Trend line

KPI Helmet use among Cyclists and Powered two-wheelers (PTWs)

Methodological guidelines

Version 2.4, 28 December 2023

Co-funded by the European Union

KPI Helmet use among cyclists and powered two-wheelers (PTWs). Methodological guidelines

Please refer to this report as follows:
Kšicová E., Moreau N., Zielinska A., Vadeby A., Areal A., Ferrer López S., João Da Silva Barros M. (2023). KPI Helmet use among cyclists and powered two-wheelers (PTWs). Methdological Guidelines. Report produced as part of the Trendline project, supported by the European Union.

Project details:	
Project start date: **Duration:** **Project name:**	15/10/2022 36 months Trendline
Coordinator:	Wouter Van den Berghe & Agnieszka Stelling | SWOV Institute Research for Road Safety Henri Faasdreef 312 – 2492 JP The Hague, The Netherlands
	Co-funded by the European Union

Report details:	
Version: **Dissemination level:** **Date:**	2.4 Public 28/12/2023

Report Author(s):

Kšicová, E., (CDV), Czech Republic
Moreau, N., (VIAS institute), Belgium
Zielinska, A., (ITS), Poland
Vadeby, A., (VTI), Sweden
Areal, A., (PRP) Portugal
Ferrer López, S., (DGT), Spain
Da Silva Barros, M.J., (ANSR), Portugal

Revision history

Date	Version	Description
31/03/2023	1.0	Version from BASELINE project for the first meeting
16/03/2023	2.0	First draft version - modifications according to the discussions and conclusions from the BASELINE project
16/06/2023	2.1	Second draft
31/10/2023	2.2	New version, incorporating small revisions
28/12/2023	2.4	Version published at the webiste

Legal Disclaimer

Table of contents

About Trendline

Trendline brings together 29 European countries (25 EU Member States and 4 countries as observers) for data collection, data analysis, delivery of road safety KPIs and for using these within road safety policies. Trendline is co-funded by the European Union and builds on the experience gained in the Baseline project. KPIs – Key Performance Indicators – are indicators that provide information about factors that are associated with crash and injury risks. At the core of Trendline project are eight KPIs:

Indicator	Definition
Speed	Percentage of vehicles travelling within the speed limit
Safety belt	Percentage of vehicle occupants using the safety belt or child restraint system correctly
Protective equipment	Percentage of riders of powered two wheelers and bicycles wearing a protective helmet
Alcohol	Percentage of drivers driving within the legal limit for blood alcohol content (BAC)
Distraction	Percentage of drivers NOT using a handheld mobile device
Vehicle safety	Percentage of new passenger cars with a Euro NCAP safety rating equal or above a predefined threshold
Infrastructure	Percentage of distance driven over roads with a safety rating above an agreed threshold
Post-crash care	Time elapsed in minutes and seconds between the emergency call following a collision resulting in personal injury and the arrival at the scene of the collision of the emergency services

These 8 KPIs originate from the Commission Staff Working Document 'EU Road Safety Policy Framework 2021-2030 - Next steps towards "Vision Zero" SWD (2019) 283 final.' In addition, some new experimental and complementary indicators will be tested within Trendline (provisional names):

- Driving under the influence of drugs
- Share of 30km/h road lane lengths in urban zones
- Red-light negations by road users
- Compliance with traffic rules at intersections
- Helmet wearing of PMD (Personal Mobility Devices) riders
- Self-reported risky behaviour
- Attitudes towards risky behaviour
- Use of lights by cyclists in the dark
- Enforcement of traffic regulations
- Alternative speeding indicators

For each of the original eight KPIs and the experimental KPIs, a 'KPI Expert Group' (abbreviated as KEG) has been established. Their main role is to draft the common methodological guidelines, to give feedback on questions, and to review the report of the KPI which they are covering.

Website Trendline: https://www.trendlineproject.eu/

Terms and definitions

Bicycle
A road vehicle which has two or more wheels and generally propelled by the muscular energy of the persons on that vehicle, in particular by means of a pedal system, lever or handle (e.g., bicycles, tricycles, quadricycles, and invalid carriages). Included are cycles mountain bikes, Racing bike, Cargo bike, Tricycle and bike with supportive power unit (e.g., electric bikes).

