The subcategory-specific divergences are gathered in the subcategory Section 12.4.

12.1.4 Scope

Helmet use is reported as a cluster of three subcategory indicators that share most of their methodology but differ in their target populations, legal frameworks, road types and sampling design:

- **cyclists**, including the riders of power-assisted bicycles (e-bikes)
- **riders of powered two- or three-wheelers (PTWs)**, including both mopeds and larger motorcycles
- **riders of personal mobility devices (PMDs)**, the family of micromobility devices, including in particular e-scooters.

Results should always be reported separately for each subcategory, since their exposure patterns and risk profiles differ markedly and the policy levers that act on them are not the same.

In principle, e-bike could be classified as bikes, PTWs or PMDs. In this manual, riders of e-bikes are considered as **cyclists**, alongside riders of conventional bicycles. The recommended optional variable “type of bicycle” then distinguishes electric from non-electric cycles within that subcategory. This classification keeps the three subcategories mutually exclusive, so that no rider is observed under more than one indicator and the results can be aggregated without any risk of double-counting. The boundary case to watch is the speed pedelec, which is excluded from the cyclist indicator because it is formally classified as an L1e-B vehicle (and hence PTW) rather than a bicycle.

12.2 Sampling and stratification

Because riders are a sparse and uneven share of traffic, location selection follows the sparse-rider procedure of the sampling chapter (Section 6.1.3). For helmet use the sampling frame is the part of the network each subcategory actually uses, and weighting is by the subcategory's own traffic exposure.

The minimum stratification differs between subcategories (see Section 12.4). Whatever the stratification, a traffic count is required at each location and during each observation session, even where national mobility statistics exist, because it provides the exposure measure that turns the raw per-session helmet rate into a weighted estimate and is needed to compute the confidence intervals correctly. The vehicles counted are those of the **target group(s) of the session** (cyclists, motorcyclists and/or PMD riders) riding in the same direction as those being observed. As for other behavioural SPIs, the count of the target group is what defines that session's share of riding within its stratum. If the observers manage to observe every passing member of the target group, the count equals the number of observations and no separate count is needed. Where observation cannot keep pace, a manual count of the target group is made for at least twenty minutes per session, either as a ten-minute window before and after the session, or as a continuous twenty-minute window in the middle. This count is then extrapolated to the full session duration. The numbers counted and the duration of the count are always recorded together so that the rate per hour can be reconstructed without error.

12.3 Data-collection specifics

12.3.1 Direct observation as the primary method

Direct observation is the preferred data-collection method, allowing the use of cameras where appropriate. Such observations are preferably carried out from a safe place along the road, ideally at locations where riding speed is reduced relative to the speed limit, such as intersections or roundabouts. For PMDs, sites where riders would gain high speed, such as steep descents, are avoided. The most straightforward procedure is to observe one rider, code the data, and then turn to the next passing rider. When traffic flow is too high to observe every passing vehicle, the next rider to be observed is selected at random, typically by coding the next passing rider once data entry has been completed. Specific rules could be defined (and documented) when a large group of riders is passing by (e.g. the first and last rider of the group, the riders passing closest to the observers, etc.). Quality of observation always takes precedence over quantity. Observers must be positioned so that the rider's head is clearly visible, and the helmet use of riders and, where applicable, of passengers is coded separately, since the legal framework and the observed prevalence typically differ between them.

The principal challenges of helmet observation concern the ability to distinguish full from partial use, the seasonal variability of traffic (especially for motorcycles), the small samples available on some road types, the difficulty of distinguishing vehicle types, and the difficulty of judging gender and age at riding speeds. The most operationally important is the full-versus-partial-use distinction for powered two-wheelers: a rider who is wearing a helmet but has not buckled the chin-strap is, in many jurisdictions, counted as not wearing a helmet, because an unbuckled helmet offers little protection in a crash. The decision rule applied and the cases in which the chin-strap is too small to be seen from the roadside must be specified clearly in the observation protocol and documented in the metadata.

12.3.2 Observers and training

Observations are made by well-trained, independent staff under the supervision of a coordinator. Observers should receive theoretical training (a briefing on aims, variables, definitions and procedures) and practical training (exercises on the road), in line with Chapter 7. Inter-rater reliability is checked before fieldwork starts and at intervals during the campaign. Observers must have a clear, country-valid definition of "helmet", clear guidance on how to handle weather, visibility or safety problems that prevent observations from being carried out, and should be familiar with how the shared PMD devices in operation in the country appear. Where both a cyclist and a PTW or PMD indicator are measured at the same session, a rule is defined in advance for choosing the next observation subject when traffic is heavy; it is generally preferable to assign one observer per subcategory.

Particular attention should be given in training to the reliable identification of vehicle types, since this is among the most error-prone judgements made at riding speed and it determines the subcategory to which a rider is assigned and, within a subcategory, the optional vehicle-type variable. Observers should be trained, with photographs and roadside exercises, to make the distinctions that matter for one or more subcategories::

- between conventional bicycles, e-bikes and the excluded speed pedelecs
- between mopeds and motorcycles
- between a moped and a PMD where the two resemble one another
- between e-scooters and ordinary kick scooters
- between shared and personal devices.

Where a model exists in several variants of similar appearance (for example a light motorcycle and a moped, or a pedal-assisted e-bike and a conventional bicycle) the distinguishing features should be illustrated explicitly during the briefing. The decision rules for ambiguous cases should be agreed in advance, and any residual uncertainty in the classification should be documented in the metadata. Inter-rater reliability checks should cover vehicle-type coding as well as helmet coding, so that systematic misclassification between observers is detected before it biases the indicator.

12.3.3 Field procedure: sessions and conditions

One observation session lasts at least 30 minutes and ideally one hour, excluding the time required for traffic counting and the recording of environmental data. Each location may be visited for several sessions in different time intervals, or assigned at random to a single time interval. When organising fieldwork, road types and time periods are balanced so as to avoid systematic bias. For example, one should not have motorway sessions (for motorcyclists) only in the morning, urban sessions only in the afternoon, or shared-device sites only at the weekend. Sessions should take place only in good enough conditions: no heavy rain, no snow or ice, no fog, and undisturbed traffic flow. Coding may be carried out on paper, tablets or smartphones.

12.3.4 Temporal requirements

Data collection takes place during daylight and covers the whole day. Sessions may be organised across the parts of the day (morning, midday, evening) or around peak hours, in which case each location is ideally observed across all the periods. Helmet use is observed both on weekdays and on weekend days, and results are

reported separately for the two, since both the purpose and the duration of riding differ between them. Measurements are carried out at the end of spring or the beginning of autumn; in principle all months are admissible except December and January, with July and August also excluded for the cyclist and PTW indicators owing to typical holiday periods (and June in some countries). Sessions on official feast days are avoided. Countries that have difficulty reaching the minimum sample in the recommended window may extend the measurement into the summer months. Where a country has historical series, the same period of the year is retained to preserve comparability, and where two seasonal measurements are run, the minimum sample requirements apply to the combined dataset.

12.4 Specific considerations by subcategory

12.4.1 Cyclists

Population, vehicles and helmets in scope. The cyclist indicator covers the riders of bicycles, tricycles and other muscle-powered cycles, **and the riders of power-assisted bicycles (e-bikes)**. Speed pedelecs are excluded, as they are formally classified as L1e-B vehicles. As a recommended optional variable, the type of bicycle can be recorded. This covers the distinction between non-electric and electric (e-bike) cycles, and where feasible between private and shared cycles, and between professional (delivery, postal, police) and non-professional riders. The reliable distinction between a conventional bicycle, an e-bike and an excluded speed pedelec is one of the vehicle-identification skills that should be covered explicitly in observer training (Section 12.3.2). Helmets observed within this subcategory include road and MTB helmets (the typical bicycle helmets), BMX-style multisport helmets and full-face downhill helmets.

Road types. Two road types are in scope: urban roads (inside built-up areas) and rural non-motorway roads (outside built-up areas). Motorways are not relevant for cyclists.

Age and gender. Results for children (0–14 years) are always presented separately from those for adults (15 years and over). Where national legislation imposes a different age threshold for compulsory helmet use (e.g., up to age 12) an additional category should be added so that results align with the legal threshold, with the rule documented in the metadata. For adult cyclists, results are also reported by gender (men and women), with the option of a “not determined” category where gender cannot be coded reliably.

Table 12-1. Trendline minimum requirements for the cyclist indicator.

Element	Requirement for the cyclist indicator
Total observations (overall)	≥ 2,000
Observations per road type	≥ 500 urban; ≥ 250 rural
Observations per week period	≥ 350 on weekdays and ≥ 350 on weekend days
Locations per period	≥ 10 per period; ≥ 2 per crossed road-type × period combination
Observations per age group (when reported)	≥ 500 per group

12.4.2 Motorcyclists (powered two-wheelers)

Population, vehicles and helmets in scope. The PTW indicator covers mopeds (categories L1 and L2 of the UN Consolidated Resolution on the Construction of Vehicles, R.E.3, generally with a cylinder capacity below 50 cc) and motorcycles (categories L3 to L7 of R.E.3, generally 50 cc or above). As a recommended optional variable, the type of PTW is recorded, at minimum to distinguish mopeds from motorcycles. Because a moped and a light motorcycle can look very similar from the roadside, and because the moped–motorcycle boundary is also the point at which a PTW may be confused with an e-bike, this distinction is one of the vehicle-

identification skills that should be covered explicitly in observer training (Section 7.2.3). Helmets in scope include full-face, open-face, half-head and tropical helmets.

Road types. Three road types are in scope: motorways, rural non-motorway roads and urban roads. The proportion of observations in each road type should normally exceed 20 % to ensure a sufficient sample for each stratum, even if this implies disproportionate sampling. Each road type is defined in terms of its typical characteristics (speed regime, signing, number of lanes, ...) so that any deviation from the EU definitions is made explicit.

Observation on motorways requires specific procedures. Locations are typically chosen at the last intersection before an on-ramp, the first intersection after an off-ramp, after a service-area exit, or on a bridge over the motorway. Observers must have an unobstructed view of the carriageway and must remain behind a safety barrier. Where direct roadside observation is impractical, observation from a moving vehicle may be considered, provided observation safety is preserved.

Table 12-2. Trendline minimum requirements for the PTW indicator.

Element	Requirement for the motorcyclist indicator
Total observations (overall)	≥ 2,000
Observations per road type	≥ 500 (urban, rural, motorway)
Observations per week period	≥ 500 per period
Locations per stratum	≥ 10 per road type and ≥ 10 per period; ≥ 2 per crossed combination

12.4.3 PMD riders

Population, vehicles and helmets in scope. The PMD indicator covers e-scooters, electric unicycles, one-wheels, e-skateboards and the corresponding non-electric variants such as kick scooters, skateboards and skates. E-bikes are not part of this indicator. The Trendline definition of a PMD is built on the OECD/ITF Safe Micromobility classification (2020) and the Eurostat/UNECE Transport Glossary. As a recommended optional variable, the type of device is recorded so as to distinguish shared from personal devices and to identify the specific kind of device (e-scooters, electric unicycle, onewheel, e-skateboard, and their non-electric variants). The distinction between an e-scooter and other PMD devices, and between shared and personal devices, is one of the vehicle-identification skills that should be covered explicitly in observer training. Observers should familiarise themselves beforehand with the appearance for the shared-fleet devices operating in the country. Helmets observed include the BMX-style multisport helmet, typical for PMD riders, as well as the full range of cycle helmets.

Road type. Only one road type is in scope: urban roads (inside built-up areas). Other road types may be observed only if the minimum sample per ownership type and week period is respected, and any such extension is documented in the metadata.

Ownership as a stratum, and the location structure. Type of ownership (shared versus personal) is treated as a stratum in its own right, in place of the road-type stratum used for the other subcategories. A possible approach to ensure sufficient vehicles of both types covered is to split the urban network into ten observation locations, of which five are in areas where shared devices operate and five in areas without shared devices. Locations without shared devices may lie within the same city as those with shared devices but must fall outside the licensed shared-service zones. Because PMDs concentrate around specific urban hotspots, the manual fine-selection of locations gives priority to spots where higher density is expected: tourist areas; the

surroundings of train, bus and tram stations; schools; squares; playgrounds and sports centres; and other points of interest.

Age, gender and passengers. PMD riders are classified by gender (men and women, for adults) and by age group (children 0–14, adults 15+), with a “not determined” gender category available as an option and any country-specific legal age threshold added as an extra group and documented. Monitoring of passengers is not recommended, since these devices are designed for one person and the observed prevalence of passengers is low. If a country nonetheless wishes to track passengers, the minimum sample requirements apply to that subgroup. In countries where no shared-device service operates, results on helmet use among shared-device users are not produced for that country. Where these minima are not met, the deviation should be justified in the metadata. Variables, analysis and reporting

Table 12-3. Trendline minimum requirements for the PMD indicator.

Stratum	Minimum number of PMD observed
Weekdays — with shared devices	500
Weekdays — without shared devices	250
Weekend days — with shared devices	500
Weekend days — without shared devices	250
Total — with shared devices	1200
Total — without shared devices	600
Total number of PMD	2000

12.4.4 Variables to be collected

Helmet use is coded as a binary outcome for each observed person: a value of one denotes a helmet worn correctly and a value of zero denotes no helmet or a helmet worn incorrectly. This coding is applied to the rider and, where present, separately to each passenger. The minimum requirement is this binary coding.

Other variables below may be added without compromising it, on the condition that any additional variable can be observed reliably from the roadside. Such additional optional variables include:

- the type of vehicle within the subcategory (moped versus motorcycle; electric versus non-electric cycle; shared versus personal device)
- the type of helmet (full-face, open-face, half-head, tropical, multisport, downhill)
- the correctness of fastening (in particular for motorcyclists)
- the use of other protective equipment such as gloves, a jacket, trousers or shoes
- the colour and reflectivity of the helmet
- the wearing of reflective clothing
- the use of earphones (for cyclists) or of a mobile phone
- the rider’s estimated age category and gender.

12.4.5 Weighting

The weighting proceeds in the two layers of Chapter 8. *Within* a stratum, each session is weighted by the volume of the target group it represents, so that a busy session counts for more than a quiet one. *Between* strata, the stratum estimates are combined into a national figure by weighting each stratum by its share of the target

group's national traffic volume. Ideally these are the vehicle-kilometres ridden by that group, and where these are unavailable, a combination of road-length data for the relevant road type with the average hourly counts from the sessions in that stratum. Where official road-length data are themselves missing, expert estimates from the relevant administrative services should be requested, and where no traffic-volume or road-length information of any kind can be obtained, the strata are kept separate and not aggregated.

Because sampling is nested in locations, two-stage stratified (cluster) models are recommended, with the first stage being the road type (or, for PMDs, the ownership condition) and the second the time period; approximations that assume simple random sampling may be used for the point estimates only if results are weighted by traffic volume, but, as set out in Chapter 8, are not acceptable for the confidence intervals, which must reflect the nesting of observations within sessions and would otherwise be unrealistically narrow. If a country allows e-scooters to be ridden on pavements, that riding must be reflected in the exposure measure.

12.4.6 Reporting requirements

The main indicators to be reported are the percentage of riders wearing a helmet across all times and locations, reported separately for cyclists, for PTW riders and for PMD riders. For each, a point estimate and a 95% confidence interval are to be reported, both overall and for the minimum-required breakdowns:

- road type (three levels for PTWs; two for cyclists)
- ownership type (for PMD riders)
- week period (weekday and weekend)
- age group (children and adults, for cyclists and PMD riders)
- gender (for adults).

Riders and passengers are reported separately where passengers are monitored. In some countries the number of passenger observations is very low; those countries are not included in the passenger results. Specific estimates for combinations of levels, with their confidence intervals, are recommended where the sample permits, and a regional breakdown (for example by NUTS 1 region) may be added as an option, in which case the minimum requirements apply within each region and results are weighted by regional traffic volume.

As for the other behavioural SPIs, results are to be delivered to the Trendline Secretariat using the standardised data-submission template, and the three levels of data delivery described in Chapter 8 apply (see Table 23-6 in Part D).

Each delivery is accompanied by metadata, which for the helmet indicators are particularly important because comparability depends on choices invisible in the headline figures. The metadata should describe at least:

- the national legislation on helmet use for each vehicle type, including any age-specific provisions, and for cyclists the accompanying note on the legal position
- confirmation that e-bikes were counted under the cyclist indicator (and how they were distinguished from speed pedelecs and from conventional bicycles)
- the observation method (roadside observers, observers in moving vehicles, cameras, or automated computer-vision detection)
- where cameras or automated detection were used, how privacy was protected, the image source and its licensing, and how the method was validated against human-coded ground truth
- the location-selection procedure and the fieldwork period and months
- the chin-strap (full versus partial use) decision rule for PTWs
- the weighting version and confidence-interval method
- whether bootstrapping was applied to the road-length-based aggregation
- any deviation from the minimum requirements.

12.5 Minimum sampling and reporting requirements

Table 12-4 consolidates the minimum and optional requirements across the three subcategories. It mirrors the format used for the other behavioural SPIs and can be read as the helmet-specific specialisation of the common requirements set out in Chapters 6 and 7. The subcategory-specific sample-size and location minima are given in Table 12-1 to Table 12-3. The table below summarises the dimensions along which the subcategories agree and differ. As elsewhere, the minima may imply disproportionate sampling relative to traffic in smaller Member States, which is acceptable and corrected at the weighting stage; what is not acceptable is to fall below the minima without explicit justification in the metadata.

Table 12-4. Trendline minimum requirements and optional additions for the helmet indicators.

Aspect	Cyclists	Motorcyclists (PTW)	PMD riders
Vehicles in scope	Bicycles and e-bikes (speed pedals excluded)	Mopeds (L1–L2) and motorcycles (L3–L7)	E-scooters (and related devices); e-bikes excluded
Road types	Urban and rural (no motorways)	Motorways, rural and urban	Urban only
Primary stratum	Road type; age group	Road type	Ownership (shared / personal)
Time period	Weekday and weekend Daytime; late spring or early autumn	Weekday and weekend Daytime; late spring or early autumn	Weekday and weekend Daytime; late spring or early autumn
Total observations	≥ 2,000 (≥ 500 urban, ≥ 150 rural; ≥ 325 per period)	≥ 2,000 (≥ 500 per road type and per period)	≥ 2000 (≥ 500 per ownership type); ≥ 250 per cell
Locations	≥ 10 per period; ≥ 2 per crossed cell	≥ 10 per road type and per period; ≥ 2 per crossed cell	10 urban (5 shared / 5 non-shared); ≥ 2 per cell
Required breakdowns	Road type, week period Age (children / adults) Gender (adults)	Road type, week period Rider vs passenger	Ownership, week period Age (children / adults), gender (adults)
Optional additions	Bicycle type (electric vs non-electric); region; helmet type and fastening; other protective gear	Moped vs motorcycle; region; helmet type and fastening; other gear	Device sub-type; region; other gear; passenger monitoring gear

Where a country cannot meet a particular minimum, for example because the minimum traffic flow cannot be reached for a stratification level, or because no shared-device service operates, the shortfall and its reason are to be reported in the metadata and the affected indicators are the reported in disaggregated form rather than aggregated into a national figure.

12.6 Using alternative measurement methods

12.6.1 Use of cameras for observations

Cameras may be used where appropriate. The advantages are clear in terms of reliability and the possible duration of observation sessions, but the possible disadvantages must also be recognised:

- privacy and data-protection (GDPR) issues, since faces must never be recorded in identifiable form

- the risk of missing key variables, in particular distinguishing children from adults or estimating gender
- the technical challenge of placing cameras safely while maintaining adequate image quality.

The technology must be tested and validated before use, and cameras must not be installed exclusively on a single road type or in a single area, as this would introduce selection bias. Each country must comply with its national and international requirements on ethics and the protection of private life.

12.6.2 Automated detection of helmet use

The standard methodology of this chapter assumes coding by human observers, optionally assisted by cameras whose images are coded afterwards. A more far-reaching alternative is automated detection by computer vision, in which a trained algorithm identifies riders and their helmet use directly from images. This is not part of the standard Trendline methodology, and most current implementations do not yet satisfy the representativeness requirements set out above. Yet it is described here because the technology is developing rapidly. It is already the most practical option for some settings, in particular for motorcycle helmet use in low- and middle-income countries with huge numbers of mopeds on the road. Where any such method is used, its limitations should be made explicit in the metadata and its values presented as complementary to, not a substitute for, the standard observer-based indicator.

Earlier automated approaches relied on hand-crafted image features: detecting the wheels of a motorcycle, extracting the region above them as the head, and classifying it as helmeted or not. These were of low accuracy, struggled to count riders correctly, often failed to attribute helmet use to the right rider position, and degraded badly in crowded scenes. Modern deep-learning image analysis does not depend on such features: the algorithm is trained on images in which every active motorcycle (one with at least a rider) is marked with a bounding box, and each box is annotated with the number of riders, their positions and their helmet use. Validated against human coding, such models produce helmet-use assessments that do not differ significantly from observer-generated data and are markedly more efficient. They can also distinguish driver from passenger positions. This is an important capability, since helmet use differs systematically between the two and by time of day and day of week. The same set of techniques can in principle be applied to cyclists and e-scooter riders, although the published validation work to date concerns motorcycles.

Two quite different data sources can feed such an algorithm. The first is purpose-collected roadside video, coded automatically instead of by hand: this improves the efficiency of coding but offers no gain in geographical scale, since the imagery must still be gathered specifically for the survey. The second is imagery not collected for road-safety purposes to which a trained algorithm can be applied to generate helmet-use estimates across a whole city or country at low marginal cost. Examples of such image sources are crowd-sourced street-view platforms such as Mapillary, or commercial street-view and CCTV sources. Recent projects, applying a trained algorithm to street-view imagery, have already produced pilot produced street-level helmet-use maps and multi-year trend lines for several regions and countries. Because licensing terms differ between platforms, the legal conditions of the image source must be checked; some widely used street-view services do not permit the creation of derived datasets.

The decisive technical issue for the street-view route is the robustness of the algorithm when it is applied to images from sites on which it was not trained. Detection accuracy is known to fall when an algorithm trained at particular sites, camera angles and road layouts is transferred to new ones, and on street-view data essentially every image comes from a “new site”, since the camera is mounted on a moving vehicle. Robustness is therefore the prime concern, and is addressed by training on a large and varied set of sites and by fine-tuning the model to each country: a regional subset of images (in the published work, on the order of 10,000 annotated images per country) is hand-annotated to provide country-specific training labels, the model is fine-tuned on these and validated on a held-out test set, and only then is it applied to the full image set.

Apart from efficiency in the data processing, a major advantage of such methods is that, provided the image database is large enough, they can provide SPIs not just at the level of a country, but also of a region within a country or even cities. A drawback is, however, the lack of guarantee of representativity and the uneven volume of image data across years, making annual trend monitoring less accurate.

Whatever the image source, an automated estimate is credible only if it is validated against human-coded ground truth. The recommended practice is for partners to collect roadside video at a cross-section of the network's road types, hand-annotate it to obtain a reference helmet-use rate for each road type and region, and compare this with the algorithm's output on both an absolute basis (does the overall percentage agree?) and a relative basis (does the algorithm reproduce real patterns, such as the rise and fall of helmet use over the day, or a trend over years?). The validation method, the image source and its licensing, the fine-tuning and annotation effort, and the absolute and relative agreement achieved should all be documented in the metadata.

13 SPIs for distraction (by a handheld mobile device)

13.1 Introduction

13.1.1 Rationale

Driver distraction is a contributory factor of growing importance in road crashes, driven by the spread of mobile devices and especially smartphones over the past two decades. The widespread use of texting and other interactive applications has compounded the older problem of handheld phone calls, so that a driver may now be visually, manually and cognitively distracted at the same time. Using a handheld mobile device while driving takes the driver's eyes off the road, reduces attention for the driving task, and is associated with markedly increased crash risk. Because the cognitive component of distraction cannot be observed at the roadside, the use of a handheld mobile device is adopted as an observable proxy for the wider distraction problem: it is the visible behaviour that best stands in for a risk that is otherwise difficult to measure directly.

The Distraction SPI captures this behaviour and is directly relevant to policy, since handheld device use is the object of legislation and enforcement in nearly all Member States. Monitoring its prevalence provides a leading measure of how well that combination of legislation, enforcement and awareness is working, well before any effect becomes visible in crash statistics.

13.1.2 Indicator definition and theoretical population

The distraction SPI is defined as:

Percentage of drivers not using a handheld mobile device while driving.

The use of the term "device" rather than "phone" makes the indicator future-proof: it covers smart-phones, tablets, personal navigation devices and digital cameras. Although several Member States ban only handheld mobile-phone use, others have extended the prohibition to mobile electronic devices more generally. Countries participating in Trendline are therefore required to provide metadata on the applicable national regulations, so that the indicator can be read against the legal context in each country.

The theoretical population is the total of all journeys made by the targeted vehicle types over the national territory, expressed in vehicle-kilometres driven. By weighting observations by traffic volume, the percentage of drivers not using a handheld device may equivalently be interpreted as the percentage of kilometres driven without a handheld device in use, which is the policy-relevant exposure measure.

13.1.3 Position within the behavioural framework

As set out in the introduction to behavioural SPIs (Chapter 5), the distraction SPI rests on the common backbone and follows the standard observer-based procedure (optionally camera-assisted; see the automatic-detection section 13.6), including the manual traffic count. Three features give the distraction indicator its own character and are the subject of this chapter:

- the behaviour of interest is only partly visible from the roadside, so the coding scheme is built around what can reliably be seen rather than around the legal definition
- observation is confined to moving traffic, since distraction behaviour differs sharply when a vehicle is stationary
- the minimum vehicle scope is three vehicle types (cars, light and heavy goods vehicles), reported both separately and as a single aggregated indicator.

13.1.4 Target population and vehicle types

The target group is drivers of moving vehicles of the relevant types. Three vehicle types are required as a minimum: passenger cars, light goods vehicles (LGVs) and heavy goods vehicles (HGVs). HGVs replace the buses and coaches of the earlier Baseline project as a minimum-required type, because sufficient sample sizes for buses and coaches were not always feasible. As a consequence, the aggregated distraction indicator is not strictly comparable with the Baseline figure. The minimum required outputs are the vehicle-specific and the aggregated indicators for the three vehicle types together. Buses and coaches may be added as an optional fourth category, on condition that a sufficiently large sample (at least 500 observations) is collected for that category alone.

Observations are made in flowing traffic only. Distraction behaviour is markedly different when a vehicle is stationary. For instance, when waiting at traffic lights, drivers frequently pick up a device they would not handle while moving. For this reason observations of stationary drivers are excluded, even where the location lies near an intersection. The different vehicle types and their categorisation should be clearly defined and illustrated for the observers during training, since some cars and vans share the same model (a passenger car has rear side windows and rear passenger seats; the corresponding van has neither). This borderline car/LGV distinction is the most common vehicle-typing error and should be addressed explicitly in observer training (Section 7.2.3).

13.1.5 Observable distraction categories

Although the indicator refers to handheld mobile-device use, the categorisation used in practice is based on what is visibly detectable during a roadside observation. This is the central conceptual point of the chapter: the legal target (device use) is wider than the observable behaviour (a device visibly in the hand), and the methodology deliberately codes the latter so that the observation procedure is clear, uniform and reproducible across observers and countries.

As an absolute minimum, two mutually exclusive categories must be coded for each observed driver:

- (1) **a mobile device in the hand:** the driver is holding the device, whether at the ear, against the steering wheel, or anywhere else within view
- (2) **no mobile device in the hand:** the rest category.

This binary coding inevitably underestimates true handheld use, because drivers frequently conceal a device on the lap or below the dashboard where it cannot be seen. This underestimation must be accepted as the price of a uniform, reproducible procedure; the consistent direction of the bias is itself useful for trend comparison. A more granular optional categorisation, following the FERSI recommendation (Vollrath et al., 2019), distinguishes handheld phoning (the device held at the ear or in front of the mouth), handheld texting or keying, handheld reading or watching without active operation, the operation of another mobile electronic device in the hand, and the rest category. Whatever level of detail is collected, it must always be possible to derive the binary minimum (device in hand versus not) from the data.

Coding is therefore binary at minimum: a value of one denotes a driver **not** using a handheld device (the compliant, “positive” outcome for the indicator) and a value of zero denotes a driver using a handheld device. Optional sub-categories such as calling, texting or keying, other manual interaction with a device may be recorded in addition, provided the binary minimum remains derivable.

13.2 Sampling and stratification

13.2.1 General principles and road types

The general logic of stratifying, selecting and sampling observation locations is set out in Chapter 6 and applies in full to the distraction SPI, including the three-step location selection procedure and the documentation of any substitutions in the metadata. The minimum stratification for distraction is by road type, with three categories required: motorways, rural roads and urban roads. Results are presented separately for the three road types and the three vehicle types and aggregated, after weighting, for the whole network. Where a Member State traditionally uses a different categorisation, an attempt is made to infer the three Trendline categories, and any deviation is documented in the methodology.

A distinctive feature of distraction relative to the other behavioural SPIs is its treatment of **time period**. Observations should only take place during daylight, with no weekday/weekend distinction. The Trendline minimum temporal requirement is therefore to plan sessions at mixed time intervals during daylight on normal working days, with the mix of moments balanced across the three road types to avoid systematic bias. As an option, weekday and weekend-day periods may be distinguished. In that case the minimum numbers of observations and locations apply per period. A finer optional split (weekday peak, weekday off-peak, weekend day, following the FERSI recommendation) may also be used.

13.2.2 Selection of locations

Locations are selected by the standard three-step procedure: the required number of locations is determined; a geographically well-spread set of points is drawn at random on the map of the area, without yet regard to the detailed site requirements; and the operational site is then chosen manually in the vicinity of each random point, on the basis of the road type, inclusion and exclusion criteria, and practical considerations, with the help of tools such as Google Street View. A GIS package (for example ArcGIS) is convenient for the random draw; where none is available, Google Maps, Google Earth or OpenStreetMap may be used. Bigger countries may, as a prior step, select one or more regions considered representative of the country with regard to distracted driving, in order to improve feasibility. Such a choice should be explained in the metadata.

Two site requirements are specific to distraction. First, because observation is confined to moving traffic, sites should permit free-flowing traffic and ideally not be placed directly in front of traffic lights. Observation may take place near an intersection, but only drivers who are actually driving are observed, never stationary ones. If the traffic flow at a selected site is disturbed (by road works or a crash), the observer moves to a new site on the same lane or nearby within the same road category. Complex road and traffic situations demanding the driver's full attention are best avoided. Second, the observer must have a good, safe and inconspicuous view of passing traffic; the practical demands are set out in Section 6.2.2). The basic characteristics of each location are to be recorded in the metadata, such as a unique identifier, the road type, the coordinates, the number of lanes, the target lane(s) and direction, and the visibility.

13.2.3 Obtaining a representative sample

On-road observation studies must provide a representative sample of all traffic in the study area, and representativeness is secured by combining a random draw with a relevance filter, exactly as for the other behavioural SPIs. A random selection of locations will inevitably include some low-volume sites; where the traffic flow is too low. In the selection process it is acceptable to exclude sites with ten or less relevant vehicles per hour. Because the three road types must all be covered, the sample is allocated to them so that each carries an adequate number of observations even where this implies disproportionate sampling. Where national traffic-volume data by road type are available and each type carries at least 20% of traffic, a minimum of ten locations per road type is used. The disproportion this introduces is corrected at the analysis stage by

weighting each stratum by its share of traffic volume (see Section 8.2) so that the national figure is representative even though the location sample is not proportional.

13.2.4 Minimum sample size and locations

The Trendline minimum requirements mirror those of the other observer-based behavioural SPIs. The absolute minimum is 2,000 valid observations in total, for the three required vehicle types combined in case only an aggregated indicator will be reported. In such cases; no separate minimum per vehicle type is set. Member States wishing to publish disaggregated results by vehicle type are encouraged to collect at least 500 observations per vehicle type.

Each road type must contribute at least 500 observations, and at least ten different locations are required per road type, giving a minimum of thirty locations in total (the motorway requirement is relaxed for countries with no or very limited motorway network). The same location may be re-used for different sessions at different times. Region is an optional further stratification dimension. Where it is used, the minimum location and driver requirements apply within each region, and results are weighted by regional traffic volume.

13.3 Data-collection specifics

Distraction observation follows the standard observer-based procedure of Chapter 7. The points below concern only the respects in which it specialises or diverges from that common procedure.

13.3.1 Driver selection and field conditions

Drivers are selected at random from all the drivers passing the site: after one observation is fully coded, the next passing driver in the target lane and direction is observed. No deliberate priority is given to LGVs or HGVs (a goods vehicle is coded only if it is the next to pass) which ensures random sampling within the traffic stream and the correct vehicle mix. Observers focus on drivers only; where the driver's hands are not visible, the case is coded as "not observable" and excluded from the distraction prevalence calculation, rather than guessed.

A traffic count accompanies each session as defined in Section 7.3.2. The count's unit is the relevant vehicle, counted separately by type where disaggregated results are envisaged.

Fieldwork should use the default window of the sampling chapter (Section 6.2.3). The one distraction-specific deviation is that observation is confined to daylight, since device use cannot be seen reliably in an unlit vehicle. Sessions are run only in good enough conditions: good weather (no heavy rain, snow or storm), good visibility (no darkness or fog) and good road conditions (no ice), with flowing, undisturbed traffic. Coding may be done on paper, tablets or smartphones. Studies are not carried out when traffic is far from normal (for example during a severe lockdown).

13.3.2 Observation siting and line of sight

General constraint. A clear line of sight into the vehicle is the binding practical constraint for distraction, more so than for the other behavioural SPIs, because the behaviour is small, fast and easily hidden. Observing the drivers of HGVs is harder than observing car drivers, because of the high seat position and windows; observers therefore need a sufficiently high vantage point such as a safe and stable raised platform, or simply a taller observer. The same constraint shapes camera placement in the automated method.

Urban and rural roads. Observation is carried out from a safe place along the road, in accordance with the applicable road-safety regulations and as inconspicuously as possible, so as to avoid that drivers lay down their phone when noticing the observers and hence influence the measurement.

Motorways and other high-speed roads. Observation is possible from locations along the carriageway that observers can reach safely and from which they can stand behind a safety barrier to observe undisturbed

traffic. Typically this concerns rest or parking areas; such sites are usable at least for the nearest lane and for slower vehicles such as HGVs, while lanes further away or vehicles at high speed are harder to observe.