Moped
A two or three-wheeled road motor vehicle which is fitted with an engine having a cylinder capacity of less than 50cc and a maximum authorized design speed in accordance with national regulations. Where limitations concerning the engine displacement are not applicable, a restriction in terms of motor power may be in force. This relates to categories L1 and L2 of the UN Consolidated Resolution on the Construction of Vehicles (R.E.3).

Motorcycle
A two or three-wheeled road motor vehicle not exceeding 400 kg of unladen weight. All such vehicles with a cylinder capacity of 50cc or over are included, as are those under 50cc which do not meet the definition of moped. This relates to categories L3, L4, L5, L6 and L7 of the UN Consolidated Resolution on the Construction of Vehicles (R.E.3).

Driver and passenger
The driver is the one who controls the vehicle (bicycle or PTW), the passenger just rides (on a seat, in a seat, in a wheelchair…) or just pedals (on a two-wheeler).

Motorway
(definition according to Directive 2019/1936/EC)
A road, specially designed and built for motor traffic, which does not serve properties bordering on it and which meets the following criteria:
(a) it is provided, except at special points or temporarily, with separate carriageways for the two directions of traffic, separated from each other either by a dividing strip not intended for traffic or, exceptionally, by other means;
(b) it does not cross at level with any road, railway or tramway track, bicycle path or footpath;
(c) it is specifically designated as a motorway.

Expressway
Road specially built for motor traffic, which does not serve adjacent properties, and:
a) Is accessible only from interchanges or controlled junctions;
b) Is specially sign-posted as an express road and reserved for specific categories of road motor vehicles;
c) On which stopping and parking on the running carriageway are prohibited.
Entry and exit lanes are included irrespective of the location of the sign-posts.
Urban express roads are also included.

Urban roads (or road inside built-up areas)
Public roads inside urban boundary signs.

Rural roads (or road outside built-up areas)
Public roads outside urban boundary signs, excluding motorways and expressways.

Week – daytime
Working week – Monday to Friday 6.00 a.m. to 9.59 p.m.

Weekend – daytime
Saturday to Sunday 6.00 a.m. to 9.59 p.m.

1. Introduction

1.1. Context

The Communication of the European Commission "Europe on the Move – Sustainable Mobility for Europe: safe, connected and clean" of the 13th May 2018 confirmed the EU's long-term goal of moving close to zero fatalities in road transport by 2050 and added that the same should be achieved for serious injuries. It also proposed new interim targets of reducing the number of road deaths by 50% between 2020 and 2030 as well as reducing the number of serious injuries by 50% in the same period. To measure progress, the most basic – and important – indicators are of course the result indicators on deaths and serious injuries.

In order to gain a much clearer understanding of the different issues that influence overall safety performance, the Commission has elaborated, in cooperation with Member State experts, a first set of key performance indicators (KPIs). The KPIs relate to main road safety challenges to be tackled, namely: (1) infrastructure safety, (2) vehicle safety, (3) safe road use including speed, alcohol, distraction and the use of protective equipment, and (4) emergency response. The aim of the KPIs is connected to EC target outcomes.

The Commission Implementing Decision C(2021)5763 final of 5.8.2021 concerning the adoption of the work programme for 2021-2023 and the financing decision for the implementation of the CEF foresaw a technical assistance action for the collection of Key Performance Indicators for road safety in EU Member States. The action builds on a previous CEF support action in 2020-2022 which established the Baseline project to collect 8 road safety Key Performance Indicators (KPIs) in 18 EU Member States. On the 10th of August 2022, a call was published with reference "MOVE/C2/2022-54— Technical Assistance for the development and collection of Road safety Key Performance Indicators (KPI)". A consortium of 25 EU Member States proposed the "Trendline" project to continue and elaborate the work on key performance indicators.

1.2. Purpose and basis of this document

This document presents the methodoligical guidelines for the KPI Protective equipment. It describes the minimum methodological requirements to qualify for this KPI, defined as:

Percentage of riders of powered two-wheelers and bicycles wearing a protective helmet

As data must be presented separately for PTW riders and for cyclists, these guidelines actually concern two KPI measurements: one on helmet use among PTW riders and the other on helmet use among cyclists.

The main target audience for this document are the persons in the participating countries that will collect and/or analyse the data to deliver the KPIs.

The minimal requirements set by the EC for this KPI are described in the Commission Staff Working Document SWD (2019) 283 (referred to as 'SWD'). These requirements are quantified and specified in this document. This is mainly based on expert consultation TRENDLINE Key Expert Group Protective Equipment.

Next to the specification of the minimum requirements to deliver the main KPI and the disaggregated indicators, each section will also include optional supplementary methodological recommendations. Member States can decide whether to follow the minimal requirements only or to extend (part of) their methodology, depending on available means and their own research questions.

2. Scope

2.1. Vehicles

The Eurostat/OECD/ and UNECE Transport Glossary[1] glossary contains the definitions listed in the Terms and definitions chapter for the vehicles.

Although optional, it is recommended to provide a variable "vehicle type" including different vehicle types for both KPIs. For PTWs, we recommend making at least the distinction between "moped" and "motorcycle"; for bicycles, it is at least possible to make the distinction between "bicycle" (non-electric) and "electric bike / e-bike".

2.2. Riders and passengers

The objective of the roadside observation study is to estimate the percentage of powered two-wheelers (PTWs) (motorcycles and mopeds) and cyclists (including electric bicycles) wearing a protective helmet. Therefore, the theoretical population for these two KPIs refers to the total number of kilometres ridden over the national territory by PTWs and by cyclists. Hence, by weighting the results by number of kilometres ridden (or a proxy of traffic volume), the percentage of riders wearing a helmet will also reflect the percentage of kilometres ridden with a helmet.

Some important considerations:

- Monitoring of passengers for cyclists and PTWs is only recommended.
- For PTWs, the minimum for collection is possible to divide into moped and motorcycle.
- For cyclists, data for children (0-14 years old) should be shown separately.

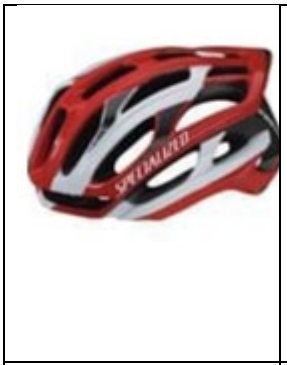
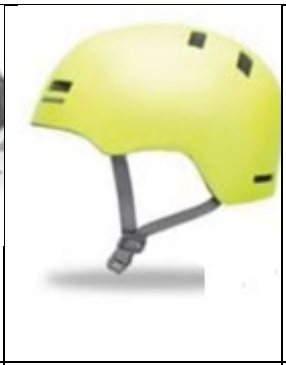
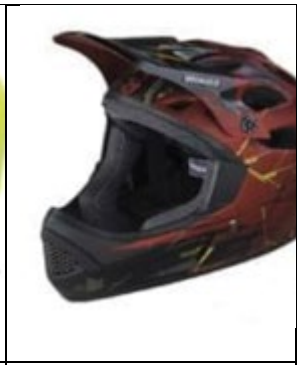
Cyclists should be grouped into men, women (including sex distribution among children) and into age groups 0-14 and 15 and over. If national legislation makes cycle helmets compulsory for children, but using another age limit (e.g., up to 10 years old), this age category should be added. Monitoring of other age categories is also possible, while maintaining the distribution (0-14, 15 and over). The rules regarding the obligation to use helmets shoyld be recorded in the metadata.

2.3. Protective helmets

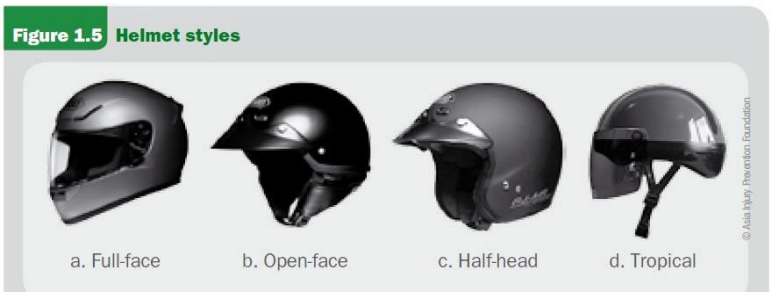
As Hakkert and colleagues (2007) have highlighted: "Under the term helmet, we understand a crash/safety helmet designed for two wheelers, whether motorized or non-motorized." Examples of types of helmets are shown in the figures below.