Moving-vehicle observation. A complementary or alternative method on motorways is observation from a moving vehicle in real traffic, with a driver and an observer on the back seat. This approach allows vehicles to be observed on different lanes and at different speeds. With this method the geographical unit is a section (from point A to point B) rather than a single location, the section reflecting as far as possible the minimum 30-minute observation duration. The lanes and speeds of the observation vehicle should be varied systematically within a session so that the observations are representative. The traffic count for such a section is taken by stopping at a safe location along it. Moving observation can be combined with stationary observation of the right lane. A vehicle with a high seat position, such as a bus or coach, is preferred for the view of HGV drivers and gives the best view generally. Observation from overpasses or bridges is a further option, provided the vantage point is good and not so high as to make observers conspicuous.

13.3.3 Observation challenges specific to distraction

Several challenges are specific to distraction and shape both the coding rules and the realistic level of detail:

- Some forms of device use are largely invisible from the roadside. For example, a driver typing on a phone resting in the lap may show no hand-to-ear movement and appear fully attentive. Such behaviours are a source of systematic underestimation.
- Fast traffic shortens the observation window for each vehicle.
- Nighttime observation is severely degraded, since the hands and the device are not visible in an unlit cab. This is the main reason why in Trendline observation is confined to daylight.
- There is also a real risk of confusion between device use and other one-handed activities such as eating, smoking or drinking. Explicit decision rules for these cases must be set out in the protocol and taught in training, so that observers neither record a device where there is none nor miss one that is present.

13.4 Variables, analysis and reporting

13.4.1 Required and optional variables

Session information to be stored includes:

- the date and the start and end time of the session
- the unique location and session codes
- the durations of the observation and traffic-counting periods
- the relevant traffic counts.
- the road type
- the number of lanes
- the observation lanes and direction
- the time-period category
- weather and road conditions
- the region (optional)
- observer code (optional)

Distraction is coded as a binary outcome at driver level:

- a value of one denotes a driver not using a handheld device
- value of zero denotes a driver using a handheld device

In addition the vehicle type (passenger car, LGV, HGV) needs to be coded for every driver observed.

Optional driver-related variables that can also be coded if visibility and speed of the vehicle allow it:

- private versus professional driver

- type of activity (calling, texting or keying, other manual interaction, ...)
- activity other than holding a mobile device (smoking, eating, make-up, manipulating dashboard, ...)
- estimated age category (e.g. young 18–24, medium 25–65, older over 65)
- gender
- presence of passengers in the vehicle

13.4.2 Reporting and data delivery

For each reported indicator a point estimate and a 95% confidence interval are to be given, together with the unweighted number of drivers on which the value is based. The main indicator is the percentage of drivers not using a handheld mobile device across all daytimes and road types, computed over the three required vehicle types combined; however, reporting the SPI estimate for each of the three vehicle types separately is recommended, in particular for those vehicle types where the number of observations exceed 500. If optional vehicle types (buses, coaches, and so on) are observed, they are excluded from the aggregated indicator. Indicator values are also required for each of the three road types. Additional estimates may be provided, where sample sizes allow, for each of the optional variables as well as for combinations of these.

Results can be delivered at three levels (see Table 23-6 in Part D).. For distraction the minimum level is the headline percentage and the three vehicle-type and three road-type breakdowns with their confidence intervals. Each delivery should be accompanied by metadata describing at least:

- the study design and the vehicle types considered
- the location-selection method and rationale (sampling method, inclusion and exclusion criteria, minimum traffic flow)
- the fieldwork procedure (session planning, recording method, prerequisites for a session, days and hours, crossed designs, variables collected, months)
- the statistical techniques used for weighting, confidence intervals and analysis, including whether bootstrapping was applied to any road-length-based aggregation
- the national legislation on mobile-phone or device use
- (if cameras were used) the technical and accuracy information set out in Section 13.6..

13.5 Minimum sampling and reporting requirements

Table 13-1 consolidates the minimum and optional requirements for the distraction indicator. It mirrors the format used for the other behavioural SPIs and can be read as the distraction-specific specialisation of the common requirements set out in Chapters 5 and 6. As elsewhere, the minima may imply disproportionate sampling relative to traffic in smaller Member States, which is acceptable and corrected at the weighting stage; falling below the minima requires explicit justification in the metadata.

Table 13-1. Trendline minimum requirements and optional additions for the distraction indicator.

Aspect	Minimum requirement	Optional
Indicator	% of drivers not using a handheld device for each of the 3 vehicles Aggregated over the 3 vehicle types, with CI Same indicators per road type (3), with CI	Different distraction types (calling, texting, other) Breakdowns by vehicle type, time period, region, age, gender, private/professional
Method	Direct observation by trained observers, along the road or from a moving vehicle	Roadside cameras / automatic detection (pilot study first; same minimum requirements)
Vehicle types	Passenger cars, LGVs and HGVs	Disaggregation by vehicle type (2,000 per type) Buses/coaches ($\geq 2,000$ for that type)
Road types	Motorways (expressways merged in) Rural roads Urban roads (reported separately)	Expressways reported separately Region (NUTS 1) stratification
Time period	Mixed daytime intervals on weekdays, balanced across road types Not during holidays or hard winter	Weekday vs weekend (recommended where feasible); FERSI-recommended split: weekday-peak / weekday-off-peak / weekend
Sample size	$\geq 2,000$ observations in total (3 types combined) ≥ 500 observations per reported vehicle type ≥ 500 observations per road type ≥ 10 locations per road type (≥ 30 total)	$\geq 2,000$ observations for passenger cars ≥ 500 observations for additional vehicle types ≥ 2 locations and ≥ 500 observations per crossed stratum Larger samples for regional / crossed indicators
Conditions	Moving traffic only (stationary excluded) Good weather, visibility and road condition; daylight ≥ 10 relevant vehicles/hour; clear line of sight	Automatic traffic counter alongside the manual count

Where a country cannot meet a particular minimum (e.g., because a road type does not exist, or because a stratification level cannot reach its threshold) the shortfall and its reason are reported in the metadata, and the affected indicators are reported in disaggregated form rather than aggregated into a national figure.

13.6 Automatic detection via roadside cameras

13.6.1 Camera types and the recommended pilot

Observation can also be done by using cameras, either traditional cameras of which the images are analysed by researchers, or by smart cameras, in which images of passing vehicles are captured and the presence of a handheld device is determined from the image. The automated approach offers reliability, longer effective measurement periods and potentially nighttime use, but the technology is relatively new and must be tested and validated before use. It is also more challenging to find a representative set of locations. It is described here as an accepted alternative to direct observation rather than as part of the standard procedure.

The generic conditions for camera-based observation (local validation, GDPR compliance and the masking of faces, coverage of more than one road type, and documentation of detection accuracy) are the same as those mentioned in Chapter 7 (fieldwork). The optical and siting requirements are given in Appendix E. This section keeps only the points specific to distraction:

Cameras may be fixed (unmanned, installed for weeks or months) or mobile (manned, easily moved between locations), and they can vary in “intelligence”. A *not-intelligent* camera photographs every passing vehicle for manual checking; a *partly intelligent* camera flags presumed device users for manual confirmation; and a *fully*

intelligent camera identifies device use fully automatically. Experience with such cameras for research, as opposed to enforcement, remains scarce, and so far mostly mobile and not- or partly-intelligent systems have been used. Because of this, the method should first run a pilot study to evaluate the national, regional and local regulations that apply; the GDPR constraints; the camera's false-positive and false-negative rates; the technical feasibility, including positioning to observe both low vehicles (cars, vans) and higher vehicles (trucks); and the labour required to confirm images. Privacy protection is essential: faces must not be captured in identifiable form, and where a passenger is detected, masking of that part of the image is recommended.

13.6.2 Practical and technical considerations

There are still several technical challenges that need to be addressed when using cameras:

- A single camera position can rarely serve all vehicle types equally well. The steeper and more downward the viewing angle, the deeper the view. The viewing angle must be set to see into the vehicle and must be changed to observe lower vehicles versus higher vehicles.
- Windscreen glare and poor light can make detection difficult. Polarising filters and infrared illumination can help addressing this problem.
- Not every camera can be placed on every road type.
- Cameras designed for deployment on overpasses constrain the random location sampling and may not be suited for urban and rural roads.

As with human observation, drivers who hold a device low on the lap or against the door are hard to detect, so the method shares the same downward bias as direct observation.

13.6.3 Reporting and documentation

The minimum dataset must distinguish device users from non-users, record the session variables in which observations are nested (location, road type, date, start and end time), and record the count of all passing relevant vehicles is needed for weighting.

Practical experience reported within the Trendline programme that illustrated that the detection of the use of a device is harder than detection of vehicle type, that an explicit treatment of "unclear" images and an inter-rater reliability check are important, and that the false-positive and false-negative behaviour of the method should be documented and reflected in the metadata, so that camera-based and observer-based values can be compared with appropriate caution.

When using cameras the following should be documented:

- detailed information on the technical characteristics of the camera
- information on the sampling and selection process for locations
- the vehicle types included
- the camera accuracy (false-positive and false-negative ratios)
- the data-collection and coding procedures
- the data quality and correction procedures,
- the weighting and analysis.

14 SPIs for driving under the influence of alcohol or drugs

14.1 Introduction

14.1.1 Rationale

Driving while impaired by alcohol is one of the most consistently documented risk factors in road safety. The relative crash risk rises sharply with blood alcohol concentration (BAC): the DRUID studies place it at roughly double the sober risk at 0.5–0.8 g/l, five to thirty times higher between 0.8 and 1.2 g/l, and more than twenty times higher above 1.2 g/l. Monitoring the prevalence of drink-driving among drivers in normal traffic is therefore essential, both for tracking the safety of the road-transport system and for assessing the effect of enforcement, legislation and awareness-raising.

Driving while impaired by illicit drugs was an experimental indicator in the first phase of Trendline, introduced because the available evidence consistently identifies illicit-substance use as a significant contributor to crashes, and especially to severe ones. The meta-analyses (SWOV, 2020) indicate that amphetamines multiply the risk of a fatal crash roughly fivefold and of an injury crash roughly sixfold; that cocaine roughly triples the risk of a fatal crash; that opiates roughly double it; and that the combined use of several drugs, or of drugs together with alcohol, raises the risk by an order of magnitude or more. Cannabis taken alone shows a more moderate effect with odds ratios of about 1.3 to 2.5 across studies but is by far the most frequently detected substance in roadside testing, which keeps it relevant from a public-health perspective even at its lower per-use risk.

Because alcohol and drug impairment overlap in their behaviour, in their risk and in their combination, the two indicators are treated together in this chapter. Their methodologies are largely common, and the guidelines explicitly recommend that, wherever a country measures both, it does so with a single, shared instrument.

14.1.2 Indicator definition and theoretical population

The **Alcohol SPI** is defined as:

Percentage of drivers driving within the legal limit for blood alcohol concentration (BAC).

The legal limit is country-specific, ranging from 0.0 to 0.8 g/l across Europe, with separate and generally lower limits for novice and professional drivers in many Member States. Because the limit varies, the headline indicator is not directly comparable between countries. For this reason, the methodology calls for complementary indicators anchored in fixed BAC thresholds (Section 14.7.1).

The **Drugs SPI** is defined as:

Percentage of drivers not driving under the influence of (illicit) drugs.

Its methodology is closely aligned with that of the Alcohol SPI but it differs in three material respects: the testing medium is oral fluid (saliva) rather than breath; the on-site result is essentially binary, a target substance being either detected or not, rather than a continuous concentration; and a third measurement option, random saliva testing carried out by researchers without police involvement, is admitted alongside police-led testing and self-report. Moreover, two additional indicators are recommended that are subsets of the group of drivers using drugs:

- *Percentage of multidrug-users while driving*

- *Percentage of drivers with combined alcohol and drug-use.*

For both DUI indicators, weighting the observations by traffic volume across road types and time periods allows the result to be read as an exposure measure: the share of vehicle-kilometres driven within the legal limit (for alcohol) or free of illicit drugs (for drugs) across the national network.

14.1.3 Position within the behavioural framework).

As set out in Chapter 5 (the introduction to behavioural SPIs) the DUI indicators rest on the common backbone. Three features give them a character of their own and shape the whole chapter:

- **They require physical interaction with the driver.** Unlike the passive roadside observation used for speed, restraint, helmet and distraction, an objective DUI measurement requires the vehicle to be stopped and the driver to provide a breath or saliva sample. In nearly all Member States only the police hold the legal power to stop drivers. Hence, objective roadside testing needs to be conducted within an official police enforcement operation, and researchers never stop vehicles themselves. Researchers may assist and document an operation (in particular in defining the distribution of measurement sites), but must never attempt to stop vehicles.
- **Self-report is a standard method, not merely a supplement.** Because random roadside testing is legally impossible in some countries where the police may stop drivers only on reasonable suspicion, self-reported behaviour is a recognised, standard route to the indicator rather than a fallback of last resort. It is the designated method for that legal context, with its own sampling and questioning rules (Section 14.3). The self-report route is not a degraded version of the physical test but a different and fully legitimate measurement for the countries that must use it. Its results are less accurate but considerably cheaper to obtain, and for some Member States it is the only realistic option.
- **For drugs, a third method is currently admitted.** Where police-led testing is not feasible, the Drugs SPI may also be measured by researcher-led saliva testing without the police (Section 14.4), an option not offered for alcohol.

Because the DUI SPIs can be produced by more than one method, the no-mixing rule mentioned in Section 5.3 applies with particular force here: police-led testing, period-prevalence and trip-based surveys should be reported as separate series.

A direct consequence of the first two features is that the indicators are produced by more than one method, and that SPI results derived from different methods should never be mixed. Police-led testing measures observed point prevalence; period-prevalence surveys measure self-reported behaviour over a recall window; trip-based surveys measure self-reported behaviour for a single recent journey. Each describes something different and carries different biases, so each is reported separately, with the underlying method stated.

14.1.4 Target groups

The minimum scope of the DUI indicators is the same as for the Speed SPI: drivers of passenger cars, defined according to the CARE conventions. Goods vehicles, buses and coaches, and motorcycles are optional. When included these vehicles should be reported separately. It should be understood that including additional categories requires substantially larger samples (and hence financial resources) to yield meaningful results. In countries with separate legal limits for novice or professional drivers, separate indicators for those groups are recommended. Where more than one vehicle type is tested or surveyed, vehicle type is recorded as a variable in the dataset, with the categories clearly defined for the fieldwork and in the metadata.

In self-report surveys, a respondent is treated as a car driver if they report driving a car at least a few days a month (which is the standard ESRA filter) and the same logic applies to any other vehicle type surveyed.

14.1.5 Acceptable measurement methods

The methods recognised for each indicator form a ladder ordered by accuracy; choices depend on what is legally and practically feasible. The justification for the choices is recorded in the metadata.

For the **Alcohol SPI**, two methods are recognised. The strongly preferred method is **Random Breath Testing (RBT)**. This is the so-called 'golden standard' for the Alcohol SPI, consisting of roadside breath testing of randomly selected drivers, conducted with the police on a representative sample of locations. Because every tested driver yields a direct, objective BAC measurement, RBT gives the most accurate picture of drink-driving in normal traffic. Where RBT is impossible because the police is not legally allowed to stop a car without suspicion, **self-reported behaviour surveys** provide the recognised alternative, in either of the two designs described in Section 14.3.

For the **Drugs SPI**, three methods are recognised, ranked from most to least accurate. The recommended method, or "gold standard", is **random roadside saliva testing by the police**, frequently carried out together with breath testing for alcohol. Where police-led testing is not feasible, because the law does not permit stopping drivers without suspicion, or because police resources are insufficient, an option to be considered is **random saliva testing without the police**, in which researchers approach drivers at locations where vehicles are routinely stationary (Section 14.4). When neither form of random saliva testing is feasible, the minimum acceptable method is again **self-reported behaviour**.

14.2 Roadside testing by the police

The procedures for sampling of locations and drivers, the minimum sample sizes and the fieldwork rules are common to RBT for alcohol and to police-led saliva testing for drugs. Indeed the two are often run as a single operation, a stopped driver being tested for both at once. They are therefore described together here, with the drug-specific aspects of saliva testing gathered in Section 14.2.7 All of this follows the observer-based fieldwork logic of Chapter 7, adapted to the fact that the vehicle is stopped.

14.2.1 Sampling of locations and sessions

The first phase of the design concerns the choice of control sessions: the combinations of a location and a time period at which a roadside check is set up. Location selection follows the same three-step random procedure used for the other behavioural indicators. First the required number of locations is determined by selecting points on a national or regional map, with adequate geographical spread, typically using GIS software. Second, the operational location is then chosen manually in the vicinity of each random point, taking the road type and the practical requirements of a check into account (see below). Larger countries may pre-select one or more representative regions before drawing locations, in order to keep the fieldwork manageable; this should be documented in the metadata.

The final selection respects two practical constraints: drivers must be able to be stopped safely, and the police, the research team and any tested driver must be able to park without disrupting traffic. On motorways and expressways, checks can be organised at exits or entrances, or drivers transferred to nearby rest areas; this should be documented in the metadata. Locations near bars, discotheques or other places of known high prevalence are deliberately avoided, as such proximity would bias the sample upwards. Where random selection yields several locations on the same road, one is reselected to preserve geographical diversity, and locations with very low traffic flow (fewer than ten vehicles per hour) can be excluded.

Each location is then assigned a **week period** (weekday daytime, weekday night, weekend daytime or weekend night) so that the four periods are balanced across the three road types and across the country. Allocation strictly proportional to traffic volume is not appropriate here: nighttime volumes are low while prevalence at night is high, so strict proportionality would place the widest confidence intervals on precisely the strata that

matter most. Nighttime periods are therefore deliberately oversampled, and the resulting disproportionality is corrected by stratification weighting (Chapter 8).

The minimum requirements are at least 10 different locations on each of the three road types; at least 10 different locations for each of the four week periods; and at least 2 different locations for each of the 12 crossed road-type-by-week-period strata. The minimum total is therefore 30 distinct locations, and a location may be re-used for sessions in different week periods. Member States with no, or only a marginal, motorway network are exempt from the motorway requirement.

14.2.2 Random selection of drivers and the role of the police

The second phase concerns the selection of drivers within each session. Drivers must be sampled at random and irrespective of any suspicion of impairment; any selectivity in the choice of which drivers to stop biases the sample and undermines the SPI. Voluntary participation cannot be accepted, as self-selection bias is then unavoidable. Where the law denies even the police the power of random stopping, RBT and police-led saliva testing cannot be used, and the Member State turns to the methods of Section 14.3 or Section 14.4 (drugs only).

Random testing requires the police to depart from their normal practice of selective enforcement. It is therefore strongly recommended that a researcher attend every session to assist with coding data and to verify that selection is genuinely random with correct location and time period, and without targeting of “suspicious” drivers. Where a researcher cannot attend, the data are screened afterwards for session-level signs of selectivity, such as outliers in the proportion of drivers over the limit. The ideal field rule is that every driver passing during the session is tested; where traffic volume or police capacity prevents this, the next driver to arrive after each completed test is stopped.

14.2.3 Temporal considerations

Data collection is ideally carried out in late spring or early autumn. In practice all months are admissible except December, January, July and August. Bank and school holidays and hard winter conditions are to be avoided, as they disturb normal traffic and mobility patterns. Where a country holds a historical series, the same period of the year is retained for comparability. Fieldwork can be spread over two waves, for example one in spring and one in autumn, provided the minimum sample requirements are met by the combined dataset, which then yields the overall and disaggregated indicators.

Although the typical duration of observation sessions for behavioural indicators are 60 minutes, for DUI sessions one may need shorter periods of 30 minutes (or even 20 minutes) because of social media and specific applications announcing the controls. This in turn, however, makes it necessary to have more observation sites and/or observation moments, in order to arrive at the minimum sample.

The COVID-19 experience underlined that collection should not be planned while traffic or social life is heavily distorted by emergency measures such as lockdowns, curfews or the closure of bars and restaurants. Since these circumstances bear directly on the risk factors for impaired driving; a representative context is a precondition for a representative SPI. These rules apply equally to the testing methods and to the self-report surveys.

14.2.4 Minimum driver sample

At national level, the minimum required sample for the Alcohol SPI and for the police-led Drugs SPI is 2,000 tested car drivers. Each road type and each week period should reach at least 500 drivers. For “difficult strata” (typically night and/or weekend periods for DUI alcohol and most strata for DUI drugs) the minimum can be relaxed to 250 drivers per stratum. This relaxation should be acknowledged in reporting: indicators based on 250–499 drivers are flagged as small samples, and those based on fewer than 250 as deviations from the minimum requirements. The thresholds refer to the number of valid datapoints in each stratum.

Reporting the 95% confidence interval alongside each point estimate is essential. Countries with the resources to do so are encouraged to oversample the most problematic strata in order to gain statistical power for sub-group analysis.

14.2.5 Fieldwork organisation and the roadside procedure

Each session lasts at least 30 minutes, and 30 to 60 minutes is recommended. It should be known that the longer a check remains in place, the greater the risk that drivers learn of it through real-time alerts on social media and re-route around it. The police use their standard, type-approved and calibrated devices (breath analysers for alcohol and oral-fluid screening devices for drugs) and apply the normal sanctioning procedure on a positive result. There are no weather, visibility or road-quality prerequisites beyond practical feasibility: safe stopping, sufficient parking for all parties, and no obvious escape route for drivers.

The roadside procedure itself is straightforward. The police officers stop a vehicle; for alcohol a passive “sniffer” device may first be used to detect alcohol in the cabin air, with a breath-analyser test following where alcohol is indicated, while for drugs an oral-fluid screening test is administered. The officer, or an assisting colleague, records the result together with the time and location of the stop and, where collected, the driver and trip variables. The appropriate enforcement measures follow a positive test, after which the next vehicle is stopped and the procedure repeated. Coding may be on paper or, preferably, on tablets or smartphones running a dedicated application, which records date, time and location automatically and reduces transcription error. Any such tool should be tested for usability and speed before adoption.

For analysis purposes, it is useful to collect additional driver or trip variables such as age, gender, novice or professional status, place of departure, trip motive, presence of passengers, etc. Questions are best asked before the test, kept few in number and, where possible, accompanied by a small token of appreciation, so as to limit non-response once the test result is known. A fuller, structured questionnaire on behaviour, legal knowledge, risk perception and preferred alternatives may also be administered within a police-led operation; the complete inventory is extensive and is best held as a maximalist protocol outside the chapter body.

14.2.6 Traffic counting

As for the other observer-based SPIs, a traffic count is made at each location and session because it underpins the within-session weighting and the confidence intervals (Chapter 8). Ideally every passing vehicle, including those stopped, is counted for the whole session by an additional police officer or a researcher; otherwise a count is made over ten minutes before and ten minutes after it. Both the count and its duration are always recorded together, so that vehicles per hour can be computed without ambiguity. The counting time does not form part of the minimum 30-minute session unless it runs in parallel with the testing. Where more than one vehicle type is in scope, vehicles are counted by category, allowing vehicle-specific weights to be derived.

14.2.7 Saliva testing in practice

Roadside saliva tests are screening devices: they indicate whether one or more target substances are present above the device threshold. However, they do not, in general, return a concentration. The on-site result is therefore essentially binary, and in most jurisdictions a positive screen corresponds to infringement of the legal limit. Establishing a concentration would require confirmatory blood analysis or advanced laboratory testing, which is usually too resource-intensive to apply to every driver in an SPI campaign. The police may, optionally, run blood tests on (a sub-sample of) positive screens. When such blood tests are conducted, the resulting percentages can be entered into the template for the raw data.

A saliva test can often return results on several substances at once. This is policy-relevant, since combinations of drugs and the combination of drugs with alcohol, carry far higher crash risks than any single substance. Multi-substance results should therefore be captured as a separate variable, and indicators are recommended separately for “single drug detected” and “multiple drugs detected”, as well as “drug(s) and alcohol detected”.

The panel of detectable substances varies between the devices that are currently used by the police in different countries. Therefore, the exact device, the substances it can detect and any cross-reactivity should be documented in the metadata.

It should be noted that saliva analysis is slower than breath testing (several minutes per driver is common). This limits session throughput and requires extra police capacity at busy locations, or for arranging the two tests so that the slower does not become the bottleneck.

14.3 Self-reported surveys

14.3.1 Sampling and survey mode

Self-report surveys must rest on a random sample that is representative of the national population aged 18 and over (or, where a Member State prefers, of its driver population). Random recruitment from an established online research panel is acceptable; convenience samples are not. The recommended mode is the anonymous online questionnaire, whose speed, low cost, anonymity, absence of interviewer effects and reduced social-desirability bias outweigh the alternatives for a topic as sensitive as impaired driving.

It is recommended to follow the sampling and weighting approach adopted in ESRA: interlaced quotas for gender crossed with age, alongside a non-interlaced regional quota based on national population statistics, with any residual disproportionality corrected by post-stratification weighting (Chapter 8). The SPI values are computed on the sub-sample of respondents who qualify as car drivers.

The self-report designs follow the sampling, mode and minimum-size rules that were given in Chapter 9 on using self-reported-data chapter: for period-prevalence designs a sample of at least 750 car drivers (1,000 recommended) and for trip-based designs a minimum sample of 2,500 car drivers.

When both alcohol and drugs are measured by self-report, using a single shared instrument is strongly recommended: it yields consistent samples, supports analysis of the combined use of alcohol and drugs, and reduces fieldwork cost.

14.3.2 Question wording for DUI alcohol

For the **period-prevalence** design, the minimum item is the standard ESRA3 question “Over the last 30 days, how often the respondent drove a car when he/she may have been over the legal drink-driving limit” with a five-point answering scale from “never” to “(almost) always”. The SPI is the percentage answering “never”. Respondents who report any higher frequency are routed to a follow-up recording whether this occurred in the week or at the weekend, by day or at night, and on motorways, urban or rural roads, which supports disaggregated period-prevalence indicators by week period and road type.

For the **trip-based** design, the respondent is asked whether they think they were driving over the legal alcohol limit during the sampled trip. The SPI is the percentage answering “no”. Countries may add further questions on the trip such as duration, distance, passengers, and motive. They are encouraged to include the period-prevalence question as well, so the that both forms of indicator can be reported.

14.3.3 Question wording for DUI drugs

The drug self-report design mirrors its alcohol counterpart in sampling, minimum sizes and survey mode, differing mainly in the wording and in the structure of the follow-ups. For the **period-prevalence** design, the minimum item is the ESRA3 drug-driving question: “Over the last 30 days, how often the respondent drove within one hour of taking drugs other than prescribed or over-the-counter medication” on the same five-point scale. The indicator is again the percentage answering “never”. Those reporting any higher frequency are routed to two follow-ups: one recording the week period(s) concerned, and one recording the substance(s)

used: cannabis, cocaine, amphetamines, illicit opiates or other. Both questions accept multiple answers. ESRA3 does not at present carry a road-type follow-up for drugs analogous to the alcohol one; a Member State wishing to report a road-type breakdown can add the question nationally, ideally by adapting the existing alcohol item (the reporting template for drugs offers the opportunity to report road type and substance).

For the **trip-based** design, the respondent is asked whether they think they were driving within one hour of having used such drugs during the sampled trip. The SPI is the percentage answering “no”, an affirmative answer is strongly recommended to trigger the same substance-identification follow-up. As for alcohol, a Member State using the trip-based design is encouraged to include the period-prevalence question as well.

14.4 Random saliva testing without police involvement

Where police involvement is not possible, the Drugs SPI may be measured by researchers carrying out random saliva testing themselves, at locations where drivers are routinely stationary for a few minutes, in particular petrol stations. A driver is approached, briefly informed about the study and invited to take part in a short, anonymous procedure consisting of a brief private questionnaire and a saliva test, after which the journey continues. A quota sample by age group and gender is used to approach a balanced national picture, and the driver may also be invited to complete the self-report survey of Section 14.3 (or part of it). If the driver has time to wait, he or she can also be given the result of the saliva test.

The obvious weakness of this method is participation bias: participation is voluntary, and drivers who have used drugs have a stronger incentive to decline than those who have not, so the prevalence is biased downwards. Estimates from this route should therefore be reported explicitly as minimum figures that are not directly comparable with police-led estimates from other countries.

The metadata should record the participation rate, i.e. the share of approached drivers who agreed, together with any sign of differential refusal. Aside from these caveats, the location, week-period, sample-size and traffic-counting requirements are the same as for police-led testing (Section 14.2). Because the method is comparatively novel, countries adopting it are encouraged to document their field experience fully, so that its reliability can be assessed as the evidence base grows.

14.5 Variables, analysis and reporting

14.5.1 Weighting and confidence intervals

The analytical chain follows the framework of Chapter 8 and is identical in structure for alcohol and for drugs. For the testing methods, sessions are aggregated within strata (road type × week period × vehicle type). Each observed driver carries a session weight derived from the session duration and the vehicle count. Between-strata aggregation should follow the common weighting ladder for behavioural SPIs as discussed in Section 8.2. and the confidence intervals should be calculated respecting the nested design as described in Section 8.4.

For the self-report methods, post-stratification weighting should correct any disproportionality against the national population in age × gender and region. Trip-based estimates carry an additional weight to undo the deliberate over-sampling of night trips. In every case the confidence interval must reflect the nested sampling design, as explained in Chapter 8.

14.5.2 Reporting of data on DUI Alcohol

At a minimum, the headline Alcohol SPI should be reported with its 95% confidence interval, and disaggregated indicators by road type and by week period (for RBT and for trip-based prevalence; for period-

prevalence, the headline indicator alone). SPI values by age category, by gender and by BAC level are strongly recommended, as are crossed estimates across the levels of disaggregation.

Each data delivery should include the unweighted number of drivers and of locations, and should be accompanied by the metadata, including at least:

- study type
- sampling design
- fieldwork procedure
- weighting and confidence-interval method
- the applicable BAC limits, including any specific limits for novice or professional drivers.

14.5.3 Reporting of data on DUI Drugs

Reporting parallels alcohol. Each Member State delivers at least the headline Drugs SPIs (single use, multiple use, and combined use with alcohol) with its 95% confidence interval and indicators by road type and by week period (for random saliva testing and for trip-based prevalence; for period-prevalence, the headline alone). The detected substances should also be mentioned. Strongly recommended additions include indicators by age, by gender, for private versus professional drivers and for novice drivers, together with crossed estimates where the data allow. Estimates from researcher-led testing are, as noted in Section 14.4, should be reported as minimum figures.

Metadata should include at least:

- study type
- saliva test method and detectable substances
- sampling design
- fieldwork procedure
- weighting and confidence-interval method
- any country-specific rules for professional or novice drivers
- substance-specific limits in force.

Self-report datasets should include the standard demographics, the response to the drug question (five-point scale and the dichotomised “never” / “at least once”) and, for trip-based prevalence, the full trip description, with the substance follow-up held as a set of binary variables.

14.6 Minimum sampling and reporting requirements

The two tables below summarise the minimum requirements for each indicator and method. Minimum requirements are shown in plain text; recommended options are noted where relevant. Where a cell refers to a stratification, the corresponding SPI must be delivered with its 95% confidence interval.

Table 14-1. Minimum requirements for the Alcohol SPI.

Aspect	Random Breath Testing	Self-report
Indicator	% of car drivers within the legal BAC limit (point prevalence)	Period: % "never" over the last 30 days. Trip-based: % "no" on a recent trip
Vehicle types	Passenger cars (others optional)	Car drivers (others optional)
Road types	Motorways, rural, urban — separate SPIs required	Period: by road type recommended. Trip-based: by road type required
Week periods	Four periods (weekday day/night, weekend day/night) — separate SPIs required	Period: by week period recommended. Trip-based: by week period required
Locations	≥ 10 per road type, ≥ 10 per week period, ≥ 2 per crossed stratum (≥ 30 in total)	n.a.
Sample size	≥ 2,000 car drivers; ≥ 500 per road type and per week period; ≥ 250 for difficult strata	Period: ≥ 750 (1,000 recommended). Trip-based: ≥ 2,500
Time of year	Avoid Dec–Jan and Jul–Aug; avoid holidays and severe weather	Same recommendations apply
Complementary	% drivers ≥ 0.5 g/l and by BAC band (highly recommended)	Combine trip-based and period-prevalence questions where feasible

Table 14-2. Minimum requirements for the Drugs SPI.

Aspect	Police-led saliva testing	Researcher-led saliva testing	Self-report
Indicator	% of car drivers free of illicit drugs (point prevalence)	% of car drivers free of illicit drugs (volunteer-based, minimum estimate)	Period: % "never" over 30 days. Trip-based: % "no" on a recent trip
Vehicle types	Passenger cars (others optional)	Passenger cars (others optional)	Car drivers (others optional)
Road types	Motorways, rural, urban — separate SPIs	Same as police-led	Period: recommended. Trip-based: required
Week periods	Four periods — separate SPIs	Four periods — separate SPIs	Period: recommended. Trip-based: required
Locations	≥ 10 per road type and per week period; ≥ 2 per crossed stratum (≥ 30 total)	As police-led, adapted to the petrol-station setting	n.a.
Sample size	≥ 2,000 car drivers; ≥ 500 per stratum; ≥ 250 for difficult strata	Same as police-led	Period: ≥ 750 (1,000 rec.). Trip-based: ≥ 2,500
Substance reporting	Aggregate plus separate SPIs per common group; multi-drug flag; combined use with alcohol	Same as police-led	Substance follow-up on "yes" responses
Specific caveat	Saliva analysis is slower than breath testing — plan capacity	Voluntary participation bias — report as a minimum estimate	Combine with the alcohol survey wherever possible

14.7 Beyond the legal limit: complementary indicators

14.7.1 Fixed BAC thresholds for alcohol

Because the legal BAC limit varies between countries and may change over time, the headline Alcohol SPI is not directly comparable across Europe. To make international and historical comparison more meaningful, countries using RBT are strongly recommended to report also two complementary indicators anchored in fixed thresholds:

- the percentage of drivers at or above 0.5 g/l,
- and the percentages within the bands 0.5–0.79, 0.8–1.19 and 1.2 g/l or more.

These bands track the structure of the DRUID risk evidence. As Table 14-3 shows there is a slight increase in crash risk between 0.1 and 0.5 g/l, a medium increase between 0.5 and 0.8, a high increase between 0.8 and 1.2 and an extreme increase above 1.2 g/l. The high-BAC group accounts for a disproportionate share of alcohol-related casualties, so an estimate of its prevalence is of particular policy interest. Since the modern devices used by police forces record exact BAC values, these indicators come essentially for free from the data already collected.

Table 14-3. Relative risk of being seriously injured or killed by blood alcohol concentration. Source: DRUID project (Schulze et al., 2012).

Risk level	Relative crash risk	BAC band
Slightly increased	1 – 3	$0.1 \text{ g/l} \leq \text{BAC} < 0.5 \text{ g/l}$
Medium increased	2 – 10	$0.5 \text{ g/l} \leq \text{BAC} < 0.8 \text{ g/l}$
Highly increased	5 – 30	$0.8 \text{ g/l} \leq \text{BAC} < 1.2 \text{ g/l}$
Extremely increased	20 – 200	$\text{BAC} \geq 1.2 \text{ g/l}$

14.7.2 Substance-specific and multi-substance reporting for drugs

The drug equivalent of the fixed-threshold logic is reporting by substance group and by single- versus multi-substance detection, rather than by concentration band. Substance-specific reporting is encouraged: the share of drivers positive for amphetamines, or for several substances at once, speaks more directly to crash outcomes than the headline aggregate, given the much higher relative risks attached to those categories (Table 14-4)¹. The exact device, its detectable panel and any cross-reactivity are recorded in the metadata so that comparisons can be read in context.

¹ Adapted from the SWOV factsheet on drugs and medicines (SWOV, 2020) and the Trendline Drugs Methodological Guidelines.

Table 14-4. Crash-risk increase associated with driving under the influence of selected substances.

Substance	Crash severity	Odds ratio	95% CI	Source
Amphetamines	Fatal	5.2	2.6 – 10.4	Elvik (2013)
Amphetamines	Injury	6.2	3.5 – 11.1	Elvik (2013)
Cannabis	Fatal ²	1.3	0.9 – 1.8	Elvik (2013)
Cannabis	Injury ²	1.1	0.9 – 1.4	Elvik (2013)
Cannabis	Fatal & injury	1.4 – 2.5	1.1 – 3.7	Rogeberg & Elvik (2016); Rogeberg (2019)
Cocaine	Fatal	3.0	1.2 – 7.4	Elvik (2013)
Cocaine	Injury	1.7	0.9 – 3.0	Elvik (2013)
Opiates	Fatal	1.7	1.0 – 2.8	Elvik (2013)
Opiates	Injury	1.9	1.5 – 2.4	Elvik (2013)
Multiple drugs	All crashes	5 – 30	—	Hels et al. (2011)
Alcohol + drugs	All crashes	20 – 200	—	Hels et al. (2011)

² Not statistically significant

15 SPIs for compliance at pedestrian crossings and intersections

15.1 Introduction

Compliance with traffic rules is one of the oldest and most directly evidence-based candidates for a safety performance indicator. The European Transport Safety Council recommended the indicator “failure to stop or give way at junctions or pedestrian crossings” in its first SPI report (ETSC, 2001) more than two decades ago, and successive Performance Index (PIN) reports have shown that progress towards the EU road-safety targets has been fastest in the countries that prioritised compliance with key safety rules. Analyses conducted through the pilot work (Ambros et al., 2026) found compliance rates to relate plausibly to casualty patterns. Compliance therefore qualifies, in the language of the Trendline framework (Chapter 2), as an intermediary outcome with a clear causal link to crashes, injuries and fatalities.

The SPIs are built around pedestrian safety, one of the persistent weaknesses of the European road-safety system. Pedestrians make up around 20% of all road deaths in the European Union, a proportion that has been stable for a decade; it is markedly higher among women than among men, and half of all pedestrian fatalities involve someone aged 65 or over. Most pedestrian injuries occur in urban areas, with cars accounting for more than 70% of the vehicles involved, and the great majority of crashes occur while pedestrians are crossing, frequently at pedestrian crossings - 81% of fatalities on a road segment and 11% at an intersection.

The SPI is divided into two contexts, because the relevant rules and interactions differ fundamentally:

- At **unsignalised** pedestrian crossings and intersections, compliance is governed by the priority rules between drivers and pedestrians, or among drivers themselves, and the relevant unit is the interaction between the road users involved.
- At **signalised** pedestrian crossings and intersections, the traffic signal structures the interaction and largely separates pedestrians from drivers in time; non-compliance takes the form of red-light negation: red-light running by drivers and riders, or red-light violation by pedestrians.

This manual treats both contexts in a single chapter because they share a common rationale, a common location-selection logic, common observation methods and common sampling and weighting requirements.

15.2 Indicator definition and theoretical population

15.2.1 SPIs for unsignalised contexts

Two indicators are defined for **unsignalised contexts**. For **pedestrian-driver interaction compliance** the SPI is defined as:

Percentage of pedestrian-driver interactions complying with the priority rules on pedestrian crossings, on an intersection or on a road segment.

For **driver compliance** the SPI is defined as:

Percentage of drivers complying with the priority rules at intersections that have no designated pedestrian crossing.

Because national priority rules differ, both definitions are open to adaptation and should be documented in the metadata. The unit of analysis differs between the two: the first is the pedestrian–driver interaction, the second the individual driver.

15.2.2 SPIs for signalised contexts

Two indicators are defined for **signalised contexts**. For **pedestrian compliance** the indicator is defined as::

Percentage of pedestrians complying with the traffic light signals. .

For **driver compliance** the indicator is defined as:

Percentage of drivers complying with the traffic light signals.

Both are observed independently of whether the pedestrian crossing lies on an intersection or on a road segment between intersections. Because traffic-light systems and priority rules differ between countries, both definitions are open to adaptation to national conditions, and the operational definition used by each Member State is documented in the metadata.

In practice, recording every pedestrian or vehicle is rarely feasible, since most road users travel in groups and most comply. The operational approach is therefore to record only the non-compliant cases and, in parallel, to count the total number of passing pedestrians and vehicles, so that the number of compliant cases is recovered as a residual.

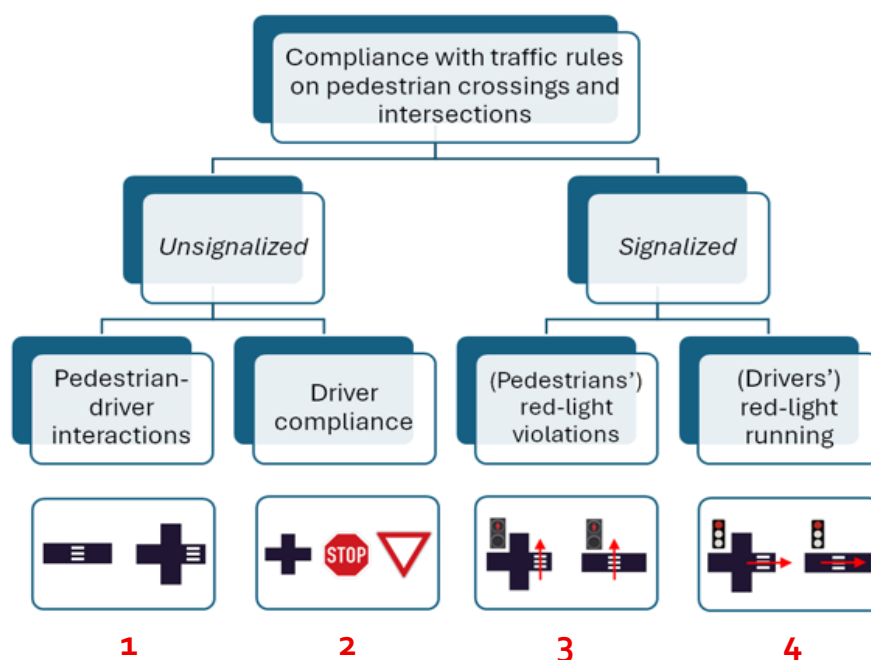
15.2.3 The compliance indicator cluster

The four SPIs are summarized in Figure 15-1:

- two “unsignalized”: (1) pedestrian-driver interactions and (2) driver compliance
- two “signalised”: (3) pedestrians’ red-light violations and (4) drivers’ red-light running

While SPIs (1), (3), (4) are observed on pedestrian crossings, which may be located on road segments (between intersections) or intersections, SPIs (2) are observed on intersections (with different priority controls), purposely without pedestrian crossings. The four SPIs are independent and not meant to be aggregated.

Figure 15-1. The compliance indicator cluster



15.3 Position within the behavioural framework

As set out in the introduction to behavioural SPIs, these indicators rest on the common backbone. Its distinctive feature is that the unit of analysis is the interaction (an encounter between road users), not a single road user. It should be noted that they can still be considered somewhat as an experimental indicator group of which the methodology is a working basis.

Three divergences are worth signalling at the outset. First, there is typically only one stratum, a set of similar pedestrian crossings in an urban context. Second, the unit of observation is not always the individual road user: in the unsignalised pedestrian context it is the interaction between a pedestrian and a driver, and compliance is assigned to the interaction as a whole. And third, the indicator is built from non-compliance events related to a separately counted exposure, rather than from a direct classification of every passing road user.

Given the fact that this was an experimental indicator in the first phase of Trendline, the methodology described here is still under development. The SPIs were assembled from two of the experimental indicators selected for testing within Trendline, red-light negation by road users, and compliance with traffic rules at intersections. The reference location characteristics, observation forms and sample structures set out below were applied in most cases during the Trendline pilots, with adaptations where local conditions required them, but some elements remain open:

- the operational definitions of compliance are deliberately left adaptable to national priority rules
- there is, as yet, no agreed surrogate-measure threshold for distinguishing compliance from non-compliance in the unsignalised context
- the sample structures are illustrative rather than prescriptive
- cyclists, motorcyclists and commercial vehicles are not yet integrated in the methodology.

The methodology is therefore expected to be refined in the light of the experience gathered during the new phase of Trendline, and the figures and forms in this chapter should be read as a working basis rather than a settled standard.

15.4 Scope

The indicators concern **passenger car drivers and pedestrians**, the two road user groups that dominate the crash and injury picture at pedestrian crossings in most European countries. Other motorised types (motorcyclists, light and heavy goods vehicles, buses and coaches) may be recorded optionally and reported separately where volumes allow; cyclists are a candidate for inclusion in a future phase. As elsewhere in this manual, the vehicle type is recorded as a variable where more than one is in scope, and the categories are defined for the fieldwork and in the metadata.

The population to which the SPI refers is the urban area of the country or, for sub-national indicators, a region. Following the general recommendation for behavioural SPIs (Chapter 5), the minimum total sample is 2,000 observations. The sample should be structured by combining a small number of conditions:

- location characteristics (based on different traffic volumes, speed limits, etc.)
- week period (weekday versus weekend)
- speed limit

15.5 Sampling and stratification

15.5.1 Selection of locations

Pedestrian risk is concentrated in urban areas, and pedestrian volumes outside built-up areas are usually too low for compliance observation to be meaningful; the SPIs therefore focus on urban locations. Within the urban network, traffic volume and speed are the critical contextual factors, both for the risk that compliance

behaviour generates and for the feasibility of observation, so locations should span a range of speed limits and traffic volumes.

Every urban network contains an enormous variety of pedestrian crossing layouts, differing in the number of lanes, the presence of central islands and cycle lanes, and their geometry. To keep the indicator comparable across countries and over time, locations should be selected from layouts that are typical but at the same time simple enough for real-time observation. The reference characteristics in Table 15-1 were applied in most cases during the pilots, with adaptation where local constraints required it; the first column applies to all compliance observations, the second and third to the signalised and unsignalised contexts respectively.

Table 15-1. Reference characteristics of observation locations (applied in most cases during the pilots, with local adaptation).

	General conditions	Unsignalised (specific)	Signalised (specific)
Recommended	Free-flow traffic Two-lane, two-way Sufficient signing and marking	No central island or median Sufficient sight conditions 3- or 4-leg intersection	No pedestrian countdown No push button 3- or 4-leg intersection Protected turn phase
To be avoided	Frequent congestion One-way Four lanes Cycle lanes Insufficient signing or marking	Central islands or medians Insufficient sight conditions Intersection with 5+ legs Roundabouts	Countdowns Push-button activation Intersection with 5+ legs Roundabouts Permissive turn phases

The locations should be chosen to reflect national or regional specificities so that the result is representative for the urban areas in the country or the region.

For the signalised indicator a typical structure combines at least ten locations, each with at least 100 pedestrians and 200 vehicles per location, giving at least 500 pedestrians and an at least 1000 vehicles per condition; moreover there should be 2,000 observations in all (which may be reached more easily with more locations). For the unsignalised indicator a typical structure uses at least ten locations, with 100 observations per location, again 2,000 in all (Table 15-2).

Table 15-2. Example sample structures (illustrative). The conditions are chosen to reflect national or regional specificities.

	Signalised (typical)	Unsignalised (typical)
Conditions	Three (e.g. low/medium/high traffic volume)	Three (e.g. different speed limits)
Locations per condition	10 locations	10 locations
Observations per location	100 pedestrians + 100 cars	100 observations
Per condition	500 pedestrians + 500 cars	1,000 observations
Total	2,000-3,000 observations	2,000-3,000 observations

15.5.2 Time period

Observations should take place in the season of highest pedestrian activity, avoiding adverse weather and poor visibility. Ideally observations should be outside the rush hours, so that pedestrian and vehicle volumes can be captured cleanly and the large, hard-to-categorise groups of the morning and evening peaks are

avoided. Where school-zone activity is the focus, its short, concentrated window should be planned for accordingly.

Each observation session lasts at least one hour and should yield at least 100 observations. In parallel with the behaviour observation, traffic and pedestrian counts are made should be made, ideally by additional observers (one for pedestrians and one for passenger cars), or from video footage. This ensures that compliant events can be related to the relevant exposure. At signalised locations this generally calls for three observers, one coding behaviour and two counting. Alternatively, like for the other behavioural SPIs, counting could take place during ten minutes before and after each observation session, or twenty minutes in the middle of it.

15.5.3 Observation methods and data sources

Behaviour at pedestrian crossings can be observed by human observers or by video, and the two involve a clear trade-off. **Human observers** have good situational insight, can adjust their position and perspective, and capture context that video does not preserve such as gender, age and informal communication such as eye contact. Their limitations are the training required and the need to record events reliably in real time. **Video** allows longer observation and detailed retrospective analysis, but the record is vulnerable to obstacles, sun glare and weather, raises GDPR and privacy concerns, and tends to lose information about road users outside the immediate study area, as well as the driver's-eye view of an event. Video analysis can be partly automated through surrogate measures of safety such as time-to-collision or post-encroachment time, but no single surrogate captures every dimension of an event. Also, in the unsignalised context especially, there is no clear threshold separating compliance from non-compliance. Trendline therefore currently retains **human observation as the standard method**, with video used where local circumstances make it more efficient.

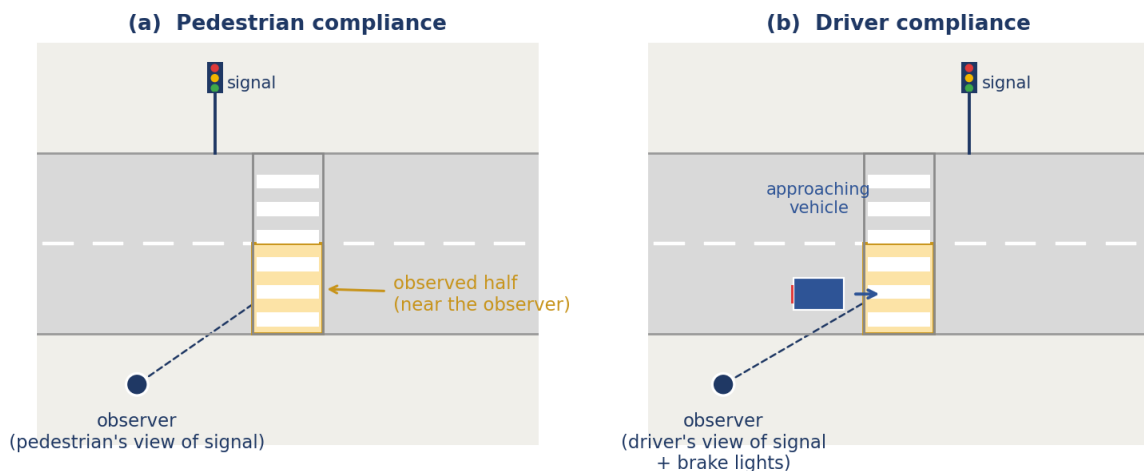
For red-light running specifically, a further source that can be used is data from fixed enforcement cameras (variously red-light cameras, safety cameras or red-light speed cameras). These typically record only non-compliant vehicles, so any indicator built on them must be supplemented with an independent count of all passing vehicles to recover the exposure denominator.

15.6 Data-collection specifics: observation approach for each of the four SPIs

15.6.1 Pedestrian compliance at signalised pedestrian crossings

Pedestrian compliance is observed at selected pedestrian crossings, on an intersection or a road segment. To keep a clear view of the signal from the pedestrian's perspective, only the half of the crossing closer to the observer is observed (Figure 15-2a). In this Figure, for pedestrian compliance (part (a)), the observer sees the signal as the pedestrian does; for driver compliance (part (b)), the observer sees the signal as the driver does and can read the brake lights.

Figure 15-2. Observer vantage points and the observed half of a signalised crossing.



For each non-compliant pedestrian the observer should record:

- gender (male or female)
- age category (0-14, 15-64, 65+)
- distraction by mobile phone, headphones or earphones

Optionally, it may be recorded whether the crossing started before the red signal turned to green (early start) or whether the crossing started after the green signal turned red (late entry).

A "group" is defined as a cluster of two or more pedestrians, whether related, such as a family or a school class, or unrelated. When such a group crosses, no individual gender, age or distraction is recorded; instead the number of pedestrians in the group is recorded or estimated. Where video is used instead, the camera is positioned so that all the required characteristics, including the signal state, are visible.

An optional variable to be considered is the number of people respecting green/red, but crossing largely outside the zebra stripes? This requires a clear convention, e.g. at least 5 meters.

15.6.2 Driver compliance at signalised pedestrian crossings

Driver compliance is observed at the same locations but from a vantage point that gives the observer the signal from the driver's perspective (Figure 15-2b). For each vehicle, the vehicle category and one of three behaviours is recorded:

- stopped on yellow or red
- did not stop on yellow or red
- started or continued on yellow or red.

Where vehicles queue at the red signal, only the first, uninfluenced vehicle is observed. Vehicles that arrive just after the green has ended and clearly cannot stop safely, given their speed and distance from the signal, are excluded from the "did not stop" category. Where video is used, the camera must capture both the brake lights and the signal. Records from red-light cameras may also be used, subject to the caveat above that they record only non-compliant vehicles and so require an independent total count for the denominator.

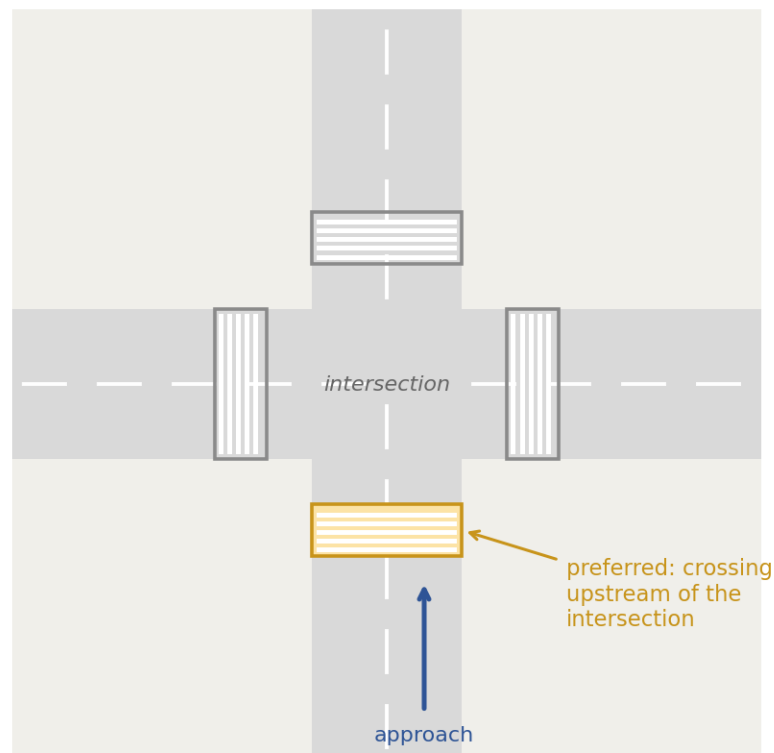
15.6.3 Pedestrian-driver interactions at unsignalised pedestrian crossings

Interactions are observed at selected pedestrian crossings, on an intersection or a road segment. For a clear view, only the half of the crossing closer to the observer is observed. In some cases, there are several pedestrian crossings on an intersection. Each crossing may have different impact on driver – it may be that he/she adjusts the speed before the first crossing (upstream of the intersection) and thus his/her speed before the second crossing (downstream of an intersection) is already lower. To ensure that observation captures the

natural pedestrian-driver interactions (uninfluenced by external influences), the crossing upstream of the intersection is preferred for observation (Figure 15-2).

An interaction is defined as any mutual reaction involving a change of speed. This could be a pedestrian stopping because a driver does not yield, or a driver braking or stopping to yield to a pedestrian. Reactions of single road users in isolation are excluded.

Figure 15-3. Schematic view of the upstream pedestrian crossing at an intersection.



Each interaction is coded by combining a pedestrian behaviour (entered or did not enter the pedestrian crossing) with a driver behaviour (yielded or did not yield), and compliance is then assigned at the interaction level. Thus an adult pedestrian given right of way by a yielding driver is a fully compliant interaction, whereas a driver who fails to yield (with the result that a child does not enter the pedestrian crossing) is a non-compliant interaction, even though the child's own behaviour was compliant. The observer records,

- for the pedestrian:
 - gender
 - age category
 - distraction (by mobile phone, headphones or earphones)
 - whether they entered the pedestrian crossing
- for the driver
 - vehicle category (at least passenger/freight vehicle)
 - whether they yielded.

The group rule is as in the signalised case. Where the volume of interactions is high enough that some are missed, traffic and pedestrian counts are made in parallel to recover the exposure denominator; where every interaction is registered, the total is known directly and no separate count is needed.

15.6.4 Driver compliance at unsignalised intersections

Driver compliance with priority rules is assessed at intersections without designated pedestrian crossings, selected to represent the different forms of priority control:

- uncontrolled intersections where priority to the right applies by default
- intersections with a give-way (yield) sign
- and intersections with a stop sign.

The aim is to observe whether drivers correctly yield according to the applicable control or default rule, in the absence of marked pedestrian infrastructure. At give-way intersections, two vehicles may arrive almost simultaneously, and compliance then depends not only on who yields but on the relative speed and distance of the vehicles. Some borderline cases can be resolved only by detailed event reconstruction, which is outside the scope of routine SPI collection. Observers should focus on clearly observable, unambiguous cases. The data reporting form therefore distinguishes only two outcomes: compliance and non-compliance. Because it is realistic to register every passing vehicle, no separate traffic count is required.

15.7 Variables, analysis and reporting

Each observed case is coded binary: a one for compliant and a zero for non-compliant behaviour (or, in the unsignalised pedestrian context, for the interaction as a whole). The number of compliant events is related to the exposure given by the parallel counts, and the stratum-level estimates should be weighted to national or regional figures or estimates following the framework of Chapter 7. Where the data permit, the indicator should be broken down by road-user characteristics (gender, age), by location characteristics (speed limit, traffic volume) and by week period. As for the other behavioural SPIs, each delivery records the unweighted number of observations and of locations and is accompanied by the metadata package; the operational definitions of compliance, being open to national adaptation, are documented there so that the indicator can be read in context.

The metadata for each observation should include at least:

- the location ID, coordinates, address
- the day and time
- the speed limit
- the type of pedestrian crossing or intersection
- the relevant marking and signing, including the green time at signalised pedestrian crossings.

15.8 Minimum sampling and reporting requirements

The table below summarises the requirements for the two contexts. As an experimental indicator, the figures are working minima rather than a settled standard.

Table 15-3. Minimum sampling and reporting requirements for the compliance indicators.

Aspect	Signalised contexts	Unsignalised contexts
Indicators	Pedestrian and driver compliance with traffic-light priority	Pedestrian-driver interactions on pedestrian crossings; driver compliance at intersections
Unit of analysis	Individual pedestrian; individual driver	Interaction (crossings); individual driver (intersections)
Coverage	Urban; range of speed limits and traffic volumes; simplified configurations	Urban; range of speed limits and traffic volumes; simplified configurations
Road users	Car drivers and pedestrians (others optional, reported separately)	Car drivers and pedestrians (others optional, reported separately)
Method	Human observation (standard); video where more efficient; red-light cameras supplementary for red-light running	Human observation (standard); video where more efficient
Sample size	≥ 2,000 observations; ≥ 100 per session; ≥ 1-hour sessions	≥ 2,000 observations; ≥ 100 per session; ≥ 1-hour sessions
Counts	Parallel pedestrian and vehicle counts for the denominator (typically two observers)	Parallel counts where interactions are missed; none needed for driver-only intersections
Timing	Season of highest pedestrian activity; off-peak; avoid adverse weather and poor visibility	Season of highest pedestrian activity; off-peak; avoid adverse weather and poor visibility
Reporting	Compliant share + 95% CI; breakdowns by user, location and week period; metadata incl. operational definition	Compliant share + 95% CI; breakdowns by user, location and week period; metadata incl. operational definition

16 SPIs on use of lighting by cyclists in the dark

16.1 Introduction

16.1.1 Rationale

Although distances travelled are shorter at night, cyclists' crash rate per kilometre is consistently higher than during the day. Twisk and Reurings (2013) and Wood (2023) have shown that cyclists are less visible and less easily identifiable as such by other road users in the dark, and quasi-experimental work by Owens and Sivak (1996) has identified lack of visibility as a major contributory factor in fatal collisions involving cyclists and motor vehicles at times more important than drivers' alcohol consumption. Functional front and rear lighting have been estimated to reduce the risk of cyclists being involved in a collision with a motor vehicle in the dark by around 17% (Kuiken and Stoop, 2012)

Bicycle use has grown markedly across Europe since the COVID-19 pandemic, driven by health, climate and economic considerations (Bouwen et al., 2022). Cycling in the dark is therefore expected to grow proportionally, even in countries where cycling remains a marginal mode of transport. Stagnating cyclist fatality figures across Europe make the indicator particularly relevant for traffic-safety monitoring. Reported figures vary widely between countries: in 2012, 44% of US cyclists reported using a front light and 34% a rear light, while a Dutch observational study in 2022 found 74% and 67% respectively (Schroeder and Wilbur, 2013; Timmermans et al., 2022)

16.1.2 Indicator definition and theoretical population

Three related indicators should be delivered. They are defined as:

Percentage of cyclists using a front light in the dark.

Percentage of cyclists using a rear light in the dark.

Percentage of cyclists using both a front light and a rear light in the dark.

In each case the denominator is the population of cyclists in the dark on urban roads (and, optionally, on rural roads).

It is essential to distinguish a "light" from a "reflector". For the purposes of this indicator, a light is any device that produces light by itself; a reflector which only reflects light when light is directed at it is not counted. The SPI measures the use of lights, not their correct use, since the legal definition of "correct use" varies by country and would limit international comparability. Each country is, of course, free to collect additional data on correct use under its national legislation.

To accommodate the practical difficulty of judging visibility from the roadside, light use is coded in three categories:

- fully visible (no doubt, no potential occlusion)
- partially visible (occluded, or where doubt arises — for example because of a low battery)
- not visible at all.

Only fully visible front and rear lights are counted as "in use" for the purpose of the headline SPIs. When lights are carried on the body, only lights on the rider are observed, not on any passenger.

The theoretical population is the total number of kilometres ridden over the national territory by cyclists during the reference period. By weighting the sample by traffic volume, the percentage of cyclists observed using lights may equivalently be interpreted as the percentage of kilometres ridden with lights, which is the policy-relevant exposure measure. Because the legal requirement to use lights in the dark differs between countries,

the applicable national rules are to be included in the metadata. As in the cyclist subcategory of the helmet SPI (Chapter 12), the cyclist population covers conventional and power-assisted bicycles (e-bikes) but excludes speed pedelecs, which are formally classified as L1e-B vehicles (equivalent to mopeds).

16.1.3 Position within the behavioural framework

This indicator rests on the common backbone of the behavioural indicators (Part B) and use a methodology that in many respects is similar as the cyclist helmet observation discussed in Chapter 12: the two share their target population, their roadside observation procedure, their sparse-rider sampling problem and their traffic-counting and weighting logic. A structural peculiarity of this SPI is that it is measured in the dark, which governs the timing of sessions.

Data collection and analysis for this SPI is not funded under the current Trendline grant phase. It is documented here because it is of clear policy relevance, has been shown feasible in pilot studies in Belgium, Poland, Croatia, Spain and the Netherlands, and may be supported in a future phase of the programme. The methodology below should accordingly be read as a tested but still-developing basis, open to refinement in the light of further experience.

16.1.4 Scope

The cyclist-lighting indicator concerns a single road-user group: people cycling in the dark. It therefore has no internal subcategory structure, but its boundaries still need to be fixed with the same care, in two respects: which riders count as cyclists, and what is meant by "in the dark".

The cyclist population is the same as in the cyclist subcategory of the helmet indicator (Chapter 12) and as set out in the indicator definition above: it comprises the riders of conventional and power-assisted bicycles (e-bikes) and excludes speed pedelecs, which are formally classified as L1e-B vehicles. A person wheeling a bicycle on foot is a pedestrian, not a cyclist, and is excluded. Only riders who are actually cycling are observed, whether they are in motion or momentarily stopped, for example at a traffic light. Where a bicycle carries more than one person, only the rider is counted, so that a passenger does not give rise to a second observation.

The condition that defines this indicator is that observation takes place in the dark. "In the dark" is the period in which daylight is insufficient and the use of cycle lighting becomes relevant for visibility (and also compulsory in almost all countries).. Observations made in daylight or in twilight fall outside the scope of the headline SPIs. Because the onset and end of darkness shift with the season and with latitude, the operational rule for deciding when a session counts as taking place "in the dark" should be fixed in advance and documented in the metadata, so that the indicator remains comparable across locations, seasons and countries. This condition also governs the timing of the observation sessions, which should be scheduled in evening or early-morning darkness as described under Sampling and stratification.

16.2 Sampling and stratification

Trendline recommends the same stratification approach as used for the helmet SPI but instead of weekday daytime, measurements should take place on an evening or early-morning darkness. Location selection should follow the sparse-rider procedure of the sampling chapter (Section 6.1.3) and the cyclist specifics of observing helmet use (Chapter 12). The Trendline requirement is to cover at least urban roads. Where countries observe both rural and urban roads, road type is added as an additional stratum. Regional stratification is added when nation-wide measurements would obscure substantial regional differences.

The minimum number of observations is 2,000 overall, with a minimum of 500 per stratum when both rural and urban observations are included. As elsewhere, the minimum refers to valid observations carrying the minimum required data, and may imply disproportionate sampling relative to traffic in smaller Member States, which is acceptable and corrected at the weighting stage. The recommended minimum number of

locations is 10 per period (and per road type if rural roads are included), with at least 2 locations per crossed combination of road type and period.

16.3 Data-collection specifics

16.3.1 Direct observation as the primary method

Direct observation along the roadside is the recommended primary method, preferably from a safe place such as an intersection or roundabout where speeds are reduced. The pilot studies undertaken in the first phase of Trendline have confirmed that light use by cyclists in the dark can reliably be measured by direct observation at the roadside. The observation procedure is the standard one: observe one bicycle, code the data, then proceed to the next. When traffic volume prevents observation of every passing cyclist, the next subject is selected at random, normally as the next passing cyclist after coding has been completed. Quality of observation should always take precedence over quantity.

Observations are made by well-trained, independent staff under the supervision of a coordinator, and receive both theoretical training (a briefing on aims, variables, definitions and procedures) and practical training (exercises on the road), in line with Chapter 7. They are given clear guidance on the road section and direction to observe, the duration of sessions, and the procedure to follow when observations cannot be performed because of weather, visibility or safety problems. Inter-rater reliability is checked before fieldwork starts and at intervals during the campaign, with particular attention to the categorical variables that depend on interpretation, such as the type of bicycle or the type of cycling infrastructure, and to the central judgement of this SPI: whether a light is fully visible, partially visible or not visible at all, and whether an apparent light source is a light or a reflector.

16.3.2 Field procedure

Data are collected in conditions of darkness or twilight, in the early morning and the evening. “Darkness” (including twilight) is defined as a natural light density below 26 lux (see Appendix E). The use of a lux meter is the most accurate way to verify this threshold. In practice, sunrise and sunset times can be used as a basis, with one or a few lux-meter readings to confirm the correct timing. In many European countries the threshold of a light density <26 lux corresponds with about half an hour before sunrise and half an hour after sunset.

The optimal measurement period is a compromise between the season, the natural light level and the reasonable probability of encountering cyclists. This indicator departs from the default fieldwork window of the sampling chapter for a stated reason: it is observed in darkness, so late autumn and winter, which combine darkness with a reasonable chance of encountering cyclists, are the preferred window. Light use is observed on weekdays and weekend days and reported separately for the two.

One observation session lasts at least 30 minutes and ideally one hour, excluding the time required for traffic counting and the recording of environmental data (it is recommended to count the bicycle traffic during the observation period, rather than before or after). Each location may be visited for several sessions in different time intervals (typically once in the morning twilight and once in the evening twilight) or assigned at random to a single interval, provided that morning and evening observations are balanced across all locations. When organising fieldwork, periods and timings are balanced so as to avoid systematic bias, for example so that weekday sessions are not all planned for the morning. Sessions take place only in good enough conditions: no heavy rain, no storm, no snow or ice, no fog, and undisturbed traffic flow. Coding may be carried out on paper, tablets or smartphones.

16.3.3 Traffic counts and weighting

Traffic counts are required at each location during each observation session. Each cyclist riding in the same direction as those being observed is counted; if all passing cyclists are observed, the count equals the number

of observations and no separate counting is needed. It is advised to count the traffic during the observation session. If this is not possible, given the volume of traffic or the number of observers, counting is conducted for at least twenty minutes per session, either as a ten-minute window before and after the session, or as a continuous twenty-minute window in the middle of the session, and is then extrapolated over the full session duration. As for the other behavioural SPIs, the traffic count is required even when national traffic-volume statistics on cyclists by road type would be available, since it underpins the calculation of the share of light users per session and the correct calculation of the confidence intervals and the post-stratification weights.

The within-stratum and between-strata weighting should follow the rules described in Section 8.2, the exposure measure being cyclist-kilometres ridden; the calculation of confidence intervals should respect the nested design as discussed in Section 8.4. The traffic count follows the approach of Section 7.3.2, the unit being the cyclist riding in the observation direction.