[1] https://unece.org/DAM/trans/main/wp6/pdfdocs/Glossary_for_Transport_Statistics_EN.pdf - see pages 39-40

Helmets for cyclists:

			
road helmet – a typical bicycle helmet	MTB helmet – a typical bicycle helmet	BMX helmet – a multisport helmet	downhill helmet – a full-face helmet

Helmets for motorcyclists (WHO, 2006):



As legislation on helmet use can vary between countries, it is requested that all countries document their legislation on helmet use regarding each type of vehicle.

2.4. Road types

The KPI should cover use of helmets on motorways, rural non-motorway roads (= roads outside the built-up areas), and urban roads = inside built-up area. Obviously, motorways are only relevant for motorcyclists and not for cyclists and moped riders. This is the assumption in the rest of this document. Rural non-motorway roads are to be interpreted as roads outside built-up areas and urban roads as roads inside built-up areas. The results should be presented separately for these three different road types.

The proportion of observations sampled for each of the three road types should be above 20% to ensure a minimal number of observations for each stratum, even if this would imply disproportional sampling. The three road types should be well defined in the methodology (e.g., typical characteristics, traffic signs, speed regimes, number of lanes, ...).

Thus, for cyclist riders, one should sample locations

- outside built-up areas,
- inside built-up area.

For motorcyclist riders the roads to be sampled include:

- Motorways,
- Roads outside built-up areas (that are not motorways),
- Roads inside built-up areas.

It is possible to monitor other types of roads with relevant traffic or special regulations, such as cycle paths, but it is necessary to observe the minimum required sample of observations for such additional road types (see next section).

3. Measurement procedure

3.1. Introduction

The methodological guidelines for all KPIs are designed to ensure international comparability between KPI values while considering feasibility and affordability. To that end the methodological guidelines have been defined in such a way that accurate and representative results can be obtained for all parameters of interest at a reasonable cost.

Obviously, the larger the sample of observations and locations for observation, the more accurate the KPI estimates for the different strata will be (e.g., a KPI value for a particular type of road, or a particular part of the week). Increasing the number of observations and locations, however, implies increasing field work costs. Statistically, the required minimum sample size depends mainly on the desired accuracy of the final estimates, for which no absolute value can be determined a priori. Therefore, for the main KPI estimates a pragmatic evaluation was made of the expected confidence intervals at different sample sizes and population parameters. The minimum number of observation per vehicle type is 2000 (see next section).

The same pragmatic logic was followed for determining the minimum number of 10 locations for observation for each of the required road types of interest. The measurement locations should be selected as randomly selected as possible. In order to ensure representativeness for the entire country larger numbers of locations might be required for larger countries. It is actually of interest to any country to increase the accuracy of the KPI estimates by boosting the number of locations and the number of observations. Decisions on the final number of locations (above the minimum of 10 per road type) are left to the discretion of the Member States.

3.2. Minimum sample size

Giving priority to feasibility and affordability, for bicycle observations, the minimum number of observations are the following:

- a minimum of 2000 observed bicycles overall,
- a minimum of 500 observations on urban roads,
- a minimum of 150 observations on rural roads,
- a minimum of 325 observations on weekdays and 325 observations on weekend days.

Similarly, for PTWs:

- a minimum of 2000 observed PTWs overall,
- a minimum of 500 observations per road type,
- a minimum of 500 observations per week period.

This should allow to identify statistically meaningful differences between countries at an affordable price. For some countries, this will imply disproportionate sampling of certain strata compared to the distribution of traffic volumes over different strata. This is however required to allow statistically meaningful international comparisons at the level of each of the strata at interest.

When considering the minimum sample size, it should be noted that this refers to the number of **observations** that include the minimum requested data (i.e., excluding observations with missing values in relation to the minimal requested data). The minimum number of observations should be understood as the minimum number of vehicles observed. Some of the vehicles observed will have not just a rider but also a passenger.

Countries that are not able to achieve the minimum requested number of observations will need to indicate the reasons in the information with the metadata.

It is possible to monitor other types of roads with relevant traffic or special regulations, but it is necessary to define them in the metadata and it is necessary to observe **a minimum sample of observations** per road type and week period.

3.3. Stratifaction and subpopulations

The SWD document requires to take into account:

- Road types (3): motorways (only for PTWs), rural roads (or roads outside built-up areas), and urban roads (or roads inside built-up areas).
- Periods (2): weekday / weekend daytime.

Within Trendline, additional requirements are:

- Age group (2): child 0-14, adults 15 and more (for cyclists).
- Sex (2): man, woman (for adult cyclists).

It is possible to classify the sex as "not determined", in case the sex is not clear to the observer. Nevertheless, the percentage of helmet use will be determined.

In case nation-wide surveys would hide major regional differences, countries might have to consider an additional stratum related to regions.

Since the overall estimate is expected to be representative for the total of all kilometres driven in a country, theoretically the optimal strategy to estimate the overall prevalence would be to sample all strata according to traffic volume of each combination of the different strata. This strategy would, however, be detrimental for the accuracy of specific low volume strata that are of interest. Certain road types could have a lower traffic volume than others, as do weekends compared to weekdays. As a result, a strictly proportional sampling would lead to much smaller confidence intervals for certain strata.

For representativeness, the recommended **minimum numbers of locations** are therefore:

- 10 locations per stratum in the first stratification level (i.e., in this KPI, 10 locations for each of the 3 road types; and 10 locations for each of the two week periods (week/weekend daytime)),
- at least 2 locations for each stratification combination (e.g., 3 road types X 2 periods = 6 combinations and a minimum of 2 locations in each combination).

If Member States aim to further distinguish age groups in reported results, a minimum number of observations shall apply to each group. It is recommended that each age group includes at least 500 observations.

3.4. Sampling and selection of locations

The selection of locations should be as **random** as possible, covering the geographical area of the country, optionally using regions like NUTS1 regions (e.g., stratified random sample). The basic process for the choice of locations consists of three steps:

(1) The required number of different locations (for the country or per region) is determined.

(2) The number of locations is randomly selected on the map using the entire area under consideration (e.g., country or region), taking a sufficient geographical spread into account. The specific requirements for each location do not have to be taken into account at this point. This step is to ensure a reasonable geographical spread of the randomly selected locations.

(3) The final locations that will be used for the observations are manually chosen in the area surrounding the locations randomly selected in the previous step. At this point, the final selection must be based on the location requirements (different road types), inclusion/exclusion criteria (if applicable), and practical considerations. This final selection may be done using Google Street View. Care should be taken to ensure that the different road types are also sufficiently geographically spread.

A convenient way of selecting locations randomly (step 2) is to use a GIS system (e.g. cartographic software like ARCView/ARCGIS) as such software can automatically select location points within defined areas randomly (e.g., https://desktop.arcgis.com/en/arcmap/latest/extensions/geostatistical-analyst/an-introduction-to-samplinghttps://desktop.arcgis.com/en/arcmap/latest/extensions/geostatistical-analyst/an-introduction-to-sampling-monitoring-networks.htmmonitoring-networks.htm). If Member States have no GIS software, step 2 can also be carried out manually using a national geographic map, e.g., Google Maps/Google Earth.

A random selection of locations will often include locations with low traffic flow for each stratum. If traffic flow is too low, it is acceptable not to include these locations. A minimum traffic flow for selecting a location can be defined as at least 10 relevant vehicles (PTWs or cyclists) per hour. Also, locations where the composition of the traffic deviates significantly from normal traffic (e.g., locations where 90% of the cyclists are sports cyclists) should be avoided.

If it is not possible to identify 10 locations with at least 10 relevant vehicles per hour for the 1st stratification level indicators, we recommend different alternative strategies to reach the minimum number of observations:

- including summer months in the measurement period to increase the probability of traffic volume for PTWs and cyclists,
- increasing the number of locations (with few vehicles per hour).

If these strategies do not allow to reach the minimum number of observations within factors of the 1st stratification level, it will be accepted to exclude these stratification level indicators. Countries facing this issue will estimate the KPI per available stratification level and no national KPI will be estimated.

The method used for location selection should be described in the information with the metadata. The rationale for choosing the observation locations should be provided. Basic characteristics of the locations should be recorded: coordinates (if possible), address or other geographical information, target lane or path and direction to be observed, traffic signs, speed regime, number of lanes, traffic flow and visibility of the traffic from the location.