16.3.4 Automatic detection via roadside cameras

The generic conditions for camera-based observation are those of Chapter 7 (Fieldwork); the optical and siting requirements are in Appendix E. This section keeps only the points specific to cyclist lighting:

A Dutch pilot study (Schmidt et al., 2025) used existing camera footage at intersections. The footage was selected to coincide with sufficient darkness (between 22:30 and 06:00 in summer months) and observers identified the visibility of the lights and whether they were on, off, or could not be established. Front-facing cameras detected front-light use in 82% of cases, and rear-facing cameras detected rear-light use in 72% of cases; the maximum detection rate from a side view was 65% for both front and rear lights. The conclusion of the SWOV team is that footage from multiple camera perspectives is needed to obtain the highest success rates for both light types simultaneously.

Several factors limit the reliability of automated or semi-automated camera observation:

- streetlights illuminating the scene make it harder to distinguish reflectors from active light sources
- well-lit overtaking vehicles can illuminate cyclists from behind, masking light use
- low frame rates and large camera-to-cyclist distances reduce image quality
- the position of the lights varies by bicycle type (bag, jacket, fork on rental bicycles)
- some lights operate only when pedalling and dim or extinguish when the cyclist stops at traffic lights
- groups of cyclists at traffic lights tend to obscure each other's lights.

Where cameras are used, all of these factors must be assessed in advance, and the technical aspects of the cameras, sampling procedures, accuracy (false-positive and false-negative ratios), data-collection and coding procedures, data-quality controls, treatment and weighting must be described in detail in the methodological report.

On the basis of the SWOV pilot, cameras can be used to measure this SPI provided that the following conditions are met:

- both a front-view and a rear-view camera are deployed per direction (so 2 + 2 cameras when both sides of a single cycle lane are measured)
- the costs of installation and footage extraction are properly estimated
- suitable camera positions are identified
- all relevant road types are covered
- the necessary permissions and privacy approvals are obtained
- an inventory of pre-installed cameras is consulted to reduce costs and allow repeat measurement at the same locations.

16.4 Variables, analysis and reporting

16.4.1 Variables to be collected

For each observed cyclist the minimum dataset should record per observed cyclist:

- the front-light status (visible / partially visible / not visible)
- the rear-light status (visible / partially visible / not visible)

Light status is recorded for the rider only, not for any passenger.

Recording both statuses for the same cyclist is what allows the three SPIs for use of lights to be derived: the front light and the rear light are each estimated separately, and, because the two are recorded jointly for every rider, the combined indicator — the percentage of cyclists using both a front and a rear light — is obtained from the same records without a further variable. The SPIs with the percentages of “in use” should be based on only fully visible lights.

Per session at least the following data should be recorded

- the location
- the position of the observers (or the cameras)
- the road type
- the date, the start and end time of the session; the session duration
- the unique location and session codes
- the duration of the traffic-counting period
- the traffic count.

Other variables may be added without compromising this minimum, on the condition that each can be observed reliably from the roadside in the dark. Such optional variables include:

- the type of bicycle (electric or non-electric; city or sport; private or public/shared; professional or non-professional rider)
- the street-lighting status at the moment of observation (no streetlights, streetlights on, streetlights off)
- the cycling infrastructure (no infrastructure, cycle lane, segregated cycle track, etc.)
- the use of reflective clothing or of reflectors (front, rear, on the pedals or on the wheels)
- aspects of correct use under national legislation (light colour, fixed versus flashing, minimum visibility distance, light location)
- the rider’s estimated age category and gender
- the presence of a passenger.
- professional riders (couriers, postal workers, police) versus non-professional riders
- public/shared bicycles versus private bicycles

16.4.2 Weighting and confidence intervals

For both the front and rear light, each observed cyclist is coded with a one for a fully visible light in use and zero otherwise. The three SPIs are computed as weighted proportions following the framework of Chapter 8, with the weighting described in Section 16.4.1. Confidence intervals must reflect the nested sampling design (observations within sessions within locations) and are calculated as set out in Chapter 8. The statistical techniques used to weight and analyse the results and to calculate the confidence intervals should be documented in the methodological report

16.4.3 Reporting and data delivery

The three SPIs — the percentage of cyclists using a front light, a rear light, and both, in the dark — are each reported with a point estimate and a 95% confidence interval, both overall and for the required breakdown by week period (weekday and weekend). Where a regional stratification is used, the indicators are also reported

by region, with the minimum requirements applying within each region and results weighted by regional traffic volume. Specific estimates for combinations of levels, with their confidence intervals, are recommended where the sample permits.

Results can be delivered at the three levels described in Part D (see Table 23-6);. Each delivery should be accompanied by metadata describing at least:

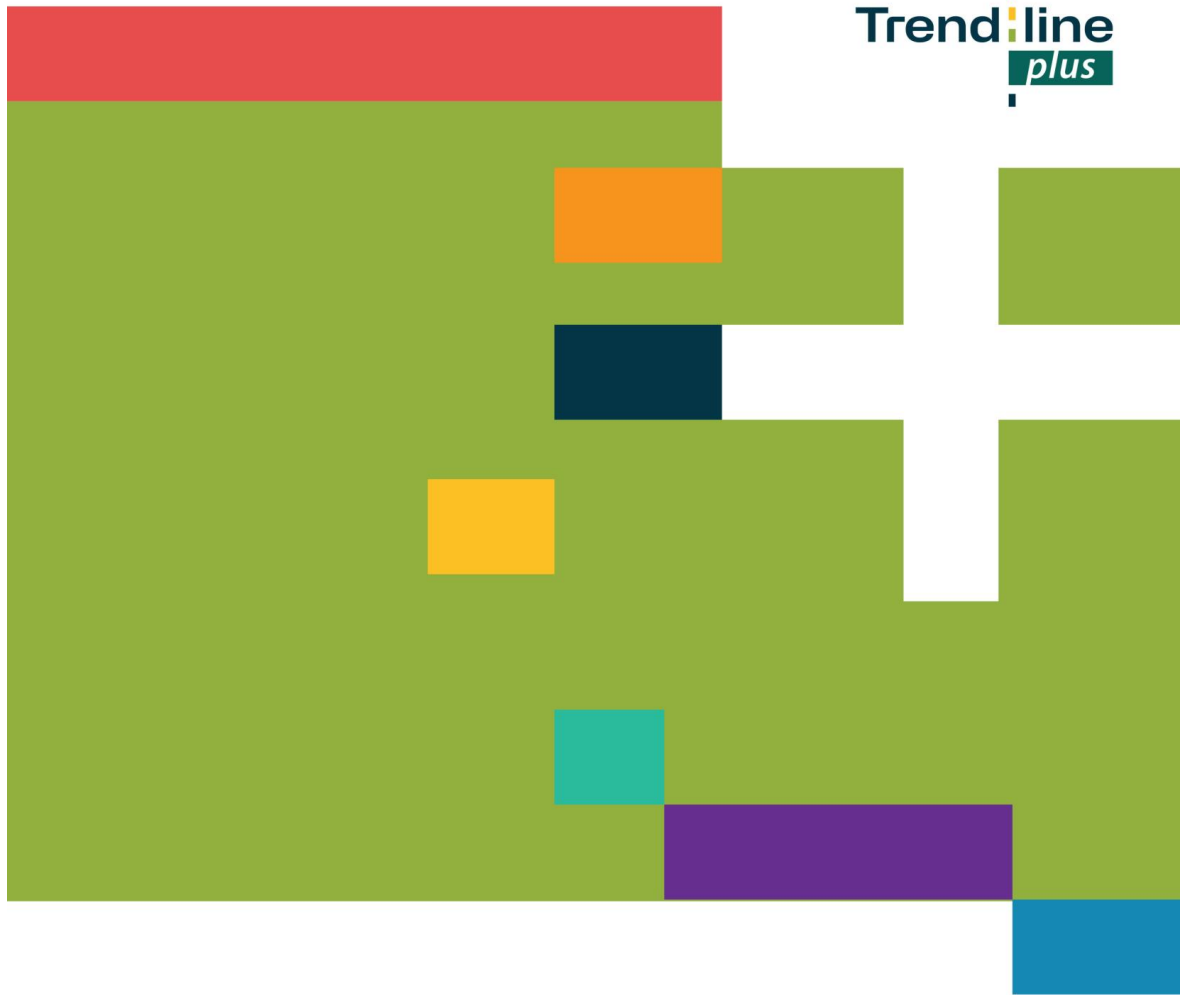
- the national legislation on cyclist lighting
- the observation method (roadside observers or cameras)
- (when cameras were used) how privacy was protected and how the method was validated
- the location-selection procedure
- the fieldwork period and months
- how darkness was verified
- the weighting version and confidence-interval method
- any deviation from the minimum requirements, with its reason.

16.5 Minimum sampling and reporting requirements

Table 16-1 consolidates the minimum and optional requirements for the cyclist-lighting indicator. It mirrors the format used for the other behavioural SPIs and can be read as the lighting-specific specialisation of the common requirements set out in Chapters 6 and 7. As this is an experimental indicator, the figures are working minima rather than a settled standard. Where a minimum cannot be met, the shortfall and its reason are recorded in the metadata, and the affected stratification level is either combined with another or omitted from reporting, no national SPI then being formed.

Table 16-1. Trendline minimum requirements and optional additions for the cyclist-lighting indicator.

Aspect	Minimum requirement	Recommended / optional
Indicators	% of cyclists using a front light; using a rear light; using both — in the dark, fully visible lights only	Specific estimates for crossed levels
Road type	Urban roads (inside built-up areas) only	Rural roads, if the minimum sample is reached for that stratum
Periods	Weekday and weekend (in darkness) — separate SPIs	Regional stratification
Sample size	≥ 2,000 observations overall; ≥ 500 per stratum if urban and rural are combined	Boost numbers for accuracy and further breakdowns
Locations	≥ 10 per period (and per road type if rural is included); ≥ 2 per crossed stratum; minimum flow 10 cyclists/hour	More locations for larger countries
Sessions	≥ 30 minutes (1 hour preferred); in darkness, < 26 lux	Lux-meter verification of timing
Method	Direct roadside observation by trained, independent observers	Validated camera-based detection (front + rear view per direction)
Traffic counts	Count of cyclists each session (≥ 20 minutes) for the exposure denominator	—
Timing of year	Season where darkness coincides with cyclist presence (end of autumn / winter)	Complementary measurement in a second season
Reporting	Point estimate + 95% CI overall and by week period; standardised template; metadata	Crossed matrix; cleaned individual-cyclist dataset



PART C – DATABASE-BASED SPIs

17 Common framework for database-based SPIs

17.1 Introduction

The behavioural SPIs of Part B are built by observing or questioning individual road users and weighting the resulting sample back to a national figure. In contrast, the indicators treated in this Part C of the manual are built in a different way. They are derived from existing structured datasets that describe aspects of the road safety system rather than the behaviour of individuals in traffic: national road inventories and network-wide safety assessment results, road authority records of speed limits and road geometry, vehicle registration registers, emergency medical service operational logs, and police or judicial enforcement databases. Because they describe system-level conditions, these indicators raise a different set of methodological questions from the behavioural SPIs, and it is convenient to treat the elements they share in a single common chapter before turning to each indicator in turn.

Five indicators are covered in Part C:

- infrastructure safety (Chapter 18), based on the network-wide safety rating of roads within the scope of Directive 2019/1936/EC;
- the share of 30 km/h roads in urban areas (Chapter 19), an experimental indicator based on urban road characteristics;
- vehicle safety (Chapter 20), based on the Euro NCAP safety rating of the passenger car fleet;
- post-crash care (Chapter 21), based on the response time of the emergency medical services; and
- enforcement of traffic regulations (Chapter 22), an experimental indicator based on enforcement activity recorded in administrative databases.

Although these five indicators describe very different parts of the road safety system, they share a common methodological backbone. The present chapter sets out that backbone:

- the general principles that apply to all database-based SPIs
- the data sources they draw on and the quality issues these sources raise
- the steps by which raw administrative data are turned into an indicator value
- the metadata that must accompany every value
- the alternative approaches that are available when the preferred data source is missing.

Each of the indicator-specific chapters that follow refers back to this chapter for the elements held in common and restates them only where the indicator genuinely diverges.

It is useful to be explicit about how database-based SPIs differ from the behavioural SPIs of Part B. Behavioural SPIs are constructed by sampling road users at the roadside or through surveys; the methodological energy of a behavioural SPI study goes into the design of that sample and the calculation of the confidence interval around it. Database-based SPIs, by contrast, are constructed by extracting and processing datasets that were created for purposes other than research and that are, in principle, already full-population datasets rather than samples.

This shift in data origin produces a corresponding shift in the methodological agenda. Where for a behavioural SPI the concern is whether its sample is representative of the underlying traffic exposure, for a database-based SPI the main concern is about whether the underlying database is complete and whether its definitions match what the indicator specification requires. Where a behavioural SPI reports a confidence interval dominated by sampling variability, a database-based SPI should report the share of the relevant system that its data cover, together with a careful account of the boundary between what the database does and does not capture. Where a behavioural SPI designs fieldwork protocols for observers to follow in real time, a database-based SPI

designs extraction and linkage procedures that usually require the cooperation of one or more third-party data owners. Table 17-1 summarises these differences.

Table 17-1. Differences between behavioural and database-based SPIs at a glance.

Aspect	Behavioural SPIs	Database-based SPIs
Data origin	Sampled observations or surveys	Existing administrative or operational registers
Coverage objective	Representativeness of the sample	Completeness and definition match of the register
Dominant uncertainty	Sampling variability	Coverage gap and definition bias
Quality control	Fieldwork supervision; observer reliability	Data linkage, cleaning and metadata documentation
Critical capability	Sample design; statistical estimation	Database engineering and stakeholder coordination
Principal partner	Field-research provider	Data-owning agency

These differences matter for the way a country organises its SPI programme. They affect who needs to be involved (data-owning agencies rather than field observers), what skills are required (database handling and the negotiation of data access rather than fieldwork supervision), where the principal risks lie (registry bias and drift in definitions rather than observer bias), and where the dominant cost falls (data access and cleaning rather than fieldwork hours). Treating a database-based SPI as merely another behavioural SPI that can be slotted into the same operational arrangements would tend to produce indicator values that are harder to defend than they need to be.

17.2 General principles

A first general principle is that database-based SPIs in general benefit from being grounded in a well-defined regulatory or methodological framework. The infrastructure indicator illustrates this clearly: its scope, its road type categories and its underlying assessment methodology all derive from Directive 2019/1936/EC and from the EGRIS Network-Wide Road Safety Assessment methodology endorsed by the Member States in November 2022. An indicator that rests on such a framework enjoys a stable definition over time and a degree of cross-country comparability that is difficult to achieve with ad hoc data collection. Where no common framework yet exists the corresponding indicators are necessarily more experimental, and the guidance has to combine flexibility with explicit methodological caveats.

A second general principle is that database-based SPIs should remain conceptually distinct from the outcome indicators on deaths and serious injuries. The Trendline framework deliberately bases the infrastructure indicator on the proactive component of the EGRIS methodology, even though the integrated assessment results — which combine the proactive and reactive components — are technically available; the reason is that an SPI should describe the conditions and processes that lead to crashes, not the crashes themselves. The same logic applies to the enforcement indicator, which deliberately measures the effort and effectiveness of detecting violations rather than the resulting level of compliance or the crash counts. As far as possible, a database-based SPI should measure system-level inputs and capacities rather than restate the outcome statistics that the policy framework already monitors directly.

A third general principle is that the value of a database-based SPI should not be interpreted in isolation from its target and its policy context. The 30 km/h indicator is explicit that its target value is not 100 %, because main arterials with higher speed limits are needed to keep through-traffic away from residential areas. The

enforcement indicator is explicit that a high level of enforcement does not in itself mean high compliance or high safety, but signals the effort a country puts into addressing specific dangerous behaviours. The vehicle safety indicator notes that it is not self-evident that a higher national score on the safety rating of the fleet translates into better national road safety performance, since that relationship has not yet been demonstrated. The manual signals these interpretive nuances where the indicators are introduced, and a country reporting a value should signal them too.

A fourth general principle is that a full-population measurement is preferred to a sample. Because the underlying datasets are typically registers rather than surveys, the natural ambition of a database-based SPI is to cover the entire relevant population without the need for weighting or extrapolation. Sampling is treated throughout Part C as a fallback to be used only where no suitable national database exists, and the methods for that fallback are set out in Section 17.6.

A fifth general principle is that each indicator must define precisely what constitutes one record and what level of aggregation is required for national reporting, because the unit of analysis and the implicit measure of exposure differ from one indicator to the next. Table 17-2 sets these out. The unit of analysis determines what is counted and cleaned; the exposure measure determines how individual records are weighted when they are combined into a national figure. Both differ markedly from the traffic-volume exposure that underlies most behavioural SPIs, and confusing the two would be a source of error.

Table 17-2. Unit of analysis and implicit exposure for the database-based SPIs.

Indicator	Unit of analysis	Implicit exposure
Infrastructure safety	Road section	Distance driven (veh-km), or road length as a proxy
30 km/h in urban areas	Road (or lane) length unit	Urban road length, or lane length / distance travelled
Vehicle safety	Registered passenger car	Number of cars in the fleet (or newly registered)
Post-crash care	Emergency intervention following a road crash	Number of injury crashes requiring a response
Enforcement	Enforcement action, ticket or installed device	Population (and optionally network length or area)

17.3 Data sources and quality issues

Database-based SPIs rely on administrative, operational or registry data. These data were collected for a purpose other than producing a road safety indicator. A vehicle registration database exists primarily to administer ownership and roadworthiness, not to characterise the safety of the national fleet; an emergency medical service log exists to coordinate emergency response, not to provide a post-crash-care indicator; a road network inventory exists to support maintenance and planning, not to feed a safety classification; an enforcement database exists to manage the penal procedure, not to gauge the effort a country devotes to enforcement. In every case the indicator is a secondary use of data whose primary purpose lay elsewhere, and this secondary-use status has three predictable consequences:

- The data may not exactly match the indicator definition. The variables that were collected, the categories that were used and the inclusion criteria of the source database were optimised for the original purpose, and any of them may differ from what the SPI requires. Calculating the indicator therefore often involves a non-trivial mapping step, in which records and variables from the source database are re-coded to match the indicator's own scheme. The post-crash-care indicator is a clear example: a national emergency database records all interventions, but the indicator needs only those relating to road crashes, only the first

medical unit to arrive, and only services that provide medical pre-hospital care, so that a sequence of selection decisions stands between the raw register and the indicator population.

- The data may contain structural biases that reflect the operational priorities of the data owner. A database that exists primarily to manage enforcement may under-record activity at locations or for offences that are rarely enforced; a register maintained for billing or administrative purposes may be more complete for some categories of record than for others. These biases are not removed by having a large number of records, because they are properties of the register rather than of the sample size.
- The data may carry institutional sensitivities. Data owners are sometimes reluctant to release operational data that could be used to assess their own performance, or that could expose individual incidents to scrutiny. The legal basis for sharing may also be more restrictive than for purpose-built research data. This is markedly so where health or privacy regulation applies, as it does to emergency medical data.

Several practical steps help to manage these consequences. The data owner should be approached early, ideally during the planning phase rather than at the moment the data are needed; the legal and operational basis for sharing should be documented (what may be shared, in what aggregation, with which named recipient and for which agreed purpose); and the lead time required for the data owner to prepare the data should be anticipated and built into the schedule, since it can be substantial when a third party is involved.

A recurring choice concerns the use of non-official sources. Official sources are the primary reference for every database-based SPI. In general, non-official sources such as open databases such (e.g., OpenStreetMap), commercial providers such as TomTom or Google for road characteristics, or the crowd-maintained databases of speed and red-light cameras that can inform the enforcement indicator, should be used only to fill gaps that the official sources leave. OpenStreetMap has the advantage of bringing many cities together in a single database and of being free, but its completeness and quality vary across countries, regions and cities; commercial providers tend to offer higher quality at a price. Where a non-official source is used, the indicator value should be flagged as approximate or indicative, and the metadata should describe the source, the way its quality was assessed, and any cross-checks performed against official figures.

Partial coverage of the relevant system is a structural feature of several of these indicators rather than an accident of a particular country's data:

- Infrastructure rating systems typically cover motorways and major rural roads but little or none of the urban local network
- The 30 km/h indicator depends on urban road databases whose completeness varies widely
- Vehicle registers exclude certain categories of car
- Emergency databases differ in whether they capture every region and every type of responding unit
- Enforcement databases differ in how they record local police activity and automated enforcement.

Whatever the indicator, a country should always report the share of the relevant target system that its data cover, and any indication that the covered and uncovered parts differ systematically in their risk profile. Extrapolation from rated to unrated roads, from cities with data to cities without, or from recorded to unrecorded activity is generally not recommended, because it can produce values that look comparable on the surface but rest on quite different empirical foundations from one country to the next.

Finally, no dataset is perfect. Common quality problems include the systematic under-registration of records, erroneous values and outliers, and missing key fields, such as the timestamps on which the post-crash-care indicator depends or the safety rating of an unrated car model. The metadata should state the proportion of records for which the required information is missing or erroneous. A particular concern is temporal consistency: speed limits, fleets and networks change, so the data must carry a clear reference year or date, and a value can be compared across years only where the definitions and the registration practices behind it have remained consistent.

17.4 Data processing

Turning an administrative dataset into an indicator value follows a recognisable sequence, even though the detail differs from one indicator to the next. The sequence is mapping, cleaning, linkage, calculation and weighting and the estimation of uncertainty (in cases where a full population is not available). Each of these steps are briefly described below:

- Mapping is the re-coding of source variables and records into the categories the indicator requires, following from the definition match. It ranges from the straightforward assignment of a road section to one of the four road types of the infrastructure methodology, through the identification of urban roads and the reading of their speed limits for the 30 km/h indicator, to the construction of an explicit conversion table that links the commercial model names in a national vehicle register to the model names used in the Euro NCAP database. Every mapping decision of any consequence should be recorded in the metadata.
- Cleaning is the removal or correction of records that are incorrect, incomplete in their essential fields, duplicated or improperly formatted, so that the calculation rests on data that are fit for use rather than on the register exactly as received. What is to be reported to the Trendline databases, is data based on a cleaned database. The cleaning rules and their effect, in particular the number and proportion of records removed or left unresolved, belong in the metadata, because they bear directly on coverage.
- Linkage is required wherever the indicator draws on more than one dataset. The vehicle safety indicator links a national register of registrations to the Euro NCAP ratings, by type approval number together with year of construction where possible and by make, model and year otherwise; the post-crash-care indicator may need to bring together the timestamp of the emergency call and the timestamp of arrival from datasets held by different owners, a link that is not always possible where common identifiers are absent. The linkage method, and the proportion of records that could not be linked, should be reported.
- Calculation then applies the indicator's formula to the prepared data. The forms recur across the five indicators. Several indicators are an exposure-weighted share in which a dichotomised criterion is applied to each record and the records are weighted by their exposure:
 - the infrastructure indicator sums the exposure of road sections rated in the safe class and divides by the total exposure
 - the 30 km/h indicator sums the length of urban roads at or below 30 km/h and divides by the total urban road length
 - the vehicle safety indicator counts the cars meeting or exceeding the star threshold and divides by the total.

The enforcement indicator is instead a normalised count of controls, tickets or installed devices, expressed per head of population, and optionally per kilometre of network, per unit of area or per unit of traffic. The post-crash-care indicator is different again: it is not a percentage but a percentile of a distribution, the value below which 95% of response times fall, and a country may in addition report other percentiles for benchmarking.

A practical rule applies wherever part of the population cannot be classified such as the registered cars for which no safety rating is available: the indicator value should be reported twice, once excluding the unclassifiable records and once including them under the unfavourable assumption, so that the effect of the gap is visible rather than hidden.

- Where a full population is available, the calculation ends here and no confidence interval arises, because the value is a property of the whole population and not an estimate from a sample. Where, by contrast, the value rests on a sample, the sample must be weighted back to the national population and an interval must be reported around the estimate. Weighting is by the appropriate measure of exposure for the indicator: by the relative number of interventions per region for post-crash care, by population or traffic volume per region for enforcement. The interval is then computed by a method that respects the sampling

design; an interval that assumes simple random sampling where the design is in fact stratified or clustered will be too narrow. For a percentile rather than a mean or a proportion, the interval is not given by a closed formula and is best obtained by bootstrapping, as set out for the post-crash-care indicator. The weighting and confidence-interval framework of Chapter 8 applies in full to these sampled situations.

Disaggregation follows the same discipline as in Part B. Breakdowns by road type, by region, by time period, by vehicle type, by city or city size are valuable for diagnosis but are reported only where the underlying data are sufficient and consistent. For a database-based SPI the relevant condition is that the coverage of the breakdown is known and adequate. A breakdown drawn from a part of the system whose coverage is unknown should not be reported, and partial coverage should be stated wherever a breakdown is given.

17.5 Metadata requirements

A database-based SPI value is only as interpretable as the documentation that accompanies it. Because the indicator is a secondary use of data created for another purpose, a bare value conveys very little: two countries can report the same figure on the basis of datasets that differ in coverage, in definitions and in quality to a degree that makes the figures incomparable. The metadata is therefore an integral part of the deliverable. It should be detailed enough to allow an independent analyst to understand, and where the data permit to reproduce, the reported value.

The following elements should be documented for every database-based SPI:

- the data source or sources and the agency that owns them, distinguishing official from non-official sources
- the reference year or date to which the value applies
- the mapping between the source variables and the indicator definition, including the categorisation and re-coding decisions of any consequence
- the coverage of the data: the share of the relevant target system that they represent, what is missing, and any reason to think the covered and uncovered parts differ systematically
- the linkage method (when more than one dataset is used), and the proportion of records that could not be linked
- the cleaning rules applied and the proportion of records that were removed, or whose required fields were missing or erroneous
- the weighting and extrapolation method and, where a sample is used, the method by which the confidence interval was calculated
- any legal or institutional limitation that affected access to the data or the form in which they could be used
- any element that falls outside the standard scope of the indicator and is therefore reported separately.

Beyond the strict requirements, contextual information specific to each indicator (e.g., the number of dispatching centres and emergency units behind a post-crash-care value, or the share of the national network that the infrastructure indicator covers) aids interpretation and benchmarking and is encouraged where it is readily available.

17.6 Alternative approaches

The preceding sections describe the preferred situation, in which a suitable national database is available and the indicator can be calculated for the whole relevant population. This situation does not hold everywhere, and Part C therefore provides for a graded set of alternatives, to be used in order of preference and always with the limitation clearly flagged in the metadata:

- The first alternative is to substitute a non-official source for a missing official one: an open or commercial road database in place of a national road inventory for the 30 km/h indicator, or a crowd-maintained

camera database for the enforcement indicator. The resulting value is treated as approximate or indicative and cross-checked against whatever official figures exist.

- The second alternative is sampling, used where no national database covers the relevant population. Sampling reintroduces the apparatus of Part B. The post-crash-care indicator sets out the fullest example: where no national emergency database can be used, a sample of at least 2,000 interventions is drawn, stratified by region (for instance at NUTS 1 level), by first selecting EMS stations within each region and then taking their interventions, with the regional results weighted by the share of interventions each region represents. The enforcement indicator follows the same logic where data are available only locally or regionally, selecting locations as randomly as possible while ensuring geographical spread, and weighting the results by population or traffic volume. In every sampled case the design must be documented and the interval computed by a method that respects it.
- The third alternative is a survey of the authorities that hold the relevant information, used where neither a database nor a workable sample of records is available. The 30 km/h indicator describes such an approach: in the absence of an exploitable road database, a national authority may survey local road authorities for the length of their urban network and the share of it at or below 30 km/h. A survey of this kind requires a clear common definition (in this example of an urban road) so that the responses can be combined, and it yields a value whose quality depends on the response rate and on the consistency of the estimates that respondents provide.
- The fourth alternative is a simplified or proxy indicator that approximates the intended quantity from data that are more widely available. The infrastructure chapter offers a length-based proxy for the exposure-based indicator where traffic volumes are unavailable, and a Baseline-style indicator for the part of the network that lies outside the scope of the Directive. The 30 km/h chapter offers, where no road database exists at all, an indicator based on the number or the population of towns whose default urban speed limit is 30 km/h. The vehicle safety chapter notes the average age of the fleet, available from external sources, as related contextual information. A proxy of this kind answers a slightly different question from the indicator it stands in for, and the difference, together with the reason for resorting to the proxy, is recorded in the metadata so that the value is not mistaken for the indicator proper.

A country is encouraged to use the most demanding approach its data allow, and to move toward the preferred full-population measurement as its data infrastructure develops. Whichever approach is used, the guiding requirement is the one that runs through this whole chapter: the value is reported together with an honest account of how it was obtained and of the limits within which it can be read.

18 SPIs for infrastructure safety

18.1 Introduction

The layout, design and maintenance of a road determine much of its safety level, independently of how the road is used or what vehicles travel on it. The infrastructure indicator is intended to provide a quantified representation of that quality for a national road network. In the EU framework it is defined as the percentage of distance driven over roads with a safety rating above an agreed threshold, and it is one of the four pillars (infrastructure, vehicle, safe road use and emergency response) around which the European Commission's original set of indicators was built.

Of the five indicators treated in Part C, infrastructure is the one that rests most firmly on an established regulatory and methodological framework, and it therefore illustrates the first general principle of Chapter 17 particularly clearly. Its scope, its road type categories and its underlying assessment method all derive from Directive 2019/1936/EC on road infrastructure safety management (RISM) and from the Network-Wide Road Safety Assessment (NWA) methodology prepared for the Commission and endorsed by the Expert Group on Road Infrastructure Safety (EGRIS) in November 2022 (European Union, 2023). An indicator grounded in this way enjoys a stable definition over time and a degree of cross-country comparability that ad hoc data collection cannot match.

The framework also fixes the indicator's boundaries. Its scope is limited to the road network for which network-wide road safety assessments are compulsory under the Directive, which generally does not include secondary rural roads and urban roads other than urban motorways. The safety of urban networks is addressed separately, through the experimental 30 km/h indicator of Chapter 19, and the part of the national network that falls outside the Directive is addressed through a complementary indicator described in Section 18.6. This chapter is concerned with the network within the scope of the Directive.

18.2 Indicator definition and scope

18.2.1 Definitions

Two definitions of the infrastructure indicator are used, differing only in the measure of exposure they apply. The primary definition expresses the indicator in terms of distance driven:

Percentage of the distance driven over roads rated in terms of safety as "class 1" of the EGRIS Network-Wide Assessment (NWA) proactive methodology.

Distance driven is the preferred measure because it allows the safety rating of high-volume roads to carry more weight in the national figure than that of low-volume roads, which is the behaviour an exposure-based indicator should have. Traffic volumes are not, however, available for all roads in all countries, and a simplified definition is therefore also provided, in which road length stands in as a rough proxy for exposure on the assumption that traffic volume is equal on all roads:

Percentage of the road network length rated in terms of safety as "class 1" of the EGRIS Network-Wide Assessment (NWA) proactive methodology.

In both definitions the percentage is calculated relative to the total exposure on all roads within the scope of the Directive.

18.2.2 Road network and road types within scope

The network within scope comprises the roads for which the Directive requires network-wide assessment: roads forming part of the trans-European road network, motorways (rural and urban), other primary roads (the roads immediately below motorways in a country's functional classification) and other roads outside urban areas that do not serve bordering properties and that were completed using Union funding. Countries provided the Commission with a list of their motorways and primary roads at the end of 2021.

For the purpose of the EGRIS proactive safety assessment, and therefore for the estimation of the indicator, four road types are distinguished:

- rural motorways;
- urban motorways;
- primary divided roads;
- primary undivided roads.

The indicator is estimated separately for each of these four types and for all roads in scope combined. Urban roads other than urban motorways, and minor and local rural roads, lie outside the scope: no commonly agreed European methodology yet exists for their network-wide assessments.

18.3 Data requirements and sources

For the calculation of the SPI values two datasets are required, both at the level of the individual road section and both separated by road type: the safety rating produced by the proactive assessment, and the traffic volume on the section. Their nature and origin are described in turn, followed by the quality and coverage considerations that apply to them.

18.3.1 Safety ratings from the NWA-proactive methodology

The safety classes on which the indicator is built come from the proactive component of the EGRIS NWA methodology (https://road-safety.transport.ec.europa.eu/news/new-guidelines-assess-safety-road-infrastructure-2023-01-16_en). The full methodology has two components applied over the same network: a proactive assessment of the in-built safety of roads, and a reactive assessment based on the analysis of crash occurrence, the two being integrated to give the final network rating. However, the Trendline infrastructure SPI indicator uses the proactive component only. This is a deliberate choice: an SPI should describe the conditions and processes that lead to crashes rather than the crashes themselves. The integrated result, which leans heavily on crash analysis, would blur the distinction between the SPI and the outcome statistics it is meant to complement.

In the proactive assessment each road section is given a score on a scale from 0 to 100, where an ideally safe section scores 100. The score is built up from a set of Reduction Factors (Equation 11), each corresponding to a design or operational characteristic of the road and each taking a value greater than zero and at most one, with one representing the safest condition.

Equation 11. Proactive assessment safety score.

$$Score = 100 \cdot RF_{1i} \cdot RF_{2i} \cdot \dots \cdot RF_{ni}$$

Different characteristics, and therefore different Reduction Factors, are used for motorways and for primary roads, reflecting their different design and operation; the parameters also differ between rural and urban motorways and between divided and undivided primary roads:

- For motorways, the in-built safety parameters are lane width, roadside characteristics (including clear zone width, obstacles and presence of barriers), curvature, interchange spacing, conflicts between

vulnerable road users and motorised traffic, and the presence of traffic operation centres or comparable mechanisms to inform users of incidents.

- For primary roads, the parameters are lane width, roadside, curvature, density of property access points, intersections, conflicts between vulnerable road users and motorised traffic, shoulder type and width, the presence of passing lanes, and the quality of signs and markings.

Several of these parameters are treated differently for urban and rural motorways and for divided and undivided primary roads, reflecting the genuine differences in their design and operation. In addition to the parameters used to compute the score, operational characteristics such as Annual Average Daily Traffic, the speed limit, the presence of automated speed enforcement and, where available, the operating speed V85, may modify the safety scoring or the final ranking of sections.

Because the parameters and scores differ, scores are not comparable between road types. On the basis of its score, each section is assigned to one of three safety classes, as shown in Table 18-1.

Table 18-1. Safety classes of the NWA-proactive methodology.

Safety class	Risk level	Colour code
Class 1	Low risk	Green
Class 2	Intermediate risk	Yellow
Class 3	High risk	Red

The use of the EGRIS methodology is not mandatory for meeting the requirements of the Directive. Where a country applies a different but compatible national assessment method, it is strongly recommended that proactive EGRIS results be provided as well, so that a comparable indicator can be produced for all countries (e.g., by running the EGRIS proactive assessment in parallel).. Another option that may be considered is to develop and apply a conversion procedure to obtain EGRIS compatible NWA proactive classes from national road safety assessment results. In this case however, both the original and the converted datasets need to be provided, together with an explanatory note describing the national methodology, the conversion procedure and any assumptions it required (see also paragraph 18.3.3).

The result of the assessment is a section-level database with the structure shown in Table 18-2. Each record describes one road section: its number, its road type, its length, its proactive score and its resulting safety class. This database is the starting point for the estimation in Section 18.4.

Table 18-2. Indicative structure of the NWA-proactive results database.

Section number i	Road type	Section length L_i (km)	NWA-proactive score (%)	Safety class S_i
1	Rural motorway	0.600	87	1 (low risk)
2	Urban motorway	0.400	74	2 (intermediate risk)
3	Primary divided	0.800	45	3 (high risk)
4	Primary undivided	1.200	41	3 (high risk)
...	...	...	...	...

18.3.2 Traffic volume data

The primary, exposure-based definition additionally requires the traffic volume on each road section, expressed as the Annual Average Daily Traffic (AADT) for the year in which the assessment was carried out and measured in vehicles per day. The count is of actual vehicles; it should not be weighted by vehicle class to give passenger-car equivalents. Where a section has been assessed per direction of travel, as is typical for motorways and divided primary roads, the AADT for that direction is used; where it has been assessed for both directions together, as is typical for undivided primary roads, the AADT for both directions is used.

Traffic volumes may be obtained from national mobility data, from toll-station counts on tolled roads, or from traffic counts on a representative sample of sections. Where volumes are available only for part of the network, a national estimate may still be derived by sampling, provided the sample is sufficiently representative and appropriate weighting factors can be defined. In that case a point estimate and a 95% confidence interval should both be reported. Where no traffic data are available and none can be estimated, the length-based definition is used instead (see Section 18.4.2).

18.3.3 Data quality and coverage

Countries were expected to complete a first network-wide assessment of their network by the end of 2024 and to repeat it at least every five years; these assessments provide the basis for the indicator. The two datasets are delivered disaggregated at section level and separately by road type, which allows an independent analyst to reproduce the value. Where privacy or other legal restrictions prevent the release of section-level data, aggregated values are provided at the highest level of disaggregation possible.

Where a national assessment method has been used in place of the EGRIS method and a conversion applied to obtain EGRIS-compatible classes, both the original and the converted datasets are provided, together with an explanatory note describing the national assessment methodology, the conversion procedure and any assumptions it required. As set out in the common framework (Chapter 17), the value is accompanied by metadata documenting the source, the coverage of the network, and how the AADT figures were derived. Comparison across years requires that the definitions and the assessment behind the figures have remained consistent.

18.4 Estimation of the SPI

18.4.1 Case 1: exposure-based estimation

The primary definition is estimated by the following procedure, starting from the results database of Table 18-2:

1. Identify all roads within the scope of the Directive and classify each into one of the four road types of Section 18.2.2.
2. For each road, carry out the EGRIS NWA-proactive assessment, producing for each section its length (L_i in km), its score and its safety class (class 1: low risk, class 2: intermediate risk, class 3: high risk).
3. For each section, obtain the AADT (in veh/day), denoted T_i , for the year of the assessment.
4. For each section, compute the exposure factor E_i as the product of section length and traffic volume (Equation 12):

Equation 12. Exposure factor

$$E_i = L_i \cdot T_i \cdot 365 \quad (\text{veh.km / year})$$

5. Dichotomise the safety rating by introducing a binary variable DS_i that distinguishes sections meeting the rating criterion from those that do not:
 $DS_i = 1$ if the section is in class 1 (low risk)
 $DS_i = 0$ if the section is in class 2 (intermediate risk) or class 3 (high risk)
6. The database of Table 18-2 now carries, for each section, three further columns: the AADT T_i , the exposure factor E_i , and the dichotomised rating DS_i .
7. Compute the indicator as the share of exposure on class 1 sections in the total exposure of the road type (Equation 13)

Equation 13. Computation of the Infrastructure SPI.

$$SPI(\%) = 100 \cdot \frac{\sum_{i=1}^N (E_i \cdot DS_i)}{\sum_{i=1}^N E_i}$$

where N is the number of sections of the road type within scope of the Directive. Following this procedure, five values are produced: one for each of the four road types and one for all roads within scope combined.

18.4.2 Case 2: length-based estimation

Where traffic volumes are not available, the simplified definition is estimated by exactly the same procedure, with the single difference that the exposure factor in step 4 is set equal to the length of the section, in other words: $E_i = L_i$

The dichotomisation and the final formula are unchanged, so that the indicator becomes the share of class 1 length in the total length of the road type. As in Case 1, five values are produced — one per road type and one for all roads in scope of Directive 2019/1936/EC.

18.5 Minimum reporting requirements

The minimum requirement is the set of five Case 1 values: the SPI value for rural motorways, for urban motorways, for divided primary roads, for undivided primary roads, and for all roads within the scope of the Directive 2019/1936/EC. Where traffic data are unavailable and cannot be estimated, the corresponding five Case 2 values may be reported instead. Where traffic volumes have been obtained by sampling, both the point estimate and its 95% confidence interval are reported.

Beyond this minimum, countries are encouraged to provide the cleaned section-level NWA-proactive dataset alongside the values, since it allows the indicator to be recalculated and checked, and to document how the AADT figures were derived. Disaggregation by region (e.g., at NUTS 1 level) is optional. Table 18-3 summarises the minimum and the recommended elements.

Table 18-3. Minimum and optional reporting requirements for the infrastructure indicator.

Aspect	Minimum requirement	Recommended / optional
Indicator values	Case 1 for the four road types and for all roads in scope (five values)	Same values broken down by region (e.g. NUTS 1)
Fallback	Case 2 (length-based) where traffic volumes are unavailable	Case 2 SPI values broken down by region.
Uncertainty	95% confidence interval where traffic volumes are sampled	—
Background data	<i>(In case a national assessment method is used and conversion is applied)</i> Original and converted NWA -proactive section-level datasets and explanatory note on national methodology, conversion methodology and assumptions	Cleaned NWA-proactive section-level dataset
Metadata	Source and reference year; coverage of the Directive network; method of deriving AADT	95% confidence interval of AADT estimates
Complementary	—	Complementary indicator for non-Directive roads; share and composition of the network (Section 18.6)

18.6 Complementary indicators and alternative approaches

18.6.1 Complementary indicator for roads outside the Directive

The indicator proper covers only the network within the scope of the Directive. Countries that wish to characterise the rest of their network may use the complementary indicator developed in the Baseline project, defined as:

Percentage of the road network length of roads either with opposite traffic separation (by barrier or area) or with a speed limit equal to or lower than xx km/h, in relation to the total road network length.

The threshold of the safety rating is here assumed to be met when a road either has opposite-traffic separation or has a speed limit at or below a defined value. The speed-limit threshold is not prescribed; in line with Safe System principles the Baseline project proposed:

- 30 km/h where collisions between a vulnerable road user and a motorised vehicle are possible
- 50 km/h for built-up roads where vulnerable road users are separated from motorised traffic and for inter-urban roads with a risk of right-angle collisions
- 70 km/h for roads with a risk of head-on collisions, while leaving the final choice to the country.

To compute the indicator, road length is used as the proxy for exposure and roads are classified into three groups: those at or below the speed-limit threshold (RL), those above the threshold without opposite-traffic separation (RH), and those above the threshold but with opposite-traffic separation (RS). Denoting the lengths of sections of each type accordingly, the indicator is computed as shown in Equation 14.

Equation 14. Complementary infrastructure indicator

$$SPI = \frac{\sum_1^N LRL_i + \sum_1^N LRS_i}{\sum_1^N LRL_i + \sum_1^N LRH_i + \sum_1^N LRS_i}$$

It is useful to report the components of this figure as well — the share of separated roads (RS) and the share of other roads with a safe speed limit (RL) in the total network length — and, where the threshold is differentiated by road type, the proportions of safe-speed length within each band, for example the share of RL30 length within the combined RL30 and RH30 length. Two cautions apply: no Trendline funding is available for the data collection behind this indicator, and because speed limits change frequently the underlying inventory must be kept up to date.

18.6.2 Other complementary indicators and breakdowns

Two further descriptive indicators help to place the main value in context:

- the share of the network within the scope of the Directive in the total national road network length
- the share of each road type (rural motorways, urban motorways, primary divided, primary undivided and urban roads) in that total.

These values, and any regional breakdown a country chooses to provide, complement the indicator but do not substitute for it, and they are added to the Trendline database as supplementary information rather than reported in its place.

19 SPIs on 30 km/h roads in urban areas

19.1 Introduction

Speed influences both the risk that a crash occurs and the severity of its outcome, and a safe travel speed is one of the pillars of the Safe System approach. Where motorised traffic, pedestrians and cyclists share the same space and conflicts between them are possible, a speed of 30 km/h or less is widely regarded as the threshold below which such conflicts are survivable. The share of the urban network on which that threshold applies is therefore a meaningful indicator of how far a country's urban infrastructure is aligned with the Safe System.

The lowering of urban speed limits to 30 km/h has gained considerable political momentum. It was among the specific recommendations of the 2020 Stockholm Declaration, subsequently reflected in a United Nations General Assembly resolution, and the European Parliament has called for a default limit of 30 km/h in residential areas and wherever cyclists and pedestrians are numerous. A growing number of cities have introduced city-wide 30 km/h limits, and a few countries have legislated for them nationally. Against this background the indicator described in this chapter measures, for a given year, the share of the urban road network on which the speed limit is 30 km/h or lower.

Two considerations are important to note. First, the target value of this indicator is not 100%: a functioning urban network needs main arterials at 50 km/h or 70 km/h, designed so that those higher speeds are themselves safe, to keep through-traffic out of residential streets. A possible target could be around 85% but the optimal target needs still to be researched. Second, the indicator complements the infrastructure indicators of Chapter 18, which covers the network within the scope of Directive 2019/1936/EC and excludes urban roads other than urban motorways.

19.2 Indicator definition and scope

19.2.1 Definition

The recommended definition of the indicator is the share of 30 km/h road length in the total length of urban roads:

Share of 30 km/h road length in the total length of urban roads.

In other words, the computation is as shown in the following equation:

Equation 15. Computation of the 30KM/h SPI.

$$SPI_{30} = 100 \cdot \frac{\text{total length of urban roads with a speed limit} \leq 30 \text{ km/h}}{\text{total length of all urban roads}}$$

The indicator is expressed as a percentage and refers to a given year. Length is used as the measure of exposure; alternative exposure measures, which answer slightly different questions, are discussed among the additional definitions in Section 19.6.

19.2.2 What counts as an urban road

An urban road, for the purpose of this indicator, is a public road inside the urban boundary signs. Urban motorways and urban expressways are excluded, because cyclists and pedestrians are not permitted on them; urban motorways are in any case within the scope of the infrastructure indicator of Chapter 18. The decisive notion is therefore the urban or built-up area, which is normally delimited by built-up-area signs placed on

every road entering it and within which the default speed limit is, in many countries, 50 km/h. For countries that do not use built-up-area signs systematically, a country-specific definition based on population density may be used instead, so that the set of urban areas is as complete and as comparable with other countries as possible.

Three further conventions keep the scope consistent across countries:

- Woonerfs, bicycle streets and pedestrian zones are included and treated as equivalent to roads with a limit of 30 km/h or lower.
- Footpaths and solitary cycle paths are excluded, for simplicity.
- Private roads are excluded.

19.3 Data requirements and sources

19.3.1 Data required for the basic indicator

The basic indicator requires two quantities: the total length of urban roads with a speed limit of 30 km/h or lower, and the total length of all urban roads. In almost all cases these are extracted from a database (or a combination of databases) of road characteristics. Two conditions must therefore be met by the data: (1) it must be possible to identify the urban roads, and (2) for those urban roads the database must hold the speed limit and the length of each road section.

How the urban roads are selected depends on what is available. Where geographical data on built-up-area boundaries exist, they can be combined with a digital road map to select the urban roads; where built-up-area signs are not used systematically, a Geographical Information System (GIS) analysis can delineate the urban areas before the same combination is made; and where neither is available but urban and rural roads carry different default limits, the speed limit itself can serve as the selection criterion. The total length of urban roads is then the sum of the lengths of the sections that fall within the urban areas, and within that set the sections at or below 30 km/h are identified and their lengths summed. Table 19-1 shows the structure such a working dataset takes.

Table 19-1. Example structure of a working dataset for the 30 km/h indicator.

Section number	Section length (km)	Speed limit (km/h)	≤ 30 km/h (1 = yes)
1	0.600	30	1
2	0.400	20	1
3	0.800	50	0
4	1.200	60	0
...	...	...	...

19.3.2 Data sources on road characteristics

Three kinds of source can supply these characteristics:

- Many road authorities keep records of the roads in their jurisdiction, including section length, the number of lanes and speed limits.
- Open data sources, in particular OpenStreetMap, provide speed limits and road characteristics across many cities in many countries in a single database.
- Commercial providers such as TomTom and Google hold comparable information.

As set out in the common framework (Chapter 17), official sources are the primary reference, and non-official sources are used to fill gaps. OpenStreetMap has the advantage of being free and of bringing many cities

together, but its completeness and quality vary across countries, regions and cities; commercial data tend to be of higher quality but must be purchased. Where a non-official source is used, the source and any cross-checks should be documented.

A survey among the countries participating in the first phases of the Trendline programme found that only a limited number are at present able to calculate the indicator from a national database in which urban roads can be identified and the speed limits are sufficiently complete and up to date. Some countries without a national database have databases held by local authorities. Where almost all local authorities maintain such a database, or comparable regional databases cover the whole country, the data can be combined into a national figure. Where, on the other hand, databases exist only for some cities or regions, combining them is hazardous: the share of 30 km/h roads can differ considerably between places, and a set of cities that happens to hold data is not a random sample. In that situation it is recommended to restrict the indicator to the cities or regions for which data exist and not to report a national value, rather than to extrapolate from the covered to the uncovered part of the network.

The quality of the indicator depends above all on the speed-limit information being complete and current. How current it needs to be depends on the pace of change in a country: where cities are actively lowering limits across parts of their network, the speed-limit data should be updated as changes occur, or at least once a year. In line with the common framework, the share of the urban network that the data cover is reported, together with any reason to think that the covered and uncovered parts differ systematically.

19.4 Estimation of the SPI

The estimation follows directly from the data preparation described above:

1. Select the urban roads, by whichever of the methods of Section 19.3.1 the available data allow.
2. Determine the total length of urban roads as the sum of the lengths of the selected sections.
3. Within the urban roads, select the sections with a speed limit of 30 km/h or lower, treating woon-erfs, bicycle streets and pedestrian zones as belonging to this group.
4. Determine the total length of those sections.
5. Compute the indicator as the ratio of the two lengths, expressed as a percentage (Equation 15)

19.5 Minimum reporting requirements

The minimum requirement is the share of 30 km/h road length in the total length of urban roads, expressed as a percentage and for a stated year, accompanied by documentation of the data on which it rests: the database or databases used, the way urban roads were identified in them, and an indication of data quality — whether the speed limits are correct and up to date, and whether all relevant urban roads are included. Breakdowns by city, by region or province, and by city size are optional but valuable for diagnosis. Table 19-2 summarises the minimum and the optional elements.

Table 19-2. Minimum and optional reporting requirements for the 30 km/h indicator.

Aspect	Minimum requirement	Recommended / optional
Indicator value	Share of 30 km/h road length in total urban road length (%), for a stated year	Breakdowns by city, by region/province, or by city size
Documentation	Database(s) used; how urban roads were identified; indication of data quality	—
Fallback	—	Survey of road authorities; town-based proxy (number or population of towns with a default limit ≤ 30 km/h)

19.6 Complementary indicators and alternative approaches

19.6.1 Additional definitions using the same dataset

Where the dataset is rich enough, several additional definitions can be derived from it. The first group replaces road length with a different measure of **exposure**:

- **Lane length instead of road length.** The share of 30 km/h lane length in the total urban lane length requires, in addition to the length of each section, the number of lanes or the total lane length.
- **Distance travelled.** The share of distance travelled on 30 km/h roads in the total distance travelled on urban roads requires traffic volumes: the distance travelled on a section is its AADT multiplied by its length.

These measures reflect actual exposure more closely than plain road length does, but they are more demanding of data and they carry a counter-intuitive property that should be kept in mind: adding lanes to a 30 km/h road, or shifting traffic from 50 km/h roads onto 30 km/h roads, improves the score, even though the policy intention is the opposite (traffic should use the 50 km/h network as far as possible, since 30 km/h roads are mainly residential).

The second group starts from the concept of a **safe speed**. A speed above 30 km/h need not be unsafe where vulnerable road users are absent or are physically separated from motorised traffic, and two definitions follow from this:

- The share of 30 km/h road length in the total length of urban roads with mixed traffic restricts. This restricts the denominator to roads on which pedestrians or cyclists are present and not physically separated or are expected to cross at road sections.
- The share of urban roads with a safe speed limit. This counts roads as safe that either have a limit of 30 km/h or lower, or do not admit vulnerable road users, or keep them physically separated and away from crossing points.

Both definitions require judgements about where mixed traffic and physical separation occur that are not straightforward; the recommendation is to make pragmatic, clearly described choices. This safe-speed reasoning is the same as that underlying the speed-limit thresholds of the complementary infrastructure indicator (Section 18.6.1).

19.6.2 Alternative approaches in the absence of a national database

Where the indicator cannot be calculated from existing databases, two alternatives are available. The first is a survey of road authorities. A national authority may ask local road authorities for the estimated total length of their urban network, the estimated length of it at or below 30 km/h, and the resulting share. The survey

must state clearly which roads are to be treated as urban, following the same logic as the database approach (a public road inside the urban boundary signs that is not a motorway) and respondents should be asked to explain briefly how they arrived at their estimates. A survey of this kind has been used in the Netherlands, and its results proved consistent with later database-based estimates.

The second is a town-based proxy, used where not even a survey of network lengths is feasible. Here the indicator is based on the number of towns whose default urban speed limit is 30 km/h or lower; a variant weights the towns by their population. The two formulas are shown below:

Equation 16. Alternative 30 Km/h indicators.

$$SPI_{30a} = \frac{\text{number of towns with a default urban speed limit} \leq 30 \text{ km/h}}{\text{total number of towns}}$$

$$SPI_{30a} = \frac{\text{population of towns with a default urban speed limit} \leq 30 \text{ km/h}}{\text{total population of the country}}$$

Both indicators are expressed as percentages and are best reported with a breakdown by town size, for example below 50,000 inhabitants; between 50,000 and 249,000; and 250,000 or more. A town-based proxy answers a coarser question than the length-based indicator, and the simpler survey it rests on need only ask whether the default limit within the built-up areas of a municipality is 30 km/h, checking that the default applies to all of those areas. As with any proxy, the value is flagged as indicative and the reason for resorting to it is recorded in the metadata.

20 SPIs for vehicle safety

20.1 Introduction

Cars with better passive and active safety systems both protect their occupants and surrounding vulnerable road users in a crash and help to avoid the crash in the first place. The vehicle safety indicator is intended to capture how far the cars in use in a country embody this protection. It does so through the Euro NCAP safety rating, the established European measure of a new car model's safety performance. In the EU framework the indicator is defined as the percentage of new passenger cars with a Euro NCAP safety rating at or above an agreed threshold, and vehicle safety is one of the four challenge areas around which the original set of indicators was built.

What makes this indicator methodologically distinctive, and what shapes the whole of this chapter, is that it cannot be drawn from a single dataset. Where the SPIs on infrastructure 30 km/h each rest on one road dataset, the vehicle safety indicator is formed by combining elements from two separate databases held by two different owners:

- a national register of the passenger cars registered in the country, held by the national vehicle-registration authority
- the database of Euro NCAP safety ratings, held by Euro NCAP.

Neither database alone yields the indicator. The registration data say how many cars of each make and model are on the road but nothing about their safety rating; the Euro NCAP data give a safety rating for each model tested but say nothing about how many such cars a country has. The indicator exists only once the two are linked. The central methodological effort therefore goes into the linkage of two administrative datasets and into the coordination with the two data owners that supply them.

An important consideration is that, although it is likely, it has not yet been demonstrated that a higher national score on the safety rating of the fleet translates into better national road safety performance. The relationship between Euro NCAP ratings and real-world outcomes has been shown for the passive safety of older models, but a general evaluation that also takes account of modern active safety features has not been carried out. The indicator should therefore be read as a measure of a system-level capacity rather than as a direct measure of safety outcomes.

The chapter sets out the standard indicator — based on the cars newly registered in a year — and two recommended extensions of it to a larger part of the fleet, which are described in Section 20.6.

20.2 Indicator definition and scope

20.2.1 Definition and thresholds

The standard indicator is defined as:

Percentage of new passenger cars with a Euro NCAP safety rating equal to or above a predefined threshold.

Two thresholds are used in parallel: a minimum threshold of four Euro NCAP stars and a strong threshold of five stars. The indicator is reported for the cars registered in the reference year. A Euro NCAP rating is given on a five-star scale and is valid for six full years after the year of the test; the Euro NCAP coverage area comprises the 27 EU Member States and the United Kingdom.

Two further indicators extend the same idea to a larger part of the fleet and are described in Section 20.6: one covering all passenger cars aged between zero and five years, and one covering the whole passenger-car fleet.

The latter requires rescaled ratings and is treated separately because Euro NCAP ratings expire after six years; however, this rescaling is beyond the scope of Trendline.

20.2.2 Vehicles included and excluded

A passenger car, for the purpose of the SPI, is a motor vehicle with three or four wheels used mainly to carry people and with seating for no more than eight occupants; taxis and motor caravans are included (see also the vehicle terminology in Section 3.5). Two categories of passenger car are excluded, from both the numerator and the denominator, because they cannot be linked to a Euro NCAP rating and their exposure on the road is small:

- cars adapted for drivers with disabilities, which have been modified to an extent that the Euro NCAP assessment of the original model no longer applies and which are re-approved under a new approval number
- passenger cars derived from commercial vehicles or vans but registered as passenger cars, such as a van fitted with passenger seats, which are tested by Euro NCAP but not given a star rating.

Cars adapted for drivers with disabilities take two forms: those that acquire a new make and model name, and those that keep the original make and model but are distinguished by a different approval number. Where a country cannot identify them, for instance because its data hold no approval numbers, it is acceptable to leave them in the indicator, provided this is recorded in the metadata.

20.3 Data requirements and sources

The indicator draws on two datasets from two different owners. They are described in turn, followed by the quality and coverage considerations that apply once they are brought together.

20.3.1 The vehicle-registration data

The first dataset is the national record of registered passenger cars, giving the number of cars by make and model and, from it, the total number of cars. Where possible it should also carry, for each car, the type approval number and the year of construction, because both improve the accuracy of the linkage described later on in Section 20.4.1. This dataset is obtained from national sources, typically the public authority or agency that registers cars. It may also be available from national associations and federations or from international commercial providers such as JATO and Carsalesbase, from which it may have to be purchased. Figures from different sources can differ slightly, depending on the classification and counting method used.

20.3.2 The Euro NCAP ratings data

The second dataset is the database of Euro NCAP safety ratings, held by Euro NCAP and made available to the Trendline partners in spreadsheet form. For each tested model it gives the make, the model, the type approval number, the year of the test, the year production started, the Euro NCAP vehicle category, the star rating and the underlying passive and active scores. Because ratings are valid for six years, the database provided to the Trendline partners covers models tested for the last from 2016 onward.

Where the rating of a model differs between a standard version and a version with an additional safety pack, the rating for the standard version is used, and the general rule is that a country uses the single rating given for each model in the supplied list rather than attempting to reflect version-by-version differences, which would make the calculation disproportionately complex.

20.3.3 Data quality and coverage

The coverage objective differs by indicator: the standard indicator should cover all cars newly registered in the reference year, the 0–5-year indicator all cars of that age on the reference date. Two recurring gaps should be

documented in the metadata: cars missing from the registration data because it is incomplete, and cars that are present but cannot be linked to a Euro NCAP rating, each expressed as a proportion of the total.

A further, structural difficulty affects comparability: for tax reasons some manufacturers register cars in one country and immediately export them as used cars with zero kilometres to another. No minimum-age cut-off is applied to deal with this, because the necessary information is rarely accessible, but where it is available a country may recalculate the indicator taking exported and imported zero-kilometre cars into account and report the result alongside the standard value.

20.4 Estimation of the SPI

The estimation of the SPI requires two stages: linking the two datasets, which is the substance of the work, and then a straightforward calculation on the linked result.

20.4.1 Linking the two datasets

The national registration dataset and the Euro NCAP ratings are linked using whichever of three keys the registration data support, in order of preference:

- by type approval number together with the year of construction
- by make and model together with the year of construction
- by make and model together with the year of registration, used as a proxy for the year of construction.

The linking method actually used is recorded in the metadata. The type approval number is the most precise key. It has a structured form (for example E1*2007/46*1892*06) in which the last two digits identify a particular version of the model. Since the Euro NCAP rating is valid for all versions, these last two digits are ignored when linking, and the approval numbers in the supplied Euro NCAP list are likewise given without them.

Not every record can be linked by approval number, for three reasons: some approval numbers are used for more than one model; cars approved individually carry a specific approval number that cannot be matched; and the Euro NCAP list does not give an approval number for every model. The records that remain are linked by make and model together with the year of construction. The year of construction matters because a car registered in a given year may be a model of that year or of the year before, with different ratings. When the year of construction is not available, the year of registration is used as a proxy and this is noted in the metadata.

Linking by make and model is less simple than it may appear, because the make-and-model scheme used by Euro NCAP is a simplification of the formal definition of a vehicle by make, type, variant, version and commercial description. It is often necessary to build a conversion table between the commercial names in the national database and the model names used by Euro NCAP.

Table 20-1 shows part of such a table. Finally, where the Euro NCAP list holds two ratings for the same model (e.g., a model retested some years after its first test) the rating matching the car's year of construction is used.

Table 20-1. Example of a conversion table linking national commercial names to Euro NCAP model names.

Make	Commercial name (national data)	Model (Euro NCAP database)
Audi	A1 Sportback	A1
BMW	420i	4-Series
Mercedes-Benz	A 160	A-Class
Nissan	Nissan Qashqai	Qashqai
Renault	Clio	Clio
Seat	Arona	Arona
...	...	...

20.4.2 Calculation of the indicator

Once each registered car carries a Euro NCAP rating, the indicator is the share of cars whose rating meets or exceeds the threshold, with the threshold set at four stars and at five stars in turn (see Equation 17).

Equation 17. Computation of Vehicle Safety SPIs

$$SPI(\%) = 100 \cdot \frac{\text{number of passenger cars with a Euro NCAP rating} \geq \text{threshold}}{\text{total number of passenger cars}}$$

The treatment of cars that could not be linked to a rating gives rise to two versions of each value, both of which are required. In the first version those cars are ignored, so that the indicator is calculated on the rated cars only; in the second they are retained in the denominator and treated as falling below the threshold. The second version is therefore always somewhat lower than the first. The standard indicator thus comprises four percentages for the reference year: the four-star and five-star thresholds, each in the ignoring and the including version.

20.5 Minimum reporting requirements

For the standard indicator the values to be delivered are:

- the counts of cars (the total, the number with zero-to-three, four and five stars, and the number that could not be linked to a rating), given once ignoring and once including the unlinked cars
- the four indicator percentages described in the previous section.

These values are accompanied by metadata including:

- the source of each of the two datasets
- the cars or types of car missing from the registration data and their estimated share
- the cars that could not be linked and their share
- a description of the linking method used
- a note on whether and how cars adapted for drivers with disabilities were excluded
- any difficulties met in linking
- the share of new cars in the total fleet.

Table 20-2 summarises the minimum and the optional elements.

Table 20-2. Minimum and optional reporting requirements for the vehicle safety indicator.

Aspect	Minimum requirement	Recommended / optional
Indicator values	Standard indicator: four-star and five-star thresholds, each ignoring and including unlinked cars, for the reference year	The 0–5-year and whole-fleet indicators (Section 20.6); a second reference year
Counts	Total cars; numbers at 0–3, 4 and 5 stars; number not linkable, ignoring and including unlinked cars	
Metadata	Source of each dataset; linking method; missing and unlinkable cars as shares; treatment of disability-adapted cars; share of new cars in the fleet	Recalculation accounting for zero-kilometre imports and exports

20.6 Complementary indicators and alternative approaches

20.6.1 Extending the indicator to the fleet

The standard indicator covers only the cars registered in the reference year. Two recommended extensions widen it to a larger part of the fleet. The first covers all passenger cars aged between zero and five years on the reference date, including those imported at that age; it follows exactly the same linking and calculation procedure as the standard indicator and is straightforward to produce, because Euro NCAP ratings remain valid for six years. The population is the cars of that age still registered on the reference date, which is not the same as all cars registered over the preceding five years, since the latter would include cars no longer in the country. The computational formula is the same as in Equation 17, but now including the six most recent years (rather than just the last year).

The second extension, not yet implemented in Trendline but expected to become so over the next years, covers the whole passenger-car fleet, regardless of age. Because Euro NCAP ratings expire after six years, the current ratings cannot be linked to older cars, and a rescaled set of ratings is required instead. Attempts for such rescaling have been made, but there is no consensus yet over an approved rescaling system that could be applied across Europe.

But once such a rescaling becomes officially available, the national register of all cars (still) registered in the country (by make, model and year of first registration) could then be linked to the rescaled ratings in the same way as before (Equation 18). Two values are reported for this indicator: the mean rescaled rating of the fleet, and the share of cars with a rescaled rating of four or five stars.

Equation 18. Proposed vehicle fleet safety indicator.

$$\text{Mean fleet rating} = \frac{\sum(\text{rescaled Euro NCAP rating over all cars})}{\sum(\text{cars})}$$

As with the standard indicator, both the mean and the share should be reported once ignoring and once including the cars that could not be linked.

20.6.2 Related contextual indicators

Two further measures provide context without forming part of the indicator. The average age of the passenger-car fleet, available from external sources such as Eurostat or ACEA, was considered as an alternative

indicator in the Baseline project but is retained only as contextual information. The proportion of roadworthy cars was likewise considered and set aside, because the evidence linking periodic technical inspection to road safety outcomes is inconclusive. Neither substitutes for the vehicle safety indicator, but either may be reported alongside it where readily available.

21 SPIs for post-crash care

21.1 Introduction

Post-crash care refers to the medical treatment given after a crash, whether at the scene, during transport or subsequently. This activity has a direct bearing on whether an injury proves fatal. A large share of deaths from road crashes occurs within minutes, at the scene or in transit, and evidence cited in the EU framework suggests that shortening the interval between a crash and the arrival of emergency medical services from around 25 to around 15 minutes could reduce deaths substantially. Emergency response is, accordingly, one of the four challenge areas around which the original set of indicators was built, and the speed of the emergency response is the aspect of it that this indicator captures.

The indicator measures the response time of the emergency medical services and reports it as a single summary value, the 95th percentile. Two features mark it out from the other database-based SPIs and shape this chapter. First, it is not a share or a count but a percentile of a distribution of times, which changes how it is calculated and, in particular, how uncertainty around it is expressed. Second, although the natural data source is a single operational database, the emergency medical services, and although a complete national database yields the indicator as a full-population value, this is the one indicator in Part C for which a fully developed sampling fallback is provided, to be used where no national database can be accessed. That fallback brings the weighting and confidence-interval machinery of Part B back into play, and Section 21.4.2 refers to it rather than restating it.

Several factors beyond the response time itself influence the eventual outcome of post-crash care: the quality of medical treatment delivered at the scene, the speed and quality of transport to a hospital, the trauma-care capacity of the receiving hospital, and the longer-term rehabilitation system. The SPI does not attempt to capture these other factors, and it should not be over-interpreted as a comprehensive measure of post-crash care quality. It captures one specific, internationally measurable, time-based component of that broader system, and it is most informative when read alongside other contextual data on the EMS structure and the hospital network.

21.2 Indicator definition and scope

21.2.1 Definition

The original indicator is defined as:

Time elapsed, in minutes and seconds, between the emergency call following a road crash resulting in personal injury and the arrival of the emergency services at the scene of the crash, reported as the value of the 95th percentile.

The response time so defined runs from the emergency call, as received at the dispatching centre, to the arrival of the emergency services at the scene. It does not include the interval between the crash itself and the emergency call, which is outside the scope of the indicator. The value is reported in minutes and seconds and is representative of all road crashes across the whole territory of the country.

Following the experience within Baseline and the first phase of Trendline, it is recommended to also calculate the 50th percentile (also known as the median value).

21.2.2 Which interventions count

Turning a record of emergency interventions into the indicator population requires a sequence of selections, because an emergency database records far more than the indicator needs:

- Only interventions relating to road crashes are counted. In many countries these are a minority of all emergency interventions, and their response times may differ from those of interventions as a whole, so the indicator is derived from road-crash interventions alone. A road crash here is a crash on a public road involving at least one moving vehicle, whether or not a motorised vehicle is involved.
- Only services that provide medical pre-hospital care are counted. This includes the emergency medical services, and, where they provide such care, the fire and rescue services or the police. Both advanced and basic life support are included, with and without an emergency physician on board.
- Only the first emergency medical unit to arrive at a scene is counted; records for any second or later unit should be removed.
- The timestamp for the call is the moment the call begins, not the moment it ends; the timestamp for the arrival is the moment the unit reaches the scene, not the moment the first treatment is given.

21.3 Data requirements and sources

The indicator needs two timestamps for each qualifying intervention: the receipt of the emergency call and the arrival of the unit at the scene. These correspond to the first and the fourth of the steps that make up an intervention: receipt of the call, dispatching, travel to the scene, arrival and care at the scene, and transfer of the patient. The most convenient situation is a single dataset in which both timestamps are recorded, so that no linkage is needed. Where, instead, the two timestamps are held in different datasets and by different owners (a health ministry, individual hospitals, ambulance services, the police, individual dispatch centres, a dedicated agency or regional authorities) the datasets should be linked. However, this is not always possible when these databases share no common identifier.

Access to these data is often the binding constraint. Reasons are:

- privacy and health regulation can make access difficult for researchers
- the monitoring of interventions may be organised differently between regions
- part of the data may not be held electronically
- lead times can be long when approval is required, when the data owner has no resources to prepare the data, or when a third party is involved.