4. Observation method

4.1. Direct observtion

SWD prescribes **direct observation** as the data collection method and allows the use of **cameras** if appropriate. Direct observation should preferably be carried out along the roadside (or another convenient place). If the use of cameras is adopted, they should not be installed exclusively on one type of road so as to avoid selection bias.

For both KPIs, observations of helmet use on urban and rural roads can be carried out from a safe place along the road, preferably at locations where driving speed is reduced relative to the speed limit, such as intersections. For PTWs on motorways, observations of helmet use could be carried out at the last intersection before on-ramps, at the first intersection after an off-ramp, after the exit to a petrol station, or from the bridge over the motorway.

4.2. Observe procedure

For both KPIs on helmet use, the most straightforward approach involves observing one bike or PTW, encoding the data, and then observing the next passing bicycle/PTW. When it is not possible to code the observational data for all the road users who pass by, cyclists and PTW riders should be randomly selected from all the possible road users at the observation location. The easiest and most efficient way is after coding one observation to observe the next passing target road user.

The observations must be made by well-trained independent observers (not uniformed police or other officers) under the supervision of a coordinator. Observers should receive rigorous theoretical and practical training and be given clear guidelines about the road section and traffic direction they should observe, the duration of observation periods and how to manage any potential difficulty that would hamper the data collection. They should be given clear guidance on the procedure to be followed when observations cannot be performed (due to weather conditions, concerns with visibility, safety problems, etc.) (Hakkert & Gitelman, 2007). They should also have a clear definition of "helmets" valid for each country. Wherever possible, it will be valuable to ensure consistency between observers (the inter-rater reliability) before the start of and during the fieldwork.

It is recommended that the KPIs on helmet use are measured by two observers (one for PTW riders and one for cyclists). If data for both KPIs are collected during the same measurement by only one observer, a rule should be defined to determine the next observation subject (cyclist or PTW) in the case of high traffic volumes.

One observation session should last at least 30 minutes (ideally 1 hour), excluding the time needed for counting traffic and collecting environmental data (see section 5.2). Each observation location can be used for different observation sessions (at different time intervals) or each location can be assigned (randomly) to a specific time interval.

The fieldwork procedure should be described in the information with the metadata.

4.3. Temporal requirements

Data collection should be carried out during daylight hours; observations should cover all the daytime. One may organize the observation sessions during different periods of the day (e.g. morning, noon, evening) or taking account of peak hours (e.g. 07:30 to 10:30 (AM peak), 12:00-15:00 (inter-peak) and 16:00 to 19:00 (PM peak)). In such cases, it is recommended that each location is observed during all the different periods. Dates and hours of the measurements should be reported in the meta-data.

Helmet use has to be observed both during **weekends** and during **weekdays**, because the purpose and duration of riding may vary considerably between weekdays and weekends. The KPIs should be presented separately for weekdays (excluding bank holidays) and weekends.

When planning the observation periods, one should ensure a balanced combination of road types and time periods, in order to avoid a systematic sample bias (e.g. all motorway observations at the weekend and all other roads on weekdays; or all motorway sessions in the morning and all urban sessions in the afternoon). The same balance should be sought across all combinations of periods and other time considerations, avoiding, for example, all the sessions during the weekdays being planned for the morning only.

It is recommended to implement the measurement at the **end of spring or at the beginning of autumn**. In principle, all months are allowed except December, January, July, and August (in some Member States June also). However, for countries facing difficulties in reaching the minimal number of observations, the measurement can be extended to summer months. In the interests of representativeness, sessions during official feast days and holidays should be avoided.

Member States willing to measure helmet use during two seasons (in late spring and early autumn) can apply the minimal sample size requirements for the two measurements together. The data from both sets of observations can be combined to deliver the main and disaggregated indicators. When Member States have historical series of measurements, it is recommended to use the same period(s) of the year as for the earlier measurements.

4.4. Requirements for automatic detection via roadside cameras

SWD allows the use of cameras to collect data on helmet use by cyclists and PTW riders; after recording, the still or video images can be analysed to encode the data. In some applications, helmet detection is automatically performed by the software. There are some clear advantages in using cameras instead of observers, particularly in terms of, for example, reliability and duration of the observation sessions (including night-time use, although this not relevant for the TRENDLINE project).