The data owners should be approached early and the legal basis for access should be set out clearly. Where a national database cannot be accessed at all, a fallback option is sampling, which often raises fewer access problems (see and Section 21.4.2 for the approach to be followed).

Three quality problems appear to emerge recurrently: missing of one the two time stamps; a systematic under-registration of interventions, which biases the result; and erroneous values or outliers, which require careful cleaning before the data can be used. The metadata should state the proportion of interventions for which a timestamp is unknown or erroneous. Where the value obtained has an incorrect scope or does not meet all the criteria of Section 21.2.2, a recalculated or estimated value is provided alongside the original, with the estimation method documented. Comparison across years requires that the definitions and the registration practices behind the figures have remained consistent.

21.4 Estimation of the SPI

The indicators are the 95th percentile and the median (50th percentile) of the distribution of response times. The 95th percentile is the value below which 95% of response times fall. How it is obtained depends on whether a full national database is available or the value must be estimated from a sample.

21.4.1 Calculating the percentiles from a complete database

Where a national database covers the whole territory, the indicators are computed directly from the full set of road-crash response times. The 95th percentile and the median are read from the cumulative frequency distribution of the times, taking the smallest response time whose cumulative share reaches or exceeds, respectively 50% and 95%. Because the values then rest on the whole population rather than on a sample, no confidence interval arises, in line with the common framework. Where it is useful for benchmarking, other percentiles (the 25th, 75th, 85th and 99th) can be obtained by exactly the same method.

21.4.2 Estimating from a sample

Where no national database can be used, the indicator can be estimated from a sample of emergency interventions. In this case the full sampling apparatus of Part B applies. The sample is drawn so as to be representative of the national population of road-crash response times, with a minimum total size of 2,000 interventions; a smaller sample is accepted for small countries with few injury crashes. The distribution of response times is markedly right-skewed because the longest times, often from remote locations, lie far apart. For this reason the precision of the 95th percentile cannot be judged in advance.

The sampling proceeds in two stages within a regional stratification, typically at NUTS 1 level, each region forming a stratum:

- Within each region, either all emergency medical service stations or a random sample of them are selected. Where sampled, at least ten stations per region where possible and at least 5 % of all stations should be selected. They should be chosen so as to neutralise the influence of the road type, the area type and the type of service each station mainly handles.
- Within each selected station, all interventions are taken where possible; where this is not feasible, a further random sample of at least 5 % of interventions is drawn.

The regional results are then weighted to the national figure, each stratum carrying a weight equal to its share of emergency interventions:

Equation 19. Weight factor post-crash care SPI based on a stratified sample.

$$\text{Weight of stratum } s = \frac{\text{number of emergency interventions in stratum } s}{\text{total number of emergency interventions}}$$

The 50th and 95th percentile are then computed on the weighted sample, and likewise for any weighted breakdown. Because the indicator is a percentile of a skewed distribution rather than a mean or a proportion, its confidence interval is not given by a closed formula and is obtained instead by bootstrapping, with the BCa (*bias-corrected accelerated*) bootstrap interval as the preferred variant. The bootstrap procedure is straightforward in principle: from the original sample of n response times, draw repeated samples of the same size with replacement (typically 1,000 to 10,000 such samples), compute the 95th percentile in each one, and use the resulting distribution of percentile estimates to construct the confidence interval. The percentile method takes the 2.5th and 97.5th percentiles of the bootstrap distribution as the lower and upper bounds; the BCa method adjusts these for bias and skewness in the bootstrap distribution and is generally more accurate, especially for high percentiles. More details are provided by Efron and Tibshirani (1993, p. 184).

A useful auxiliary technique is simulation. This requires drawing repeated samples from the observed sample at deliberately reduced sample sizes, in order to estimate how the confidence-interval width scales with sample size. Analyses within Trendline have shown that, for a representative response-time distribution, a sample of 2,000 produces a confidence interval that is already quite narrow.

The general weighting framework of Chapter 8 applies in full; the bootstrap for a percentile is its specialisation to this indicator. The metadata records the number of bootstrap samples used, or otherwise how the interval was determined.

21.4.3 Possible breakdowns

Where the database allows it, further breakdowns of the SPI values should be made for the following variables or combination of variables:

- by road type (motorway, urban, rural)
- by area population density
- by time of day
- by day of week
- whether e-call was used or not
- type of emergency service
- whether a physician was presence or not.

21.5 Minimum reporting requirements

The minimum deliverables are the 50th and the 95th percentile of the response time for the whole country, in minutes and seconds, for a stated year. The year referred to should be the most recent year available. Where possible, SPI values for earlier years should be added on a consistent basis, beginning with the Baseline value where one exists. Where the value has been estimated from a sample, a 95% confidence interval is reported as well, together with the unweighted number of interventions overall and per stratum and the statistical techniques used. A breakdown by road type should be provided where the data allow. Table 21-1 summarises the minimum and the optional elements.

Table 21-1. Minimum and optional reporting requirements for the post-crash-care indicator.

Aspect	Minimum requirement	Recommended / optional
Indicator value	50 th and 95 th percentile of response time (minutes and seconds), whole country, stated year	Earlier years on a consistent basis Other percentiles (25, 75, 85, 99)
Uncertainty	95% confidence interval where the value is estimated from a sample	—
Breakdowns	—	By road type (motorway, urban, rural) or by area population density by time of day, day of week, use of e-call, type of service presence of a physician
Metadata and background data	Method of calculation and, for a sample, the weighting and the bootstrap details Share of interventions with missing or erroneous timestamps	Cleaned per-intervention dataset (identifier, timestamps, weighting variables) Contextual information (Section 21.6.2)

21.6 Complementary indicators and additional information

21.6.1 Complementary and alternative presentations

The dataset also supports alternative presentations of the response time.

- As noted in Section 21.4.1, additional percentile values such as the 25th, 75th, 85th and 99th percentile values can also be used.
- The distribution can also be expressed as the share of interventions arriving within fixed time bands (e.g., within 10, 12, 15 and 20 minutes). This is often more intuitive for a policy audience than a single percentile value.

Two complementary indicators have been proposed at international level:

- the share of injury and fatal crashes for which no emergency care was provided
- the share of such crashes for which the response time did not exceed the national target.

It should be noted that the definition of a serious injury varies between countries; where available the MAIS₃₊ definition should be used.

21.6.2 Contextual information

A correct reading of the response time, and any benchmarking against other countries, is helped by contextual information about the emergency system and the data: the number of dispatching centres, of emergency transport units and of stations; the annual number of emergency calls in total and the number relating to road crashes, with the latter as a share of the former; and, as a measure of data quality, the proportion of interventions for which the call timestamp, the arrival timestamp, or both, are unknown.

22 SPIs for enforcement of traffic regulations

22.1 Introduction

Enforcement of traffic regulations is the whole penal procedure through which road users are persuaded to obey the traffic laws by the threat of detection and the imposition of a penalty. It is among the most effective ways of achieving a substantial improvement in road safety over a short period, particularly for risky behaviours such as speeding, the non-use of seat belts, the non-use of helmets, the use of a mobile phone while driving, and driving under the influence of alcohol or drugs. Systematic checks at well-chosen locations and times raise the level of enforcement that road users perceive, which changes their behaviour, reduces law-breaking, and consequently reduces the number and severity of crashes.

This indicator is the most heterogeneous of the five in Part C, mainly because enforcement methods and procedures, and the underlying traffic laws, vary widely between countries, and so therefore do the data that describe them. It is not possible to define a single enforcement indicator that all countries can produce on a comparable basis. Instead, Trendline offers four alternative options, of which a country calculates at least the one that best fits its data and its national road safety priorities.

In addition, the use of self-reported data could be considered. Surveys of the kind described in the chapter on self-reported data in Part C (the ESRA series being the established international example) ask road users directly about their exposure to enforcement, and so capture the indicator from the perspective of those being checked rather than from the records of those doing the checking. Two distinct quantities can be obtained. The first is the *perceived* likelihood of being checked: respondents are asked how likely it is, on a typical journey, that they will be checked by the police (including cameras and radars) for drink-driving, drug use, speeding, seatbelt use and handheld mobile phone use, on a seven-point scale that is dichotomised into a "likely" category. The second is the *experienced* level of enforcement: respondents are asked how many times in the past twelve months they have actually been checked for a given infringement (for alcohol and for drugs in ESRA), with the answer dichotomised into "at least once" against "never". The strengths of self-report based enforcement SPIs are that they are available where no enforcement database exists, that they are directly comparable internationally, and that they capture the deterrent effect on road users that the database options cannot reach.

A high level of enforcement does not in itself mean high compliance or high safety. Rather it indicates the effort a country puts into enforcement and the seriousness with which it addresses specific dangerous behaviours. The indicator measures enforcement activity; the four options measure different facets of that activity. Like the other database-based indicators it is built from existing administrative data rather than from a new survey; uniquely among them, it offers a choice of options so that a country can match the indicator to the data it holds, and for the camera-based options it gives non-official data sources a recognised, though secondary, role.

22.2 Indicator definition and scope

22.2.1 The four options

There is no single definition of the enforcement indicator. Instead, it is measured by one of four alternative indicators, **each expressed per head of population** (for example, per 100,000 inhabitants):

1. **The number of police checks per infringement**, for infringements such as speeding, seat-belt use, helmet use, distraction, drink-driving and red-light running.

2. The number of tickets per infringement, for the same set of infringements (a ticket is defined as "Any recorded offence for which a person receives a demand to pay a fine.").

3. The number of red-light cameras on the urban network.

4. The number of fixed speed-enforcement cameras, or of section-control stretches.

Option 1 captures the effort a country devotes to enforcement and, in combination with an analysis of its main road safety problems, helps to identify gaps in that effort. Option 2 reflects the effectiveness of enforcement in identifying violators, and read together with Option 1 gives an indication of enforcement as a preventive measure. In Options 1 and 2 police controls and tickets refer to enforcement activities taken up by Police officers in certain periods and locations. Radar sets operated by police officers or enforcement cameras that can be moved between different locations are taken into account; tickets arising from the cameras handled by the police should be counted in Option 2. Continuous enforcement using fixed cameras is not taken into account for this SPI. Options 3 and 4 measure the level of automated enforcement and indicate potential gaps. Option 4 includes all installed camera boxes, even if the camera in the installation is not temporarily working. Table 22-1 summarises the four options.

Table 22-1. The four enforcement SPI options at a glance.

Option	Indicator (per population)	What it captures	Typical source
1	Police checks per infringement	Enforcement effort	Traffic-police records
2	Tickets per infringement	Effectiveness in detecting violators	Traffic-police / judicial records
3	Red-light cameras on the urban network	Level of automated enforcement at high-risk sites	Official records; open camera databases
4	Fixed speed cameras or section-control stretches	Intensity of systematic speed enforcement	Official records; open camera databases

The per-population form is the one required for comparability; a country may in addition express any option in other units where these suit its data and interests (e.g., per kilometre of network, per unit of area, or per unit of traffic volume).

22.2.2 Selecting an option

The minimum requirement is to calculate at least one of the four options at national level. If Option 1 or Option 2 are chosen, as many infringements as possible should be covered, and preferably at least three. Which option a country chooses should follow from its most pressing road safety problems, the targets set in its national policy, and the data it can obtain. The options differ in what they demand of the data: Options 1 and 2 require a well-organised enforcement system in which every stage is recorded and accessible, whereas Options 3 and 4 can be assembled from a wider range of sources, including non-official ones, and are therefore often easier to produce. The infringements addressed correspond to the behaviours measured by the behavioural SPIs of Part B so that the enforcement indicator and those indicators can be read together.

22.3 Data requirements and sources

The enforcement indicator is calculated from data that already exist. The data are normally held by the traffic police of each country. In many countries they are published and freely available, whether by the police directly or through a ministry or the national statistical office. Where they are not freely available, they should be requested from the competent authority by means of a comprehensive questionnaire covering all aspects of the chosen option.

What is available differs by country, by option and by infringement.

- For Option 1 (police checks), data on police checks for speeding, seatbelt and helmet use and drink-driving exist in many countries, whereas checks for distraction are largely confined to mobile-phone use and data on red-light checks are limited.
- For Option 2 (tickets issued) the pattern is similar.
- For Options 3 and 4 the number of red-light and speed cameras is often widely available, from official sources and also from open databases that are maintained and regularly updated by anonymous users. Official sources should be the primary reference, and non-official sources should be used only to fill gaps with their trustworthiness checked and the resulting value flagged as approximate or indicative.

The level of detail in the source data determines how far the indicator can be broken down. Where a database holds only aggregate figures rather than a record per check or infringement, stratified results cannot be produced and the indicator is reported at the most detailed level the data allow. Because methods, procedures and laws differ between countries, cross-country comparison carries the caveats discussed in the common framework, and the metadata accordingly documents the regulations and procedures behind the value — in particular the legislation governing the behaviours addressed — alongside the source and coverage of the data.

22.4 Estimation of the SPIs

22.4.1 Computing the indicator

Each option is a normalised count: the relevant enforcement measure divided by the population, expressed per head of population (see Equation 20).

Equation 20. Computation of the Enforcement SPIs

$$SPI_{enforcement} = \frac{\text{the chosen enforcement measure}}{\text{population}}$$

The count in the numerator refers to

- the number of police checks for the infringement (Option 1), or
- the number of tickets for the infringement (Option 2), or
- the number of red-light cameras on the urban network (Option 3), or
- the number of fixed speed cameras or section-control stretches (Option 4).

For Options 1 and 2 the indicator is computed for each infringement covered, and, where the data are detailed enough, presented separately by road type, road user, vehicle type and time period.

The per-population denominator may be supplemented by the alternative denominators noted in Section 22.2.1, but the per-population form should always be reported so that values remain comparable across countries.

22.4.2 Estimating from a sample

Where no national enforcement database exists and the data are available only locally or regionally, the indicator can also be estimated from a sample of locations. The sample should be designed to be representative of the national network and to cover the whole geographical area of the country, using well-established statistical techniques. Areas are selected as randomly as possible in three steps:

- (1) the number of areas needed for the country or region is fixed;
- (2) areas (e.g., municipalities) are then chosen at random on the map, with attention to geographical spread,

(3) within each area all data are collected (e.g. for Option 3, all red-light cameras are counted).

The results are then weighted to the national figure by population (or by size of the network or by traffic volume, depending on the exact definition used. A design of this kind is most useful for Options 3 and 4.)

22.5 Minimum reporting requirements

The minimum deliverable is one of the four options at national level, with at least three infringements where Option 1 or Option 2 is chosen. The SPI values are given separately by road type, road user, vehicle type and time period wherever this is relevant to the option and the data allow. When using a sampling method, the number of areas and the number of road users behind the figure are reported before weighting.

The metadata should describe the methodology and the statistical techniques used, and documents the regulations and procedures for the behaviours addressed. The data may be delivered at three levels: as overall estimates for the chosen option and infringements; as a crossed matrix of all the disaggregations with their confidence intervals; or, ideally, as the cleaned underlying dataset. Table 22-2 summarises the minimum and the optional elements.

Table 22-2. Minimum and optional reporting requirements for the enforcement SPIs.

Aspect	Minimum requirement	Recommended / optional
Indicator	One of the four options at national level; for Options 1 and 2, at least three infringements	Additional options; more infringements; additional measurement units (per km of network, per area, per traffic volume)
Breakdowns	—	By road type (motorway, expressway, rural, urban); road user; vehicle type; time period (week-day/weekend, day/night, special days)
Coverage	Representative at national level	Regional (e.g. NUTS 1) breakdown
Data delivery	Overall estimates for the chosen option(s)	Crossed matrix of disaggregations with confidence intervals; cleaned dataset
Metadata	Methodology and statistical techniques; legislation and procedures for the behaviours addressed	—

22.6 Complementary indicators and alternative approaches

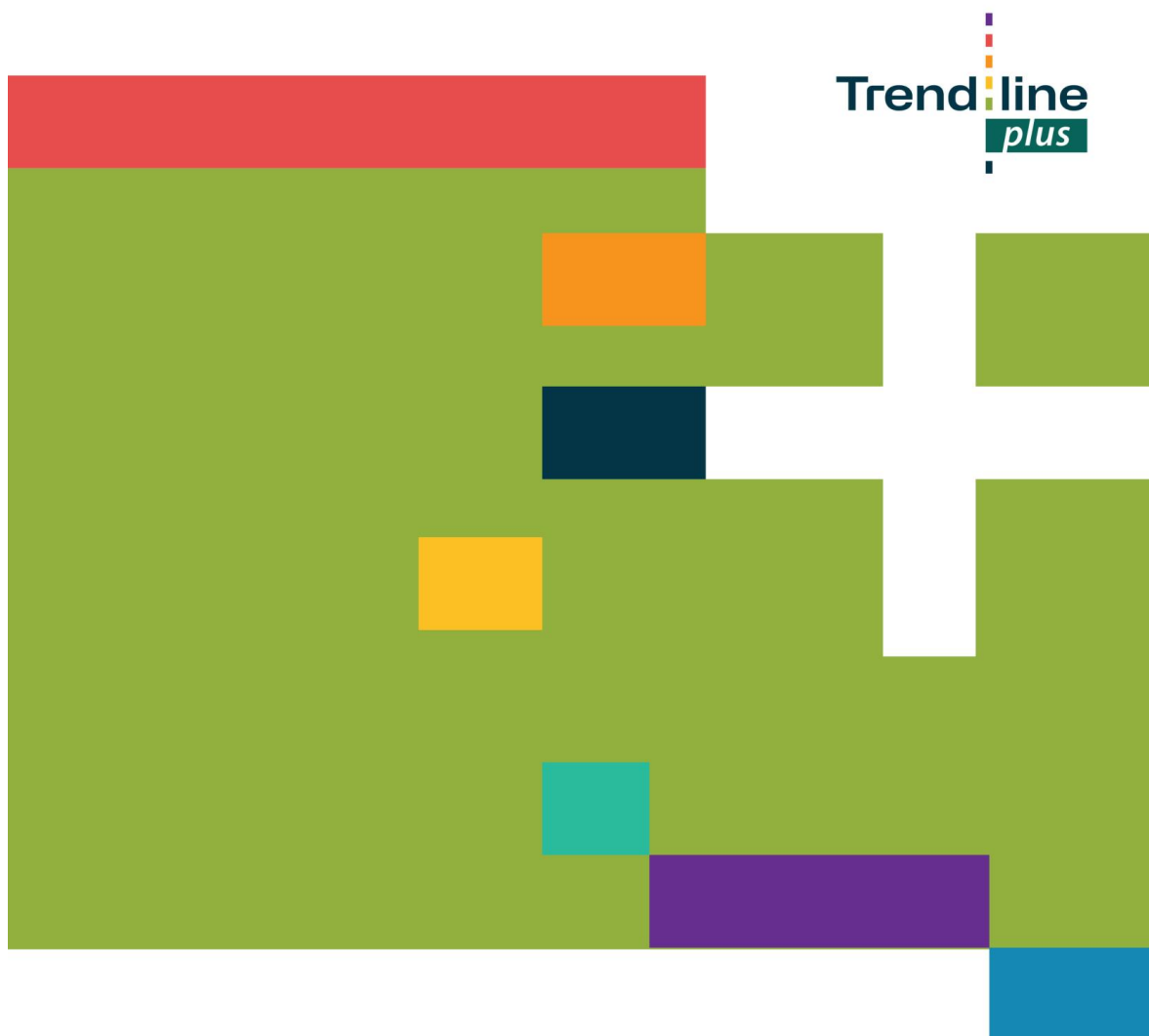
Beyond the single option that the minimum requires, a country may compute further options and cover additional infringements under Options 1 and 2. A country can additionally express any option in additional measurement units — per kilometre of network, per unit of area, or per unit of traffic volume — where these better reflect its situation.

Computing more than one option is itself informative: Options 1 and 2 read together, for instance, say more about the preventive value of enforcement than either does alone. These additions enrich the picture, but the per-population form of the chosen option is always reported so that the core value remains comparable across countries.

Using a sampling approach changes the methodology, but not the form and definition of the four options. The same breakdowns and combinations of breakdown are possible when using sampling methods, provided the samples are not becoming too small.

As mentioned in Section **Chyba! Nenalezen zdroj odkazů.**, self-reported indicators may also be considered, in particular when no national databases are available. One of the additional advantages of such SPIs is that

they allow for additional breakdowns such as by gender, age group, use of transport modes, support for enforcement and subjective safety.



PART D: IMPLEMENTATION GUIDANCE

23 Planning and managing an SPI study

23.1 The SPI study as a project

23.1.1 Introduction

The earlier parts of this manual set out what safety performance indicators are and how they are used (Part A) and how each indicator is measured (Parts B and C). This part is the practical counterpart to all of them. It addresses how to organise, resource, run and deliver an SPI study as a project, from the first scoping decisions to the submission of the results to the Trendline coordination team. It is deliberately generic: the same management discipline applies whether a country is running a roadside observation study, a survey, or a database-based indicator, and the chapter notes where the two families differ rather than repeating the methodology of Parts B and C.

Two realities of how studies are actually run shape this chapter. The first is that a single organisation rarely does everything: part of the work, and the fieldwork in particular, is often subcontracted, which is the subject of its own section below. The second is that the team may change across the life of a study: the people who design the sampling may not be the ones who analyse the data or may later communicate the results and use them to influence policy. This underlines the importance of good project coordination and documentation.

Please note that the operational, field- and office-level steps are set out as the checklists and other tools in the Appendices. This chapter and the methodology chapters explain the reasoning, while the appendices provide the step-by-step actions, the templates, the equipment specifications and the formulas. The numbering of the appendices used throughout this chapter is set out in the final section.

23.1.2 Roles and responsibilities

23.1.3 An SPI study involves several actors, and clear, durable responsibilities among them are one of the conditions for good data quality and for the sustainability of the system from one round to the next. Defining the scope

The first task in an SPI project is to define the scope. The commissioning authority or the study team decide which indicators to measure, and for indicators that offer definitional or methodological choices, which of them. Other elements of the scope to fixed are the reference period; the geographical coverage; the breakdowns required; and the level of data delivery aimed at (the reporting section below). Scope is set against the budget and, critically, against what data can be provided: for a behavioural indicator whether a sampling frame and the field capacity exist, and for a database indicator whether the necessary dataset exists and can be accessed. The choice of indicators should follow the country's road safety priorities rather than the ease of measurement, as Part A cautions. The selection and target-setting of indicators is supported by Checklist 1.

Table 23-1 (next page) sets out the principal roles. These roles are spread across the phases of the study rather than active throughout it, and they are not necessarily held by people of the same organisation: in a small in-house study the coordinator and the analyst may be one person, while in a larger study the fieldwork may sit with a subcontractor and the communication with a separate team. What matters is that every responsibility has an explicit owner and that the hand-over points between them are planned.

23.1.4 Defining the scope

The first task in an SPI project is to define the scope. The commissioning authority or the study team decide which indicators to measure, and for indicators that offer definitional or methodological choices, which of them. Other elements of the scope to fixed are the reference period; the geographical coverage; the

breakdowns required; and the level of data delivery aimed at (the reporting section below). Scope is set against the budget and, critically, against what data can be provided: for a behavioural indicator whether a sampling frame and the field capacity exist, and for a database indicator whether the necessary dataset exists and can be accessed. The choice of indicators should follow the country's road safety priorities rather than the ease of measurement, as Part A cautions. The selection and target-setting of indicators is supported by Checklist 1.

Table 23-1. Principal roles in an SPI study.

Role	Responsibility	Often held by
Commissioning authority	Defines the policy questions, sets targets, funds the study and acts on the results	Road safety agency or ministry
Study coordinator	Overall management: design, timeline, budget, quality and delivery	Research team
Data analyst	Weighting, calculation, confidence intervals or coverage, preparation of the dataset	Research team or subcontractor
Field coordinator / supervisor	Recruits, trains and supervises observers; manages field-work logistics	Research team or subcontractor
Field researchers / observers	Carry out the roadside observations or surveys	Subcontractor or hired field staff
Data owners	Supply and prepare the administrative or operational datasets (database studies)	External agencies
Police	Conduct the roadside stops for measurements of driving under the influence	Traffic police
Communication / policy team	Turns results into products for decision-makers and the public; runs campaigns	Authority's communications and policy staff
Trendline Secretariat	Receives the national deliveries and integrates the European report	Trendline consortium

23.2 Subcontracting and the division of work

23.2.1 How much is subcontracted

Most public authorities lack the in-house capacity to carry out SPI studies, particularly for roadside fieldwork at national scale, and they subcontract part of the work to a specialised research institute, a university, or a field-research agency. These organisations may in turn also subcontract part of that work to specialist organisations. Subcontracting replaces some of the internal effort with an external fee, and it adds a layer of specification, procurement and oversight that an in-house study does not carry. The commissioning authority or the lead institute remains responsible for the indicator and its use whatever it subcontracts.

The division of work between the commissioning team and a subcontractor varies along a spectrum, and the point chosen must be explicit from the start, because it determines where responsibility for verification, cleaning, weighting and quality lies:

- At one end, a **turnkey** arrangement has the subcontractor deliver the finished indicator values and carry out almost everything: design, fieldwork, data entry, cleaning, analysis and even the national report. The commissioning team's role is then to specify, oversee, verify and use the results.
- In a **partial** arrangement, the subcontractor carries out the fieldwork and delivers a cleaned dataset, and the research team performs the analysis and reporting.
- At the other end, a **raw-data-only** arrangement has the subcontractor simply collect and deliver the raw data, leaving the verification, cleaning, weighting, analysis and reporting to the research team.

When subcontracting, an important challenge is to avoid a mismatch of expectation on what exactly the subcontractor should deliver. The specification must therefore state not only what is to be collected but what state the deliverable is to be in and in what format it needs to be handed over.

23.2.2 Specifying and managing a subcontract

A subcontracted study adds project-management tasks that are worth planning explicitly:

- **A clear specification.** The tender specification should pin down:
 - the methodology: the sampling design, the definitions, the minimum sample sizes, the variables, the coverage and the required breakdowns
 - the deliverables and the state they must be in: raw records, cleaned dataset, analysis, or full report, and at which delivery level
 - the metadata to accompany them
 - the timeline: for the data collection and the delivery of the data
 - the quality assurance process: such as inter-rater reliability, documentation of deviations.

The checklists are a convenient way to make a specification precise: a tender can require the subcontractor to follow the sampling, site-selection, ethics and metadata checklists (Checklists 2, 3, 6 and 10) rather than restating their content. Checklist 12 covers the specification itself.

- **Procurement.** Writing the tender, advertising it, evaluating bids and awarding the contract take time; sometimes cumbersome procedures have to be followed. This time period should be built into the timeline. Bids should be judged on methodological quality and not on price alone.
- **Budget.** A subcontract is a distinct budget line, and the internal oversight it requires, such as specification, meetings, verification, should also be costed rather than assumed to be free.
- **Meetings and contract management.** A kick-off meeting, a checkpoint after the pilot, and regular progress meetings keep a subcontract on track. A pilot checkpoint can help to catch a methodological divergence before it contaminates the main data collection.
- **Acceptance and verification.** Before a deliverable is accepted, the research team verifies it against the specification (the coverage achieved, the sample sizes met, the metadata complete, the deviations documented).
- **Ethics and data protection.** The subcontractor is bound to the same ethical and legal requirements as an in-house team (Part B and Checklist 6), and where the police are involved for measurements of driving under the influence, the roles and the data-sharing arrangement are written into the contract.

23.2.3 Hand-over to communication and policy

The team that designs and analyses a study is often not the team that communicates its results and carries them into policy or campaigns; this work may fall to the commissioning authority's communications staff, its policy advisers or its campaign managers. This hand-over should be planned. The analysis team produces the audience-appropriate products described in Part A and briefs those who will use them. The metadata and the caveats should travel with the results across the hand-over, so that the people communicating the figures do not overstate their certainty or confuse an SPI with a crash statistic. Where a campaign is to be evaluated against an indicator, the evaluation design should be agreed before the campaign runs, not reconstructed afterwards.

23.3 Planning, timeline and resources

23.3.1 Phases and timeline

An SPI study moves through a recognisable sequence of phases, summarised in Table 23-2. The shape is common to both families, but the data-collection phase differs sharply: for a behavioural study it is roadside fieldwork, for a database study it is the acquisition, linkage and cleaning of existing data.

Table 23-2. Phases of an SPI study.

Phase	Behavioural study	Database study	Key output
Preparation and design	Sampling design, protocols, instruments	Identify sources, contact data owners, agree access	Study plan
Pilot	Test sites, instruments and observer coding	Test extraction, mapping and linkage on a subset	Validated protocol
Collection / acquisition	Roadside fieldwork	Data extraction, linkage and cleaning	Cleaned dataset
Analysis	Weighting, calculation, confidence intervals	Calculation, coverage assessment	SPI values
Reporting and delivery	National report and Trendline submission	National report and Trendline submission	Report and data delivery

Two scheduling points deserve particular attention:

- The lead time. For database studies the time a data owner needs to approve and prepare a dataset can be long. This should be anticipated from the outset. For behavioural studies the fieldwork window is constrained by season, daylight and weather. Where the work is subcontracted, the procurement period should be added to the front of the timeline.
- The value of a pilot. A small pilot almost always repays its cost by exposing problems in the protocol (or in a subcontractor's interpretation of it) before they can contaminate the main collection.

23.3.2 Resourcing and budget

The resources a study needs fall into a few categories, as shown in Table 23-3. The dominant cost differs between the two families and between in-house and subcontracted studies: for in-house behavioural studies it is field-researcher time, for database studies it is often the acquisition of the data, and for subcontracted studies it is the external fee plus the internal oversight.

Table 23-3. Main resource and budget categories.

Category	Includes
Personnel	Study coordinator, analyst, field coordinator, observers; supervision, training and oversight time
Subcontract	The external fee for fieldwork, analysis or a full study, where work is subcontracted
Equipment	Measurement devices, tablet or smartphone licences, personal protective equipment
Data acquisition	Purchase of registration, rating or enforcement data; the data owner's preparation costs
Software and other	Analysis software, secure storage and archiving, travel

When planning field capacity, it is good practice to provide for a reserve of roughly 10 to 15% more capacity than the schedule strictly requires. This allows illness, no-shows and scheduling clashes to be absorbed without rescheduling sessions.

23.3.3 Indicative effort and equipment costs

The figures in Table 23-4 and

Table 23-5 are indicative orders of magnitude of resources needed for SPI studies, to support early planning and budgeting. They assume a single national indicator of moderate scope. Effort and cost will scale with the number of indicators, the number of sites and the size of the country. Subcontracting can convert much of the fieldwork effort into an external fee plus oversight time. Table 23-4 gives indicative effort per activity. These should be multiplied with the typical cost of person-days in the country (which varies considerably across Europe) to arrive at a first estimate of the cost

Table 23-4. Indicative effort per activity (person-days, single national indicator).

Activity	Indicative person-days	Applies to
Study design and sampling or source plan	5–15	All
Project coordination (whole study)	10–25	All
Engaging data owners and obtaining access	3–10	Database
Observer recruitment and training	3–8	Behavioural
Pilot	3–5	All
Fieldwork (observation sessions)	30–100	Behavioural
Data extraction, mapping, linkage and cleaning	5–20	Database
Data entry and cleaning	5–15	Behavioural
Analysis (weighting, calculation, intervals or coverage)	5–12	All
National report	8–20	All
Trendline data submission (template and metadata)	2–5	All
Dissemination and policy briefing	5–12	All

Table 23-5 gives indicative equipment costs. Prices vary widely by country, supplier and year; renting can be cheaper than buying for a single round; and several items — LIDAR and evidential breath analysers in particular may be provided in kind by the police rather than purchased by the study. Detailed equipment specifications are given in Appendix E.

Table 23-5. Indicative equipment costs (EUR).

Item	Indicative cost	Note
High-visibility vests and basic PPE	10–30 each	Purchase
Tablet/smartphone for data-collection application	200–500 each	Application often provided centrally
Handheld tally counter	5–20 each	Purchase
Handheld speed radar	1,000–3,000	Purchase or police-owned
LIDAR speed device	3,000–10,000+	Often police-owned (in kind)
Roadside speed or observation camera setup	1,000–5,000	Purchase or rent
Floating-car / GPS speed data	Subscription, variable	Licence per area and year
Breath-alcohol screening device	200–800	Often police-owned
Evidential breath analyser	2,000–6,000	Police-owned (in kind)
Oral-fluid drug screening test	10–30 per test	Consumable
Distraction observation camera system	500–3,000+	Higher for intelligent systems

23.3.4 Staffing and training

The recruitment, training and quality control of field observers is a methodology in its own right and is treated in Part B and in Checklist 4; only the management view is given here. The training of field researchers can often be delivered as a short online session held shortly before the fieldwork window (in particular when there is a pilot roadside session planned afterwards). The material should have been distributed in advance so that the session is used for clarification rather than first exposure, and supplemented by a question-and-answer reference document for the field phase. A starter version of such a reference is provided as the field-observer FAQ in Appendix D, to be adapted and extended by each country as new borderline cases arise. Where the fieldwork is subcontracted, the responsibility for recruitment and training passes to the subcontractor, but the commissioning team still verifies that the training took place and that reliability was checked.

23.4 Quality assurance

23.4.1 Quality control during the study

Quality is built in during the study rather than checked at the end. For behavioural studies this means inter-rater reliability checks before fieldwork starts and at intervals during it, supervision of sessions, and simultaneous traffic counting where the design requires it. For database studies this implies validating the extraction and linkage on a pilot subset, checking coverage, and verifying that the mapping to the indicator definition is correct.

In both cases any deviation from the planned protocol should be recorded, so that the data can be interpreted correctly and that the protocol can be improved for the next round. The relevant checklists give the operational detail: Checklists 2 to 7 for behavioural studies and Checklists 8 and 9 for database studies.

23.4.2 Preparing and cleaning the data

Whatever its source, the data finally used are cleaned data, not the raw record exactly as collected or extracted. Cleaning removes test and out-of-window records, corrects or removes outliers and improperly formatted records, resolves duplicates, and verifies the session or record metadata; for database studies it also covers the mapping of source variables to the indicator definition and the resolution of records that could not be linked.

Where a subcontractor has delivered data, the cleaning responsibility depends on the arrangement agreed in the specification. In a raw-data-only arrangement the research team carries out the whole of the verification and cleaning, and treats the delivered records exactly as it would its own raw data; in a turnkey arrangement the team verifies that the cleaning was done and documented to specification rather than repeating it.

The cleaning rules applied, and the proportion of records they remove or leave unresolved, are documented in either case, because they bear directly on coverage and on the interpretation of the value. The data-processing steps are set out in Checklist 7, and a worked example of the analysis is provided in Appendix B.

23.5 Reporting and delivering the data to Trendline

This section sets out the operational requirements for delivering results to the Trendline coordination team. The principles of reporting and communication, and the use of the results in policy, are covered in Part A; what follows is the mechanics, and they are gathered in tick-box form in Checklist 10.

23.5.1 Defining the levels of data delivery

Results are delivered at one of three levels of increasing richness, described in Table 23-6. Defining the levels once lets each indicator chapter simply state which level it requires as a minimum. The levels were developed for behavioural indicators, where they are most detailed, but they carry over to database indicators with the confidence interval replaced by the coverage level: at the minimum level the required indicator values each with their coverage; at the medium level the disaggregations (by road type, region or the relevant strata), each with their coverage and supporting counts; and at the ideal level the cleaned record-level dataset. This means the section-level file for infrastructure, the per-vehicle or per-model file for vehicle safety, the per-intervention file for post-crash care, and the per-record file for enforcement.

Table 23-6. Levels of data delivery.

Level	What is delivered	Typical use
Minimum	Point estimates with their measure of confidence (a 95% confidence interval for sampled indicators, a coverage level for database indicators), for each required indicator and each required disaggregation taken on its own	Executive summaries, ministerial briefings, one-page products
Medium	A crossed matrix of estimates for combinations of strata, each cell with its own interval or coverage; cells below the sample-size threshold flagged rather than reported as precise	Technical reports; identifying the best- and worst-performing combinations
Ideal	The full cleaned, unweighted record-level dataset	Independent re-analysis, secondary research and re-aggregation

For sampled indicators the sample-size convention at the medium level is to report a cell fully at 500 or more observations, to flag it between 250 and 499, and to treat fewer than 250 as a deviation. Releasing data at the

ideal level requires a deliberate decision on anonymisation and licensing, taken at the start of the study rather than added afterwards.

23.5.2 Using the data-submission templates

National results are delivered using the standardised Trendline data-submission templates rather than in free form. This facilitates their integration into the European database and the European report without rekeying or reinterpretation. The procedure for obtaining the current templates (Excel sheets with multiple tabs) is available at the following website url: <https://trendlineproject.eu/templates>. Some guidance on completing the templates is provided in Appendix A. The study team should examine these data templates at the planning stage, because knowing the structure the results must take helps to shape the analysis. Also subcontractors should be required to deliver data by using these templates.

23.5.3 Producing the metadata record

Every delivery, at every level, should be accompanied by a metadata record. Metadata is contextual information; without it two similar-looking values may rest on different definitions and mislead. At a minimum the metadata record should include as a minimum:

- the operational definition of the indicator and the population covered
- the geographical coverage and the period the data refer to
- the data source or, for sampled indicators, the sampling design, including the stratification and the number of locations per stratum
- for database indicators, the coverage level and how it was determined
- the data-collection or extraction method
(for sampled indicators) the weighting basis and the confidence-interval method
- (for database indicators) the classification and mapping rules
- the data-cleaning rules and the proportion of records excluded
- any deviation from the minimum requirements, with its reason.

Each indicator additionally requires its own specific metadata. Examples are speed limits for the speeding indicators, blood-alcohol limits for driving under the influence, and age thresholds for cycle-helmet use. These are listed in the indicator chapters and gathered, in tick-box form, in the metadata and delivery checklist (Checklist 10).

23.5.4 Respecting the minimum requirements for data delivery

Drawing the above together, the minimum a country delivers to the coordinator or technical secretariat is:

- the SPI values with their measure of confidence
- the required disaggregated results with their intervals or coverage and their bases
- the common metadata core together with the type- and indicator-specific additions
- a methodological accompaniment, in particular wherever there is any deviation from the standard methodology
- the cleaned record-level dataset where the indicator chapter requires the ideal level or where national policy is to release it.

23.6 The appendices and how to use them

The appendices turn the principles of this manual into directly usable tools. While the body of this manual explains the reasoning and the logic, the appendices provide the actions, the templates, the equipment detail and the formulas. They are referenced by the numbering below throughout this chapter:

- **Appendix A — Data templates and guidance.** References to the standardised Trendline data templates and a guide to using them.
- **Appendix B — Worked example of data analysis.** A cleaned dataset taken through weighting to its final values with their confidence interval.
- **Appendix C — Checklists.** A toolbox of checklists covering both behavioural and database studies:
 - Checklist 1, selecting and target-setting SPIs
 - Checklist 2, sampling and selection of observation locations
 - Checklist 3, observation-site requirements
 - Checklist 4, observer selection, training and quality control
 - Checklist 5, questionnaire and survey design
 - Checklist 6, ethics, privacy and legal requirements
 - Checklist 7, data processing and analysis
 - Checklist 8, data sources and data access for database SPIs
 - Checklist 9, data linkage and coverage for database SPIs
 - Checklist 10, metadata and data delivery
 - Checklist 11, reporting and communicating results
 - Checklist 12, subcontracting and tender specification.
- **Appendix D — Frequently asked questions for field observers.** The borderline cases that a methodology cannot fully anticipate, to be extended by each country as its experience accumulates.

Appendix E — Equipment specifications. The devices used in behavioural measurement, with the specifications behind the indicative costs of

- Table 23-5.

The checklists can be adapted to national circumstances. Countries are encouraged to extend them as their own experience accumulates, so that each round of measurement leaves the tools better than it found them.

REFERENCES

Methodological guidelines developed in Baseline and the first phase of Trendline

- Ambros, J., Šípek, M., Ljubotina, L., Krizsik, N., Deane, P., Kreicbergs, J., Vieira, S., Karantanos, M., Mondésir, H., & Folla, K. (2025). *KPI Compliance with traffic rules on signalized pedestrian crossings and intersections. Methodological Guidelines.*
- Ambros, J., Šípek, M., Ševrovič, M., Krizsik, N., Deane, P., Vieira, S., Karantanos, M., Mondésir, H., & Folla, K. (2025). *KPI Compliance with traffic rules on unsignalized pedestrian crossings and intersections. Methodological Guidelines.*
- Boets, S., Folla, K., Houwing, S., Forsman, Å., Klipp, S., Areal, A., Jankowska-Karpa, D. & Meesmann, U. (2023). *KPI Driving under the Influence of Alcohol. Methodological Guidelines.*
- Dragomanovits, A. & Van den Berghe, W. (2023). *KPI Infrastructure. Methodological Guidelines.*
- Kšicová, E., Forsman, Å., Areal, A., Laiou, A., Valentová, V., Temmerman, P., & Boets, S. (2023). *KPI safety belts and child restraint systems. Methodological Guidelines.*
- Kšicová, E., Moreau, N., Zielińska, A., Vadeby, A., Areal, A., Ferrer López, S., & João Da Silva Barros, M. (2023). *KPI Helmet use among cyclists and powered two-wheelers (PTWs). Methodological Guidelines.*
- Kšicová, E., Musil, L., Krizsik, N., Heffernan, S., Capkin, S. O. K., Zielińska, A., João Da Silva Barros, M., Mondésir, H., Salathé, M., & Ferrer López, S. (2024). *KPI Helmet use among PMD riders. Methodological Guidelines.*
- Laiou, A., Mondésir, H., Mesimäki, J., Deane, P., Kome, S., Cardoso, J., & Vadeby, A. (2025). *Alternative Speeding Indicators. Methodological Guidelines.*
- Laiou, A., Silverans, P., Mondésir, H., Malin, F., Jankowska-Karpa, D., Torbay, A., & Goldenbeld, C. (2025). *Enforcement of traffic regulations. Methodological Guidelines.*
- Laiou, A., Vadeby, A., Riguelle, F., Van Schagen, I., Cardoso, J., Tutka, P., Usami, D. S., & Wardenier, N. (2023). *KPI Speed. Methodological Guidelines.*
- Moreau, N., Doumen, M., Marunica, A., Mataija, M., & Mondésir, H. (2023). *KPI Light use by cyclists. Methodological Guidelines.*
- Silverans, P. & Meesmann, U. (2025). *Experimental KPI Self-reported behaviour and attitudes. Methodological guidelines.*
- Stelling, A., Boets, S., González Hernández, B., Jankowska, D., Larsson, P., Schumacher, M., Vieira, S., & Ziakopoulos, A., (2023). *KPI Distraction. Methodological Guidelines.*
- van der Kint, S., Stelling, A., Van den Berghe, W., Mondésir, H., Corado, A., Areal, A., Torbay, A., Forsman, Å., Klipp, S., & Folla, K. (2025). *KPI Driving under the influence of drugs. Methodological guidelines.*
- Wardenier, N., Van den Berghe, W., Schram, R., Gaillet, J.-F., Thomas, P., Folla, K., Cleij, D., Rizzi, M., & Silverans, P. (2024). *KPI Vehicle Safety. Methodological Guidelines.*
- Weijermars, W., Van den Berghe, W., Bijleveld, F., Nuytens, N., Forsman, Å., Folla, K., & Segui Gomez, M. (2024). *KPI Post-crash Care. Methodological Guidelines.*

Weijermars, W., Van den Berghe, W., Uijtdewilligen, T., Van Petegem, J. H., Odijk, M., Dragomanovits, A., Mondésir, H., Salathé, M., Georgieva, M., Mesimäki, J., Rodrigues, E., Torbay, A., Vidigal, A., Amaro, A., Rosa, F., Sternlund, S., Ekman, L., & Peterson, M. (2025). *KPI 30 km/h on urban roads. Methodological Guidelines*.

Scientific publications referred to in the Manual

Ambros, J., Vieira, S., Nagy, P. D. (2026). *Compliance on Pedestrian Crossings and Intersections: Lessons from the Trendline Project*. Transport Research Arena 2026, Budapest.

Bouwen, L., Dons, E., & Schoeters, A. (2022). *L'impact du vélo sur la santé, le climat et l'économie en Belgique - Revue de la littérature et analyse coût-avantage d'une augmentation de l'utilisation du vélo en Belgique*. Vias institute, Brussels.

https://www.vias.be/publications/Impact%20van%20fietsen%20op%20gezondheid,%20klimaat%20en%20economie%20in%20Belgi%C3%AB/impact_du_v%C3%A9lo_sur_la_sant%C3%A9_le_climat_et_l'economie.pdf

Diependaele, K. (2015). *Sleepy at the wheel: Analysis of the extent and characteristics of sleepiness among Belgian car drivers*. Report 2015-R-06-EN. Vias institute, Brussels. <https://www.vias.be/publications/Slaperig%20achter%20het%20stuur/Sleepy%20at%20the%20wheel%20-%20Analysis%20of%20the%20extent%20and%20characteristics%20of%20sleepiness%20among%20Belgian%20car%20drivers.pdf>

Efron, B., & Tibshirani, R. (1993). *An Introduction to the Bootstrap*. Chapman & Hall/CRC, Boca Raton.

Elgner, J., Andrášik, R., Bijleveld, F., Boets, S., & Van den Berghe, W. (2024). *Suggested approach for weighting sample data and calculation of statistics*. <https://trendlineproject.eu/media/pages/publications/32e9d29d7c-1719921954/statistical-analysis-of-kpi-data.pdf>

Elvik, R. (2013). Risk of road accident associated with the use of drugs: A systematic review and meta-analysis of evidence from epidemiological studies. *Accident Analysis & Prevention*, 60, 254-267.

Elvik, R., Vadeby, A., Hels, T., & van Schagen, I. (2019). Updated estimates of the relationship between speed and road safety at the aggregate and individual levels. *Accident Analysis & Prevention*, 123, 114-122.

ETSC (European Transport Safety Council, 2001). *Transport safety performance indicators*. <https://archive.etsc.eu/documents/perfindic.pdf>

European Commission (2020). *Next steps towards 'Vision Zero': EU road safety policy framework 2021–2030*. Luxembourg: Publications Office of the European Union. <https://data.europa.eu/doi/10.2832/391271>
<https://data.europa.eu/doi/10.2832/391271>

European Commission (2022). *Road safety thematic report – Safety Performance Indicators*. European Road Safety Observatory. Brussels: European Commission, Directorate General for Transport. https://road-safety.transport.ec.europa.eu/document/download/2b27e9e4-74fa-4dde-a732-b077a284da4c_en?file-name=Road_Safety_Thematic_Report-Road_Safety_Performance_Indicators.pdf

European Parliament (2022). *EU Road Safety Policy Framework 2021-2030 — Recommendations on next steps towards 'Vision Zero'*. <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52021P0407>

European Union (2023). *Network Wide Road Safety Assessment: Methodology and Implementation Handbook*. <https://road-safety.transport.ec.europa.eu/system/files/2023-11/NWA-Handbook8.pdf>

- Fondzenyuy, S. K., Mobogshikeh, D. M., Tezong, S. L. F., Shingo Usami, D., & Persia, L. (2024). *Evaluating Speeding Safety Performance Indicators in an Urban Area of a LMIC: A Case Study of Yaoundé, Cameroon*. Preprints.Org. <https://doi.org/10.20944/preprints202402.1703.v1>
- Goldenbeld, C., & De Craen, S. (2013). *The comparison of road safety survey answers between web-panel and face-to-face; Dutch results of SARTRE-4 survey*. *Journal of Safety Research*, 46, 13–20.
- Hakkert, A. S., & Gitelman, V. (2007). *Road Safety Performance Indicators: Manual*. Deliverable D3.8 of the EU FP6 project SafetyNet. Vehicle Safety Research Centre, Loughborough.
- Hels, T., Bernhoft I. M., Lyckegaard, A., Houwing, S., Hagenzieker, M., Legrand, S.-A., Isalberti, C., Van der Linden, T., & Verstraete, A. (2011). *Risk of injury by driving with alcohol and other drugs*. DRUID project deliverable 2.3.5.
- Holló, P., Henézi, D., & Berta, T. (2018). Comparison of Self-reported and Observed Road Safety Performance Indicators. *Periodica Polytechnica Transportation Engineering*, 46, 117–121.
- Korn, E. L., & Graubard, B. I. (1998). Confidence Intervals For Proportions With Small Expected Number of Positive Counts Estimated From Survey Data. *Survey Methodology*, 24, 193–201.
- Kroesen, Maarten, Susan L. Handy, Caspar Chorus (2017) *Do Attitudes Cause Behavior or Vice Versa? An Alternative Conceptualization of the Attitude-Behavior Relationship in Travel Behavior Modeling*. Transportation Research Part A 101. <https://doi.org/10.1016/j.tra.2017.05.013>
- Kuiken, M., & Stoop, J. (2012). *Verbetering van fietsverlichting. Verkenning van beleidsmogelijkheden*. https://www.fietsberaad.nl/CROWFietsberaad/media/Kennis/Bestanden/Kuiken%202012%20Verbetering%20Fietsverlichting_def.pdf?ext=.pdf
- Lajunen, T., & Özkan, T. (2011). Self-Report Instruments and Methods. In Porter, B. E. (Ed.). *Handbook of Traffic Psychology*, 43–59.
- Meesmann, U., & Wardenier, N. (2024). *ESRA3 methodology*. Report 2024–R–09–EN. Vias institute, Brussels. <https://www.esranet.eu/storage/minisites/esra3-methodology-report.pdf>
- OECD/ITF (2018). *Speed and Crash Risk*. <https://www.itf-oecd.org/sites/default/files/docs/speed-crash-risk.pdf>
- Owens, A. D., & Sivak, M. (1996). Differentiation of Visibility and Alcohol as Contributors to Twilight Road Fatalities. *Human Factors*, 38, 680–689.
- Parker, J. D., Talih, M., Malec, D. J., Beresovsky, V., & Carroll, M., Gonzalez Jr., J. F., Hamilton, B. E., Ingram, D. D., Kochanek, K., McCarty, F., Moriarity, C., Shimizu, I., Strashny, A., Ward, B. W. (2017). National Center for Health Statistics Data Presentation Standards for Proportions. *Vital and Health Statistics*, 2, 175.
- Rao, R. K., & Bharadwaj, N. (2026). A systematic review of traffic safety data collection methods and challenges: From crash databases to AI-augmented sensors. *Journal of Safety Research*, 97, 612–627.
- Rogeborg, O. (2019). A meta-analysis of the crash risk of cannabis-positive drivers in culpability studies — Avoiding interpretational bias. *Accident Analysis & Prevention*, 123, 69–78.
- Rogeborg, O., & Elvik, R. (2016). The effects of cannabis intoxication on motor vehicle collision revisited and revised. *Addiction*, 111, 1348–1359.
- Schmidt, F. A., Mons, C., & Doumen, M. J. A. (2025). *Pilot camera use for monitoring KPI light use by cyclists in the dark*. SWOV, internal report, unpublished.

- Schroeder, P., & Wilbur, M. (2013). *2012 National Survey of Bicyclist and Pedestrian Attitudes and Behavior. Volume 2: Findings Report*. <https://www.nhtsa.gov/sites/nhtsa.gov/files/811841b.pdf>
- Schulze, H., Schumacher, M., Urmeew, R. & Auerbach, K. (2012). *DRUID Final Report: Work performed, main results and recommendations*. DRUID project deliverable. <https://www.vias.be/publications/DRUID%20-%20Final%20Report%20-%20Work%20performed,%20main%20results%20and%20recommendations/DRUID-Final%20Report.pdf>
- Silverans, P., & Vanhove, S. (2023). *Baseline conclusions and recommendations*. https://road-safety.transport.ec.europa.eu/document/download/b4b82a6b-9e1f-4934-97bf-c0754ef19278_en?file-name=Baseline_conclusions_and_recommendations.pdf
- SWOV (2020). *Drugs and medicines*. SWOV fact sheet, https://swov.nl/sites/default/files/bestanden/downloads/FS%20Drugs%20and%20medicines_o.pdf
- Thomas, P., Chambers-Smith, D., Morris, A., Yannis, G., Evgenikos, P., Duchamp, G., Vis, M., Jahi, H., Vallet, G., Dupont, E., & Martensen, H. (2009). *SafetyNet Final Activity Report*. Vehicle Safety Research Centre, Loughborough.
- Timmermans, E., Prey, A., & Laurens, J. (2022). *Lichtvoering fietsers 2021/2022*. <https://open.rijkswaterstaat.nl/@165594/lichtvoering-fietsers-2021-2022/>
- Transport Community (TC, 2024). *Final Report – Technical Assistance for collecting Road Safety KPIs on Seat-belt and Child Restraint Systems in the Western Balkan region*. https://www.transport-community.org/wp-content/uploads/2024/04/Final-Report-Westbelt_ONLINE.pdf
- Twisk, D. A. M., & Reurings, M. (2013). An epidemiological study of the risk of cycling in the dark: The role of visual perception, conspicuity and alcohol use. *Accident Analysis & Prevention*, 60, 134-140.
- Van den Berghe, W., & Stelling, A. (2025). *Trendline 2022-2025 – Data collection and analysis of road safety KPIs in Europe*. https://road-safety.transport.ec.europa.eu/document/download/da1b9324-9bfe-4405-8032-e9c1021f4964_en?filename=Trendline_final_report.pdf
- Van den Berghe, W., Fleiter, J. J., & Cliff, D. (2020). *Towards the 12 voluntary global targets for road safety. Guidance for countries on activities and measures to achieve the voluntary global road safety performance targets*. https://www.vias.be/publications/Towards%20the%2012%20Voluntary%20Global%20Targets%20for%20Road%20Safety/Towards_the_12_Voluntary_Global_Targets_for_Road_Safety.pdf
- Van den Berghe, W., S. Hakkert, V. Gitelman (2024). *Safety Performance Indicators*. In Laureshyn, A., Agerholm, N. & Sucha, M. (eds.), *Traffic safety data - the way to better knowledge and better traffic safety*. London, UK: The Institution of Engineering and Technology.
- Vollrath, M., Schumacher, M., Boets, S., & Meesmann, U. (2019). *Guidelines for assessing the prevalence of mobile phone use in traffic*. FERSI technical paper. <https://fersi.org/wp-content/uploads/2019/11/Guidelines-prevalence-mobile-phone-use.pdf>
- Wood, J. M. (2023). *Improving the conspicuity and safety of pedestrians and cyclists on night-time roads*. *Clinical and Experimental Optometry* (Vol. 106, Issue 3, pp. 227–237). Taylor and Francis Ltd. <https://doi.org/10.1080/08164622.2023.2174001>

APPENDICES

Appendix A: (reference to) data templates

Appendix B: worked out example

Appendix C: Checklists

Appendix D: FAQ for field observers

Appendix E: Equipment specifications.

Appendix A — Data templates and guidance

National results are delivered to the Trendline coordination team using SPI-specific data -submission templates. The structure of each template depends on the type of SPI and on the way in which the data are collected.

In general, two broad types of templates can be distinguished:

- **roadside-based templates**, mainly used for behavioural SPIs collected through roadside observations or surveys;
- **desk-based templates**, used for SPIs calculated from administrative databases or other existing national data sources.

The purpose of the templates is to standardise the delivery of data and results across Member States, while allowing for differences in data-collection methods between SPIs.

SPI-specific template structure

The template is an Excel workbook with a small number of sheets. The templates are not identical across SPIs. Each SPI has its own template structure reflecting:

- the data source
- the required level of detail
- the stratification variables
- the weighting procedure
- the required output tables
- the metadata needed for interpretation.

For example, the template for the distraction SPI is designed for roadside measurement data. It contains sheets for individual-level observations, traffic-count calibration data, stratum-level weighting information and aggregated results. The template for vehicle safety, by contrast, is a database template and only requires aggregated information on newly registered passenger cars and their safety rating categories.

Structure of roadside-based templates

Roadside-based templates may contain several sheets corresponding to different levels of data processing. The exact structure depends on the SPI, but roadside behavioural templates typically distinguish between source data, count period data, weighting information, results and metadata.

Where traffic count period data are available, it is used to estimate the number of relevant vehicles passing during observation session ($ETN_j(k)$), taking into account the duration of the observation session and the counting period. Roadside-based templates may also include sheets with information at the level of the basic stratum. Such information is used to calculate the stratum-level weight characteristic, for example Sl_{ij} , which represents the relative importance of a given stratum and vehicle type in the final aggregated SPI. Where direct traffic-volume or vehicle-kilometre information is not available, fallback information such as road length or the relative length of considered time periods may be used.

The results sheets are used to report aggregated and semi-aggregated SPI values. Depending on the template and the available data, these sheets may be filled directly by the Member State or completed by the Trendline coordination team from the submitted source data.

Structure of desk-based templates

Desk-based templates are generally simpler than roadside-based templates because they do not require individual observation records or session-level count period data. Instead, they are designed for aggregated data extracted from national databases or administrative sources.

For example, the vehicle safety template contains a sheet with aggregated results and a metadata sheet. The main data items include the total number of new passenger cars, the number of new passenger cars without rating, the number of 0–3-star passenger cars, the number of 4-star passenger cars, the number of 5-star passenger cars, and KPI percentage thresholds for 4-star and 5-star ratings. The metadata are therefore particularly important for understanding the coverage, definitions and limitations of the national data source.

Completing the template

Member States may submit either calculated results or the source data needed for the Trendline coordination team to calculate the results. For roadside-based SPIs, this means that Member State may either fill the aggregated and semi-aggregated result sheets directly or provide the underlying raw data, count period data and other characteristics needed for weighting procedure. For desk-based SPIs, the submission usually consists of aggregated data and metadata.

When completing a template, Member States should follow these general principles:

- use the correct SPI-specific template;
- complete all required metadata fields;
- keep predefined sheet names, column names and category names unchanged;
- document any deviations from the recommended methodology;
- respect reporting thresholds and sample-size requirements;
- avoid mixing data from different collection methods.

Metadata should be completed carefully, as they are necessary for interpreting the submitted results. They should describe the data source, reference period, coverage, deviations from the recommended methodology, missing data, assumptions used in weighting or aggregation, and any limitations affecting comparability.

Common errors to avoid

Typical problems in submissions include:

- missing measures of confidence intervals
- an incomplete metadata sheet
- national categorisations substituted for the standard ones without explanation
- incorrect calculation of SPI values, confidence intervals, weights or other reported characteristics
- methods mixed in one series
- the structure of the template overwritten or rearranged, which breaks the automated integration.

Values should be calculated according to the relevant SPI-specific methodology, including the required weighting, stratification, sample-size thresholds and confidence-interval calculation. Where applicable, the template also asks for a brief description of the weighting formula and CI calculation. Checklist 10 turns these requirements into a final tick-box check before delivery.

See also: Part D for reporting and delivering the data and Checklist 10 for metadata and delivery.

Appendix B – Worked out example

This appendix provides an example of the calculation of the distraction SPI. The example illustrates how individual observation data, combined observation weights and the survey design specification can be used to calculate weighted SPI estimates and confidence intervals. The example is based on Czech data for the distraction SPI. The response variable used in this example is NotUsed, where:

- NotUsed = TRUE means that the driver was not observed using a distracting device or behaviour;
- NotUsed = FALSE means that the driver was observed using a distracting device or behaviour.

Thus, the SPI in this example represents the weighted proportion of drivers who were not observed as distracted with a mobile device.

Data structure

The Czech distraction dataset used in this example contains 19,672 individual vehicle observations collected during 72 roadside observation sessions. Each row represents one observed vehicle. Observations are classified by region, road type, time period, session, location, vehicle type, driver group, gender and age group and contain a distraction-related response variable and a final combined observation weight W_x . The final combined observation weight W_x incorporates the within-session factor and the stratum-level aggregation factor, as described in Section 8.2. It is assigned to each individual observation. Table B.1 shows an extract of the individual-level dataset. For clarity, only variables relevant for the calculation are shown.

Table B.1 Illustrative example of the Czech distraction dataset

Region ID	Road type	Time period	Session ID	Location ID	Vehicle type	Gender	NotUsed	W_x
r000001	rural road	weekday/daytime	23411	1155	passenger car	male	TRUE	0.0000837
r000001	rural road	weekday/daytime	23411	1155	heavy goods vehicle	male	TRUE	0.0000917
r000001	rural road	weekday/daytime	23411	1155	passenger car	male	TRUE	0.0000837
r000001	rural road	weekday/daytime	23411	1155	passenger car	male	TRUE	0.0000837
r000001	rural road	weekday/daytime	23411	1155	passenger car	male	TRUE	0.0000837
r000001	rural road	weekday/daytime	23411	1155	passenger car	female	TRUE	0.0000837
r000001	rural road	weekday/daytime	23411	1155	passenger car	male	TRUE	0.0000837
r000001	rural road	weekday/daytime	23411	1155	passenger car	male	TRUE	0.0000837
r000001	rural road	weekday/daytime	23411	1155	light goods vehicle	male	FALSE	0.0000734
r000001	rural road	weekday/daytime	23411	1155	passenger car	male	TRUE	0.0000837

Please note that each row represents one observed vehicle. The variable Not distracted corresponds to NotUsed = TRUE, i.e. the driver was not observed using a mobile device.

Survey design specification

The variance estimation accounts for the stratified and clustered structure of the roadside observation data. Observation sessions are treated as clusters, because observations collected during the same session may not be statistically independent. The design is stratified by road type and time period.

In R, the survey design was specified as follows:

```
library(survey)

design_w_cz <- svydesign(
  ids = ~Session_ID,
  strata = ~Road_type + Time_period,
  weights = ~Wx,
  data = data_distraction_cz,
)
```

This defines a stratified one-level cluster sampling design with observation sessions as primary sampling units.

SPI values and confidence intervals

The SPI is calculated as the weighted average of not using mobile device according to Equation 6 for each stratum defined in the template aggregation sheets. Confidence intervals for SPI estimates were calculated using the `svyciprop()` function from the R *survey* package with the “beta” method (Korn and Graubard, 1998). The beta method uses the incomplete beta function, with an effective sample size based on the estimated variance of the proportion. This approach is suitable for proportions close to 0 or 1, which is often the case for road safety performance indicators.

Considering all dataset, the proportion of non-distracted drivers was calculated directly:

```
> svyciprop(~NotUsed, design = design_w_cz, method = 'beta')
              2.5% 97.5%
NotUsed      0.977 0.972 0.982
```

The following R code was used to calculate weighted SPI estimates by selected breakdown variables road type, time period and vehicle type:

```
results_distraction_cz <- svyby(
  ~NotUsed,
  ~Road_type + Time_period + Vehicle_type,
  design = design_w_cz,
  svyciprop,
  method = 'beta',
  vartype = c('se', 'ci')
)
```

Table B.2 shows the estimated proportion of drivers not observed using a mobile device divided by breakdowns. Sample size refers to the number of individual observed vehicles in the corresponding subgroup. Estimates are shown as NA where the sample size does not exceed 30 observations.

Table B2. Distraction SPI results by road type, time period and vehicle type

Road type	Time period	Vehicle type	Sample size	SPI: not distracted	95% confidence interval
motorway	weekday/daytime	passenger car	2065	96.4%	95.3%–97.4%
motorway	weekday/daytime	light goods vehicle	231	93.9%	90.6%–96.3%
motorway	weekday/daytime	heavy goods vehicle	781	93.5%	90.7%–95.6%
motorway	weekend/daytime	passenger car	2730	98.6%	97.7%–99.3%
motorway	weekend/daytime	light goods vehicle	114	97.5%	88.1%–99.9%
motorway	weekend/daytime	heavy goods vehicle	232	96.0%	88.3%–99.3%
rural road	weekday/daytime	passenger car	3757	97.4%	96.5%–98.2%
rural road	weekday/daytime	light goods vehicle	217	93.0%	88.9%–96.0%
rural road	weekday/daytime	heavy goods vehicle	456	94.2%	92.9%–95.4%
rural road	weekend/daytime	passenger car	2137	98.7%	98.3%–99.1%
rural road	weekend/daytime	light goods vehicle	8	NA	NA
rural road	weekend/daytime	heavy goods vehicle	27	NA	NA
urban road	weekday/daytime	passenger car	4771	97.9%	97.0%–98.7%
urban road	weekday/daytime	light goods vehicle	95	97.6%	92.5%–99.6%
urban road	weekday/daytime	heavy goods vehicle	32	96.4%	87.2%–99.6%
urban road	weekend/daytime	passenger car	2015	98.8%	98.1%–99.2%
urban road	weekend/daytime	light goods vehicle	4	NA	NA

In this example, the estimated SPIs is high across most reported subgroups. For passenger cars, the SPI ranges from 96.4% on motorways during weekday daytime to 98.8% on urban roads during weekend daytime. This suggests that observed distraction among passenger car drivers was relatively rare in all road and time-period combinations. For goods vehicles, the estimated SPI values are generally lower than for passenger cars (for example, during weekday daytime). This indicates that, in this example, drivers of goods vehicles were more often observed as distracted than passenger car drivers. Some subgroups are not reported because the number of observations does not exceed the minimum reporting threshold of 30 observations. These estimates are shown as NA to avoid reporting unstable results based on very small sample sizes. The confidence intervals show the statistical uncertainty around the weighted estimates. Wider intervals, especially for some goods vehicle categories, reflect smaller sample sizes and/or higher variability in the survey design. Therefore, differences between subgroups should be interpreted with caution, particularly where confidence intervals overlap.

24 Appendix C – Checklists

The checklists in this appendix turn the methodology of the manual into operational, field-usable steps. Each one opens with a short statement of when to use it, sets out its checks in stages as printable tick-box tables with a notes area for each stage, points to the related checklists, and — where hard minimum standards apply — closes with a short summary of them. They are written to be adapted to national circumstances and extended as experience accumulates.

Checklist 1 — Selecting and target-setting SPIs

Use this checklist when deciding which safety performance indicators to measure and when setting targets for them, at the start of an SPI programme or a measurement round.

A. Choosing the indicators

✓	Item to check
	Confirm the behaviour or condition is causally linked to crash occurrence or injury severity.
	Confirm it can be influenced by policy or enforcement, so that a target would be meaningful.
	Confirm it can be measured with a standardised, reproducible protocol.
	Confirm it can be reported with a measure of uncertainty — a confidence interval or a coverage level.
	Prioritise indicators by the country's main road safety problems, not by ease of measurement.
	Check alignment with the EU and Trendline indicator set, for comparability.
Notes	

B. Setting targets

✓	Item to check
	State the baseline: the most recent national value for each indicator.
	Choose the target type: an absolute value, a relative improvement, or a trajectory.
	Make explicit the minimalist (floor), idealist (aspirational) and pragmatic (realistic) values.
	Set a target that is ambitious yet credible, and supported by the key stakeholders.
	Give the target a horizon of at least two to three years, with annual monitoring.
	Link the target explicitly to the interventions planned to reach it.
Notes	

See also: Part A for what SPIs are and for target-setting; Checklist 11 for reporting results against targets.

Checklist 2 — Sampling and selection of observation locations

Use this checklist when planning where and when to carry out roadside observations for a behavioural SPI. Sound sampling is the single biggest determinant of whether the results are representative. Work through the four stages in order before fieldwork begins; the on-site suitability of each candidate location is screened separately with Checklist 3.

A. Define the sampling frame

✓	Item to check
	Define the study population: which road users or vehicles the SPI concerns (for example car occupants, powered-two-wheeler riders, or all motorised vehicles).
	Set the geographical coverage: national, regional or specific corridors, and decide whether every administrative region must be represented.
	Fix the road-type categories: as a minimum urban, rural and motorway, recording expressways separately where they are used.
	Choose the stratification variables: at least road type; add time period and vehicle type where the SPI requires them.
	Define the strata: each combination of stratification variables is a stratum, with at least ten locations aimed at in each.
	Set the minimum sample sizes: at least 2,000 observations overall and 500 per road-type category, as lower bounds rather than targets.
Notes	

B. Draw the sample of locations

✓	Item to check
	Compile the sampling frame: a list or map of eligible road segments for each road type, from national road-network data.
	Select locations at random within each stratum, by simple random or systematic random selection.
	Meet the minimum of at least ten locations per stratum.
	Spread the locations geographically across regions, rather than clustering them in a few areas.
	Vary traffic volume and road function across the selected sites, and exclude sites with very low relevant traffic.
	Prepare a reserve list of 20–30 % additional candidate sites to replace any that prove unsuitable.
	Screen each candidate site for on-site suitability before confirming it (Checklist 3).
Notes	

C. Plan the observation sessions

✓	Item to check
	Choose the observation season, avoiding school holidays and weather extremes unless a seasonal SPI is intended.
	Spread sessions across the times of day relevant to the SPI, including nighttime sessions for alcohol.
	Cover weekday and weekend; for alcohol, cover all four weekday/weekend by day/night combinations.
	Set a minimum session length sufficient to record enough relevant vehicles (as a working rule at least 100 vehicles, or 30 minutes).
	Plan a simultaneous, independent traffic count at every session.
Notes	

D. Weighting and documentation

✓	Item to check
	Define a weighting factor for each stratum based on its traffic exposure (vehicle-kilometres), using road length as a proxy where exposure data are missing.
	Check that the weighting does not give undue influence to very small strata, and merge thin strata where necessary.
	Document the sampling design, the stratification, the weighting basis, and every deviation from the plan.
Notes	

Minimum standards at a glance

Standard	Threshold
Locations per road-type stratum	At least 10
Observations per road-type category	At least 500
Observations overall	At least 2,000
Breakdown cell reporting	Full at 500 or more; flagged at 250–499; deviation below 250
Traffic count	Simultaneous and independent at every session
Time coverage	Weekday and weekend; all four day/night combinations for alcohol

See also: Part B for the sampling and fieldwork methodology and the confidence-interval framework; Checklist 3 for on-site suitability screening; Checklist 7 for data processing; and Appendix C for the weighting and confidence-interval formulas.

Checklist 3 — Observation-site requirements

Use this checklist during reconnaissance to decide whether a specific candidate location is suitable for roadside observation. A site must pass every section before it is confirmed; a site that fails on safety or visibility is replaced from the reserve list drawn up with Checklist 2.

A. Observer safety

✓	Item to check
	A safe off-road standing position (verge, footpath or lay-by) away from live traffic lanes.
	A safe arrival and parking arrangement that does not itself create a road safety risk.
	High-visibility clothing available for observers, unless wearing it would change the behaviour being observed.
	An identified, practicable emergency exit route.
	The site approved on safety grounds by the coordinator before deployment.
Notes	

B. Visibility and observation quality

✓	Item to check
	A clear, unobstructed sightline giving a few seconds' view of each vehicle as it approaches.
	A suitable observation angle: the driver's window for belt and phone use, the rider's head for helmet use.
	Adequate light for the planned session times, including any evening or night sessions.
	A traffic speed compatible with reliable observation; very high-speed roads need camera equipment.
Notes	

C. Traffic conditions and representativeness

✓	Item to check
	A minimum flow of about 20 relevant vehicles per hour during the planned session.
	Free-flow conditions: no congestion for speed observations, no atypical queuing for others.
	No speed camera, checkpoint or other enforcement point nearby that would change behaviour.
	Traffic not dominated by a single non-representative vehicle type.
	The site genuinely represents its stratum; any substitution is documented.
Notes	

Minimum standards at a glance

Standard	Threshold
Minimum traffic flow	About 20 relevant vehicles per hour
Unobservable (unknown) rate	Below about 15–20 % of passing vehicles
Traffic conditions	Free-flow, with no enforcement point nearby

See also: Checklist 2 for the sampling design that selects the sites; Part B for the fieldwork methodology; and Appendix E for equipment where a camera is needed.

Checklist 4 — Observer selection, training and quality control

Use this checklist when recruiting, training and supervising the field observers, or when verifying that a subcontractor has done so.

A. Selection

✓	Item to check
	Observers are adults with adequate eyesight and the attention the task requires.
	Enforcement-agency staff are not used as observers, unless the police is required (alcohol and drug testing).
	Enough observers are engaged for the schedule, plus a reserve of about 10–15 %.
Notes	

B. Training

✓	Item to check
	Theoretical training is given: aims, variables, definitions and procedures.
	Practical training is given: roadside exercises before live fieldwork.
	The methodology and observation protocol are distributed in advance of the session.
	A field-phase question-and-answer reference is provided for borderline cases (Appendix E).
Notes	

C. Quality assurance

✓	Item to check
	Inter-rater reliability is checked before fieldwork starts.
	Reliability is re-checked at intervals during fieldwork.
	Sessions are supervised and early data reviewed for coding problems.
	Any deviation from the protocol is recorded as it occurs.
Notes	

See also: Part B for the observer-training methodology; Checklist 3 for site requirements; Checklist 6 for ethics; and Appendix E for the field-observer FAQ.

Checklist 5 — Questionnaire and survey design

Use this checklist when designing a questionnaire for a self-reported SPI or for an attitudinal study that accompanies an observed indicator.

A. Content and structure

✓	Item to check
	Define the construct each question measures: (past/ intended) behaviour, attitude, perception or knowledge.
	Use validated items or scales where they exist, for comparability.
	Keep questions clear, single-topic focused, and free of leading wording.
	Order the questions to reduce priming and social-desirability effects.
	Include an integrity or consistency check where a socially sensitive behaviour is asked about.
Notes	

B. Response options and scales

✓	Item to check
	Use consistent, balanced response scales and define the meaning of intermediate values.
	Provide "prefer not to say" or "not applicable" options where appropriate.
	Pre-code the response options for analysis.
Notes	

C. Piloting and translation

✓	Item to check
	Pilot the instrument and revise any ambiguous items.
	Translate and back-translate for cross-national comparability, check meaning of intermediate scale values.
	Record the administration mode (online, telephone, face-to-face) in the metadata.
Notes	

See also: Part B for self-reported data and its interpretation; Checklist 6 for ethics and consent; Checklist 7 for data processing.

Checklist 6 — Ethics, privacy and legal requirements

Use this checklist before and during any data collection involving road users, whether observation, survey or police-supported testing. The requirements apply to every study, regardless of whether a formal ethics approval is needed.

A. Legal basis and data protection

✓	Item to check
	The legal basis for the data collection is established.
	Only the personal data strictly necessary for the SPI are collected (for example gender, age category or vehicle type, only if used for breakdowns).
	No direct identifiers are recorded such as licence plates, names, or images that allow identification.
	The data are used only for the stated research and monitoring purpose, and are not shared for enforcement, insurance or other secondary use.
Notes	

B. Consent and special groups

✓	Item to check
	Consent is obtained where required, in particular for surveys and focus groups.
	Parental or guardian consent is obtained for participants under 18.
	Professional drivers are protected from any individual performance or disciplinary use of the data.
	For alcohol and drug testing, only the police stop vehicles; researchers never stop vehicles themselves.
Notes	

C. Police-supported measurement and reporting

✓	Item to check
	Roles, data-sharing and confidentiality are agreed with the police before fieldwork.
	The anonymous research record is kept separate from the police enforcement record.
	Unfavourable results are reported with the same accuracy as favourable ones, and any conflict of interest is disclosed.
Notes	

See also: Part B for the full ethical and legal requirements; Checklist 4 for observers; Checklist 9 for data access in database studies.

Checklist 7 — Data processing and analysis

Use this checklist when turning collected field data into weighted SPI values with their confidence intervals.

A. Cleaning

✓	Item to check
	Remove test records and any records outside the fieldwork window.
	Detect and resolve outliers and implausible records.
	Resolve duplicates and double-counted sessions.
	Verify the session and location metadata and the road-type classification.
	Document the cleaning rules and the proportion of records removed.
Notes	

B. Weighting

✓	Item to check
	Apply weights by the appropriate exposure measure (in general, traffic volume) per stratum.
	Apply weights by the appropriate exposure measure (in general, traffic volume) when combining strata.
	Use road length as a proxy where exposure data are missing, and document it.
	Check the weights do not over-amplify small strata, and merge thin strata where necessary.
Notes	

C. Estimation

✓	Item to check
	Compute the weighted SPI value for the population and for the required breakdowns.
	Compute confidence intervals by a method that respects the clustered design, not simple random sampling.
	Flag breakdown cells below the sample-size thresholds rather than reporting them as precise.
	Prepare the cleaned dataset and the metadata for delivery.
Notes	

See also: Part B for the data-analysis and confidence-interval methodology; Appendix C for the statistical formulas; Appendix B for the worked example; Checklist 10 for metadata and delivery.

Checklist 8 — Data sources and data access

Use this checklist when planning a database-based SPI, to identify the data needed, find the owner, and secure access in good time. Access lead times are often the longest single item in a database study's timeline.

A. Identify the data needed

✓	Item to check
	Define the indicator's data requirements: which variables, which population, which reference period.
	Identify the authoritative source for each variable (registration register, rating database, enforcement records, EMS records, road register).
	Establish whether one source suffices or several must be combined, noting when they have different owners.
	Check the source's coverage of the target population and reference period.
Notes	

B. Secure access

✓	Item to check
	Identify the data owner and the contact point.
	Establish the legal basis and any data-sharing agreement required.
	Agree the format, variables and reference period to be delivered.
	Agree the lead time, and start the request early
	Establish whether the data are free or must be purchased, and budget accordingly.
	For non-official sources, record the provenance and treat the values as indicative.
Notes	

C. Confirm fitness for purpose

✓	Item to check
	Confirm the source can be mapped to the indicator definition.
	Confirm the source's classification (road type, vehicle type) can be reconciled with the Trendline categories.
	Confirm the update frequency suits the reporting cycle.
Notes	

See also: Part C for the common framework for database SPIs; Checklist 9 for linkage and coverage; Checklist 6 for ethics and data protection; Checklist 10 for metadata and delivery.

Checklist 9 — Data linkage and coverage

Use this checklist when combining datasets and assessing how completely the data cover the target population, for a database-based SPI.

A. Mapping to the indicator

✓	Item to check
	Map each source variable to the indicator definition.
	Reconcile the source's classification with the Trendline categories, and document any approximate mapping.
	Define the inclusion and exclusion rules in line with the definition.
Notes	

B. Linkage, where sources are combined

✓	Item to check
	Identify the linkage keys (for example type-approval number, make/model/year, or location reference).
	Define the matching rule and how partial or ambiguous matches are handled.
	Record the match rate and the treatment of unmatched records.
	Build and check the linkage on a pilot subset before the full run.
Notes	

C. Coverage

✓	Item to check
	Determine the coverage level: the share of the target population the data represent.
	Document how the coverage level was determined.
	Identify and describe any systematic gap (regions, periods or subpopulations not covered).
	Report the coverage level with the value, as the database analogue of a confidence interval.
Notes	

See also: Part C for coverage and metadata; Checklist 8 for data sources and access; Checklist 10 for metadata and delivery.

Checklist 10 — Metadata and data delivery

Use this checklist when preparing the final values, dataset and documentation for delivery to the Trendline coordination team. It applies to behavioural and database studies alike.

A. Values and disaggregations

✓	Item to check
	Deliver the indicator value(s) with the measure of confidence (95% confidence interval for sampled indicators, coverage level for database indicators).
	Deliver the required disaggregations, each with its interval or coverage and its base.
	Flag any breakdown cell below the sample-size threshold rather than presenting it as precise.
	Deliver at the level the indicator chapter requires: minimum, medium or ideal.
Notes	

B. Metadata record

✓	Item to check
	Operational definition and population.
	Geographical coverage and reference period.
	Data source or sampling design, including stratification and locations per stratum.
	For database indicators, the coverage level and how it was determined.
	Collection or extraction method; weighting basis and confidence-interval method (sampled); classification and mapping rules (database).
	Cleaning rules and the proportion of records excluded.
	Indicator-specific metadata (speed limits, blood-alcohol limits, age thresholds, and so on).
	Any deviation from the minimum requirements, with its reason.
Notes	

C. Delivery

✓	Item to check
	Use the standardised Trendline data-submission template (Appendix A).
	Deliver the cleaned dataset where the ideal level is required or national policy releases it, with the anonymisation and licensing decision made.
	Include the methodology note describing the techniques used and any deviations.
Notes	

Minimum standards at a glance

Standard	Threshold
Measure of confidence	Reported with every value (95% interval for sampled; coverage level for database)
Breakdown cell reporting	Full at 500 or more; flagged at 250–499; deviation below 250
Metadata	A complete metadata record accompanies every delivery

See also: Part D for reporting and delivering the data; Appendix A for the templates; Appendix C for the formulas; Checklist 11 for reporting and communicating.

Checklist 11 — Reporting and communicating results

Use this checklist when preparing the national SPI report and the products derived from it for different audiences.

A. The report

✓	Item to check
	Include an executive summary of no more than two pages, with headline values, confidence or coverage, comparison to target, and the key conclusions.
	Include a methodology section: coverage, design and sizes for sampled indicators; source, extraction and classification for database indicators.
	Report every value with its measure of confidence, overall and by stratum, against the target and the previous rounds.
	State the statistical significance of any difference explicitly.
	Make the limitations section prominent, not buried in an appendix.
	End with clear, actionable recommendations — answer "what should we do next?".
Notes	

B. Products for different audiences

✓	Item to check
	Produce the audience-appropriate products (ministerial brief, technical report, press release, ...).
	Explain the measure of confidence in plain terms in every product.
	Give changes in absolute terms ("from 72 % to 78 %"), not only in relative terms.
	Distinguish observed, database and self-reported indicators, and never combine methods in one series.
Notes	

C. Hand-over and dissemination

✓	Item to check
	Brief the communication and policy team, ensuring the metadata and caveats travel with the results.
	Publish the report and send it directly to key stakeholders; archive a cleaned, documented dataset.
	Submit the results to the relevant international programmes.
Notes	

See also: Part A for reporting, communicating and using SPIs; Checklist 10 for metadata and delivery; Checklist 1 for targets.

Checklist 12 — Subcontracting and tender specification

Use this checklist when part or all of an SPI study is to be carried out by a subcontractor, to make the specification clear and to manage the contract. The most damaging error is leaving the state of the deliverable ambiguous, so fix it first.

A. Decide the division of work

✓	Item to check
	Decide where on the spectrum the subcontract sits: turnkey, partial (cleaned dataset), or raw-data-only.
	State explicitly the required state of the deliverable (raw, cleaned, analysed, full report) and the delivery level.
	Decide who carries out verification, cleaning, weighting and reporting.
Notes	

B. Write the specification

✓	Item to check
	Specify the methodology: sampling design, definitions, minimum sample sizes, variables, coverage and breakdowns (Checklists 2 and 3).
	Specify the deliverables, the timeline, and the metadata required (Checklist 10).
	Specify the quality criteria: inter-rater reliability and documentation of deviations (Checklist 4).
	Require delivery in the Trendline templates (Appendix A).
	Set out the ethics and data-protection obligations and, where relevant, the police arrangement (Checklist 6).
Notes	

C. Procure and manage

✓	Item to check
	Allow time for tendering, evaluation and award, and judge bids on methodological quality, not only on price.
	Hold a kick-off meeting, a pilot checkpoint, and regular progress meetings.
	Verify each deliverable against the specification before acceptance: coverage, sizes, metadata, deviations.
	Budget the internal oversight effort, not only the external fee.
Notes	

See also: Part D for subcontracting and the division of work; Checklist 10 for metadata and delivery; Checklist 6 for ethics.

Appendix D — Frequently asked questions for field observers

This appendix answers the questions that field observers most often raise — the borderline cases and practical situations that a written protocol cannot fully anticipate. It is a starting point, distilled from field experience, and is meant to grow: each country should add the cases its own observers meet, so that the reference improves from one round to the next. Where an answer here and the study protocol differ, the protocol prevails, and a query that this appendix does not settle is taken to the coordinator rather than guessed at. The guiding principle throughout is that consistency matters more than completeness: a behaviour you cannot observe reliably is recorded as unknown, never guessed.

A. Before you start

The site I was assigned feels unsafe — what do I do? Do not set up. Observer safety comes before data. Contact the coordinator and move to a site from the reserve list, recording the substitution.

Should I wear the high-visibility vest? Yes, for safety, unless you have been told that being visible would change the behaviour you are observing. In that case position yourself discreetly and safely rather than removing the vest in a hazardous spot.

Where exactly should I stand? Off the road, with a clear view of the driver or rider for a few seconds as each vehicle approaches, at the angle your protocol specifies for the behaviour. Never stand where you would distract drivers or put yourself near live lanes.

B. What to record and what counts

What specifically do I do? First calibration (traffic counting for 10 minutes), then observation, and finally another calibration for 10 minutes. Specific conditions depend on each SPI.

A vehicle passes but I only glimpsed the driver — do I record it? If you could not reliably observe the behaviour, record it as unknown. Do not guess. An honest unknown is more useful than a wrong observation.

Do I record every vehicle, or only some? Record according to the protocol — usually every relevant vehicle, or every relevant vehicle in a defined flow. Do not skip vehicles because they are hard to see; record those as unknown.

Front-seat passenger is partly hidden — how do I code the belt? Record what you can reliably see. Code the driver and the passenger separately; if you cannot see one of them, that one is unknown.

Handsfree or handheld phone — which counts? The behaviour of interest is handheld use. A phone held to the ear or in the hand counts; a conversation through a hands-free system, which you usually cannot observe anyway, does not.

C. Borderline cases

Is a small van a car or a goods vehicle? Small vans are genuinely hard to tell from cars. Follow your protocol's rule; where it is truly ambiguous, code as the protocol directs and note it. Consistency between observers matters more than any single call.

The helmet looks like a novelty or non-standard helmet — worn or not? If a helmet is on the head, record it as worn unless your protocol distinguishes helmet types, in which case follow that rule. Note the borderline case.

Two occupants, one clearly belted and one not — how do I record it? Record each occupant separately, exactly as you see them. The indicator is built from individual occupants, not whole vehicles.

A motorcyclist and pillion passenger — do I record both? Yes, record the rider and the passenger separately for helmet use, following the protocol.

D. Difficult conditions

Traffic is too sparse to reach the target — what do I do? Extend the session if the protocol allows and record the shortfall. Do not move to a location that was not in the sample without recording the substitution and telling the coordinator.

Heavy rain or poor light — do I continue? If observation quality drops below the point where you can record reliably, pause. Record the conditions and consider rescheduling. Reliable unknowns and an honest note are better than poor data.

Vehicles are travelling too fast to observe — what then? Record those vehicles as unknown, flag the site as needing camera equipment, and inform the coordinator. A site where most vehicles cannot be observed is not a suitable site.

E. When you are not sure

I cannot decide between two codes — which do I pick? Record unknown rather than choose between guesses. The aim is not to fill every cell but to record only what was actually observed.

I made a mistake mid-session — how do I correct it? Use the correction procedure in the data-collection application and note what happened. Do not leave a knowingly wrong record in the data.

F. Talking to the public and the police

A driver or passer-by asks what I am doing — what do I say? Explain plainly that it is an anonymous road-safety study, that no personal data, names or licence plates are recorded, and that the results are used only for monitoring road safety. Give the coordinator's contact if asked and never collect anyone's identity.

Should I ever write down a licence plate? Never. No plates, names or identifying images are recorded, in any circumstances.

The police arrive or ask about the study — what do I do? For ordinary observation you do not stop vehicles and have nothing to manage; explain the study. For alcohol or drug sessions the police lead the stopping of vehicles, and you follow the agreed roles, keeping the research record separate from theirs.

G. After the session

What do I do with the data after the session? Submit or upload the records by the agreed procedure, keep no personal notes about individuals, and report any deviation from the protocol — a substituted site, an interrupted session, a difficult coding decision — so that it can be taken into account in the analysis.

See also: Checklist 3 for site requirements; Checklist 4 for observer training; Checklist 6 for ethics and data protection.

Appendix E – Equipment specifications

A. Introduction

This appendix provides specifications for the equipment to be used to collect the behavioural SPIs, as a reference for planning, budgeting and procurement. It expands on the indicative costs given in Part D (

Table 23-5): each cost category there corresponds to one of the equipment families described below. Because the database-based SPIs of Part C draw on existing records and require no measurement equipment of their own, so this appendix is confined to the behavioural indicators.

The guiding principle is that the equipment is chosen to suit the measurement method, and the measurement method follows from the indicator. Speed indicators are measured by instruments that record passing vehicles automatically; restraint use, handheld-device use, helmet use and cyclist lighting are measured by observers, with cameras as an accepted alternative; and alcohol and drugs are measured by testing devices operated within a police enforcement operation, or, where that is not possible, by self-report surveys that need no field equipment at all. The specifications below are minimum functional requirements rather than procurement specifications for a particular make or model. The precise device choice is a decision to be taken by the project team conducting the SPI study. The make, the accuracy and the calibration status are recorded in the metadata in every case.

Several items are normally owned and operated by the police and are used in kind rather than purchased by a study team. This covers evidential breath analysers and LIDAR speed devices in particular. Where a study team does buy equipment, renting or leasing the equipment for a single measurement round is frequently cheaper than purchase. A pilot should confirm that the chosen device performs as required under local conditions.

The table below maps each behavioural indicator to its measurement method and to the equipment family that serves it; the paragraphs that follow give the specifications for each family in turn.

Indicator	Measurement method	Principal equipment
Speed	Automatic spot-speed measurement	Roadside radar or LIDAR; in-road loops or axle detectors; speed cameras; GPS / floating-car data
Restraint use; distraction; helmet use; cyclist lighting	Roadside observation, or camera-based observation	Data-collection device; observation cameras and image-analysis systems; lux meter (cyclist lighting)
Alcohol	Random Breath Testing within a police operation; self-report survey	Passive alcohol sensor; screening and evidential breath analysers
Drugs	Police-led or researcher-led random saliva testing; self-report survey	Oral-fluid screening devices and test kits; (no field equipment for self-report)
All observer-based SPIs	Simultaneous traffic count	Manual tally counters; automatic counters
All roadside fieldwork	—	Personal protective equipment

B. Speed-measurement equipment

Device families

Speed must be measured as the **instantaneous (spot) speed** of individual vehicles at a defined location, under free-flow conditions. Methods that average speed over a length of road, such as Automatic Number Plate Recognition systems, are therefore excluded, because they cannot guarantee free-flow conditions for each vehicle. Because the measurement is usually carried out over a short window and at a large number of locations, equipment that can be installed quickly and flexibly is preferred; this flexibility is also what makes a genuinely random sample of locations feasible.

In practice three families of device are used. **Roadside (out-of-road) devices** are mounted on existing poles, traffic signs or street lights beside the carriageway. The Doppler microwave radar is the most widely used and the most generally recommended: it transmits a constant wave (around 24.5 GHz) that rebounds off the vehicle, and from the modified frequency it derives the instantaneous speed, the pass-by time, the vehicle count and a length-based broad vehicle classification. It is relatively unaffected by weather and can be installed without major disruption to traffic. The frequency-modulated continuous-wave radar performs similarly at a higher cost.

LIDAR works on the same principle with a laser wave and is considerably more accurate, but its high cost and its primary use in enforcement make it less suited to routine SPI work. Active infrared sensors distinguish vehicle types more finely than microwave radar but are more expensive and more sensitive to bad weather. A camera placed above the carriageway can also derive speed, and vehicle length, from the time a vehicle takes to travel between two reference points a known distance apart.

In-road devices are embedded in the carriageway. Inductive loops detect a vehicle passing over a wire loop and yield traffic-volume data directly. Axle detectors, which may be pneumatic, piezo-electric or quartz-electric, count axles and so require vehicle-type correction factors. Their advantage is that they are permanent and continuously available. The disadvantages for SPI work are that installation is intrusive, that they tend to cover only the major network, and that they are therefore poorly suited to a representative national sample and to guaranteeing free flow.

Handheld devices such as radar guns and laser guns are flexible. But they are conspicuous, and their use is therefore not recommended: a driver who notices the device adjusts speed. Moreover, real-time alerting applications spread the position of a manual check rapidly to other road users. They are listed here only for completeness.

Minimum capabilities

Whichever device is chosen, it should at least be capable of the following:

- measuring, storing and delivering the instantaneous speed of individual vehicles
- measuring the number of vehicles (the traffic count)
- measuring the length of vehicles, in metres to one decimal place, since length is the most common basis for vehicle-type classification (another means of classification may be used instead of length where available)
- recording the pass-by time of each vehicle, accurate to the second
- working uninterruptedly and capable of storing data for at least seven days
- capable of collecting data on at least 250,000 vehicles, either internally or by sending it in real time to an external server
- being reasonably unobtrusive (not resembling an enforcement speed camera)
- having a solid, stable installation, calibrated and checked after installation
- protectable against theft and vandalism (an optional small disclaimer can state that the data are not used for law enforcement)
- capable of recording its own speed-measurement accuracy, defined by the mean error and/or the mean square error.

Installation, calibration and the treatment of measurement error

Because poor installation can compromise data quality severely, the equipment should be installed by trained and experienced personnel in accordance with national safety regulations and traffic law. The safety of such people is the dominant field-safety consideration for speed measurement. After installation the device should be calibrated and checked, ideally against a second device of a different type. Where installation is subcontracted, the requirement for installation experience should be stated in the tender.

The values recorded should be those measured by the instrument. The device error margin should **not** be subtracted from measured speeds before applying the limit. Indeed, doing so, as the police often do for enforcement, would in effect redefine a 100 km/h zone as a 103 km/h zone and would bias the share of vehicles recorded as exceeding the limit downwards. A typical 3 % radar error goes both ways, so using the raw measured speeds preserves the symmetry of the error and gives the best estimate of true compliance. The treatment of measurement error should nevertheless be documented in the metadata, so that it can be taken into account when comparing countries.

Equipment for the alternative measurement methods

The alternative methods described in the speed chapter rest on different equipment or data sources and are not part of the standard methodology:

- the stopwatch method needs only two observers, a measured baseline and a stopwatch
- GPS and floating-car data are bought as a subscription, licensed by area and year, and are delivered already aggregated
- camera-based computer-vision systems range from fixed roadside cameras to plug-and-play sensors that process the video on the device and report classified counts and speed bands
- earth-observation methods procure very-high-resolution satellite or drones imagery for defined corridors.

When any of these is used, its limitations are made explicit in the metadata and its values are reported as complementary to, not as a substitute for, the standard free-flow indicators.

Floating-car data and speed traces extracted from smartphones can be used when the underlying sample of drivers can be shown to be representative of the wider population, which in practice is rarely the case. The wider possibilities offered by such data sources are discussed separately in Section 10.9.

C. Breath-testing equipment for alcohol (Random Breath Testing)

Driving under the influence of alcohol is measured by Random Breath Testing within an official police operation, since in nearly all Member States only the police may stop a vehicle. Researchers never stop vehicles, and the devices used are the police's own standard, type-approved and regularly calibrated instruments. Three pieces of equipment are involved, used in sequence:

- A **passive alcohol sensor**, or "sniffer", may be used first to detect alcohol in the cabin air without the driver's active cooperation; it is a rapid screen that indicates whether a breath test is warranted.
- A **screening breath analyser**, typically using an electrochemical (fuel-cell) sensor, then gives a rapid roadside reading of breath-alcohol concentration. Mouth-alcohol effects are handled according to the manufacturer's instructions, and regular calibration is essential to a reliable reading.
- When a positive screen has to be confirmed for enforcement, an **evidential breath analyser** is used; this is a certified, higher-specification instrument, almost always police-owned and operated.

For the SPI, the relevant property of the modern devices is that they record the exact breath- or blood-alcohol concentration rather than a simple pass/fail. This means that the complementary fixed-threshold indicators come essentially for free from data already collected. Study teams are encouraged to capture the recorded concentration rather than only the binary outcome. The devices and their calibration status should be recorded in the metadata. The anonymous research record should always be kept separate from the police enforcement record.

D. Oral-fluid (saliva) testing equipment for drugs

Driving under the influence of drugs is measured in Trendline by random saliva testing. It is normally led by the police and frequently run as an operation alongside breath testing for alcohol. Where police-led testing is impossible, researchers may carry out the testing themselves at locations where drivers are routinely stationary. The equipment in both cases is the **oral-fluid screening device** and its consumable test kits, of the type already in police use across the Member States (for example devices supplied by Dräger, Securetec and other manufacturers).

These devices are screening instruments: they indicate whether one or more target substances are present above the device threshold but do not, in general, return a concentration. The on-site result is therefore essentially binary, a target substance being detected or not. In most jurisdictions a positive screen corresponds to an infringement of the legal limit. Establishing a concentration would require confirmatory blood analysis or advanced laboratory testing, which is too resource-intensive to apply to every driver. The police may run such confirmatory tests on a sub-sample of positive screens to enrich the national report.

Several features of the equipment bear on the methodology and must be reflected in procurement and in the metadata:

- It is very important to document the detectable **panel of substances and the cut-off levels**, since these differ between devices and between countries. Developments in saliva-testing technology are worth monitoring for the sake of comparability.
- A single test can return results on **several substances at once**, which is policy-relevant because combinations of drugs, and of drugs with alcohol, carry far higher crash risks. The kit chosen should support the multi-substance reporting recommended in the drugs chapter.
- **Saliva analysis is slower than breath testing** (several minutes per driver is common) which limits session throughput and requires extra police capacity at busy locations. Two tests should be arranged so that the slower one does not become the bottleneck.
- Drug saliva kits are **consumables with a shelf life** and a cost per test; they should be used within that shelf life.

E. Observation cameras and automated image analysis

Introduction

Cameras support the observation of behaviours that are hard for the eye to record reliably at speed, such as restraint use, handheld-device use, helmet use and cyclist lighting. They are required on high-speed roads where an observer cannot work safely or accurately. A camera-based approach can raise the reliability of the recording, increase the number of observations per session, extend the session (in some cases into the night) and allow the footage to be re-analysed if a new indicator is added later.

These advantages are to be weighed against a substantial equipment cost, the difficulty of finding positions that give a usable angle, the exposure of equipment to theft or damage, the possibility that drivers change their behaviour on noticing a camera, and the privacy questions that any recording in a public space raises. For all these reasons a camera-based method should be tested and validated locally, ideally in a pilot, before it is adopted as standard.

Use of camera images

Two arrangements are in use. In the first, ordinary cameras capture still images or video that are then **coded off-line by human raters**. In the second, **smart cameras** detect the behaviour automatically by image recognition. These can vary in "intelligence", from a not-intelligent camera that photographs every passing vehicle

for manual checking, through a partly intelligent camera that flags presumed cases for manual confirmation, to a fully intelligent camera that classifies the behaviour without manual confirmation.

Cameras may be fixed (unmanned, installed for weeks or months) or mobile (manned, easily moved between locations). Experience with smart cameras for research and SPIs, as opposed to enforcement, remains scarce do far, and so far mostly mobile and not- or partly-intelligent systems have been used.

The detail of the camera arrangement depends on the behaviour. For distraction and restraint use, a downward, frontal view into the cabin is needed. A practical configuration consists of pairing a first camera that detects the licence plate of an approaching vehicle and triggers a second, higher-resolution camera aimed at the driver, with a polarising shield fitted to the second camera to suppress windscreen reflections. For cyclist lighting, the Trendline pilot showed that a single perspective is not enough: a front-view camera detects front-light use best and a rear-view camera detects rear-light use best, so a front-view and a rear-view camera are deployed per direction, with the cameras mounted well above the ground (in the pilot, 8–10 metres) at the intersection observed.

Optical and siting requirements

The resolution and frame rate must be high enough to read the behaviour through the side or front window at the prevailing speed; a high frame rate matters most where vehicles or cyclists pass quickly or at a distance. The viewing angle is the binding constraint: the steeper and more downward the angle, the deeper the view into the vehicle. However, a single position rarely serves all vehicle types equally, since lower vehicles (cars and vans) and higher vehicles (trucks and buses) call for different angles. Windscreen glare and poor light are addressed with polarising filters and, for night or bad-weather use, infrared illumination. Not every camera can be placed on every road type, and cameras designed for deployment on overpasses constrain the random sampling of locations and may be unsuited to urban and rural roads. The siting limits of the chosen equipment should therefore be considered at the sampling stage, because they bear directly on representativeness.

Privacy and data protection

Privacy protection is essential and should be settled before any deployment. Faces are not captured in identifiable form; where a passenger is detected, masking of that part of the image is recommended. Also, licence plates and other identifiers should not be retained in the research dataset. The legal basis for recording in a public space should be confirmed, signage be displayed where the law requires it, and images be stored securely, with access restricted and prompt deletion once coding is complete. These requirements follow the ethics and data-protection rules of Part B and Checklist 6.

Validation and documentation

Because the detection of a behaviour is generally harder than the detection of a vehicle type, an explicit treatment of "unclear" images and an inter-rater reliability check are important. The false-positive and false-negative behaviour of the system should be documented so that camera-based and observer-based values can be compared with appropriate caution. A methodological report should describe technical characteristics of the cameras, the sampling and selection of locations, the vehicle or rider types included, the camera accuracy, the data-collection and coding procedures, the data-quality and correction procedures, and the weighting and analysis.

F. Light-level measurement for the cyclist-lighting indicator

The cyclist-lighting indicator is observed in the dark, defined as a natural light density below 26 lux. A **lux meter** is the most accurate way to verify this threshold, but it need not be used at every session: a reading taken once is used to fix the correct timing of sessions relative to sunrise and sunset, after which sunrise and

sunset times serve as the working basis. In most European countries the threshold is met from about half an hour after sunset in the evening and until about half an hour before sunrise in the morning. The lux meter is an inexpensive item and is the only instrument specific to this indicator.

G. Data-collection (coding) devices

For the observer-based SPIs, field data can be entered directly into a data-collection application running on a tablet or smartphone. Such an approach has advantages compared to recording on paper, because direct coding can reduce transcription error, time-stamp and geo-locate each record automatically, support real-time central collection and speed up cleaning.

A dedicated observation application should be configured to the protocol of the indicator. Offline backup capture should be provided where mobile coverage at the site is poor. Also useful is that a correction facility is available that allows a mis-coded record to be amended during the session. The device needs a battery sufficient for a full session and secure, time-stamped storage. Any such tool should be tested for usability and speed before adoption and shown to be at least as good as paper coding; paper remains an acceptable fallback.

H. Traffic-counting equipment

Every observer-based session is accompanied by a traffic count, run simultaneously and independently of the behaviour observation. A **handheld tally counter** suffices for manual counting, with a separate counter for each vehicle type where disaggregated results are envisaged. An **automatic counter** such as a pneumatic tube or similar may be used where a continuous count over the whole session is wanted.

The count should cover the same lane(s) and direction as the observation, and the count and its duration should always be recorded together so that vehicles per hour can be computed without ambiguity. For the automatic speed devices the count is produced by the same instrument that measures the speeds, so a separate count is not needed.

I. Observer safety and personal protective equipment

Observer safety governs every site, and the basic protective equipment is inexpensive and always provided. High-visibility clothing is worn unless wearing it would change the behaviour being observed — as it can for distraction, where a conspicuous observer leads drivers to put a phone down — in which case the observer is positioned discreetly and safely rather than removing the protection at a hazardous spot. Each site has a safe off-road standing position away from live lanes and a planned emergency exit route, as set out in the site-suitability checklist.

Title: The Trendline manual for data collection and analysis of road safety performance indicators

Main Author(s): Wouter Van den Berghe, Jiří Ambros, Jan Elgner

Published by: Centrum dopravního výzkumu, v. v. i., Líšeňská 33a, Brno, Česká republika

Published in: Brno, Czech Republic, First Edition

Publication year: 2026

ISBN 978-80-88583-01-1 (Online; pdf)