Possible disadvantages should however be evaluated, such as privacy/GDPR issues (identifiability of riders) and the risk of lacking key variables. This technology should be tested and validated before use. On account of privacy issues, faces should not be caught on camera. Each country will have to deal with national requirements regarding the ethics and protection of private lives.

It should be ensured that the cameras are installed on all types of roads to avoid selection bias.

4.5. Optional elements

4.5.1. Regions

Disaggregated data for regions is not part of the minimum requirements. Member States are free to choose supplementary stratifications by region within the country (e.g., NUTS 1 regions). One can consider collecting data from each region or from a representative selection of regions. Member States that want meaningful KPIs at regional level should apply the minimum methodological requirements to each region (see Section 3). If stratification by regions is used, results should be weighted according to traffic volumes by region.

4.5.2. Optional additional data collection

One may consider the observations as an opportunity to collect additional data related to helmet use or other behaviours among PTW riders and cyclists. Some examples to consider are:

- types of bicycles (electric or not, city/sport bike, ...),
- types of PTW (moped, e-moped, motorcycles of certain types, ...).

Other possible parameters to expand the data are listed in the annex.

5. Data analysis

5.1. Statistical analysis and post stratification weights

Specifications on calculating weights and confidence intervals are provided in Appendix 2 *Suggested approach for weighting sample data and calculation of statistics*. Although this appendix has been written for vehicles, the approach presented is also applicable for cyclists and PTW-riders.

5.2. Traffic volume and traffic counts

The weighting by traffic volume for cyclists and PTWs in the country is ideally based on national mobility data (driven vehicle-kilometers). If such traffic volume data is not available, it is recommended to use a combination of data on road length with traffic counted during the observations. If no official data on road length are available, it is recommended to request estimates from experts from the relevant administration services.

For both KPIs on helmet use (among PTWs and among cyclists), traffic counts should be performed at each location and during each observation session. The purpose is to count all the relevant vehicles. For KPIs on helmet use, this means that each PTW or cycle who rides in the same direction as those who are being observed will be counted. Note that if the observers succeed in observing helmet use among all PTWs (or cyclists) who pass by, there will be no need for the traffic count as it will be equal to the number of observations.

It is recommended to count the traffic of PTWs and bikes for a minimum of 10 minutes (either 5 minutes before and 5 minutes after the observation, or 10 minutes in the middle of the observation session). These counts should then be extrapolated for the whole duration of the session.

It is stressed that traffic volume should also be counted even when national traffic volume statistics according to road type are available. This information is necessary to calculate the share of helmet users.

5.3. KPI values to be calculated

The main indicator is the percentage of riders wearing a helmet across all times and locations, separately for users of bikes and PTWs, and for riders and passengers. So overall there are 2 main KPIs for which a point estimate and a 95% confidence interval is to be calculated:

- riders of bicycles (including e-bikes),
- riders of PTWs (mopeds and motorcyclists),

In some countries, the number of passenger observations is expected to be very low. Countries facing this limitation will not be included in the results on helmet use among passengers.

It is also recommended to distinguish values:

- riders of e-bikes separately,
- passengers of bicycles (including e-bikes or separately),
- age of cyclist – 0-9 / 10-14 / 15- 19 / 20 years and more or 20-60 and 60 and more,
- riders of mopeds or motorcyclists separately,
- passengers of mopeds or motorcyclists separately.

For each of the 2 main KPIs, it is also required to calculate a point estimate and a corresponding 95% confidence interval for the disaggregated levels:

- road type (3 levels) (motorways, outside the built-up area, and built-up area),
- period of the week (2 levels) (weekdays and weekend days),
- region (if applicable).

It is recommended to provide specific estimates for combinations of levels (including the confidence interval).

Member States should provide the meta-data of their data collection and deliver this together with the dataset(s). Final info on this will be provided together with the Trendline Datafile info.

References

European Commission (2017). Monitoring Road Safety in the EU: towards a comprehensive set of Safety Performance Indicators. European Commission, Directorate General for Transport https://ec.europa.eu/transport/road_safety/sites/roadsafety/files/pdf/ersosynthesis2017-detailhttps://ec.europa.eu/transport/road_safety/sites/roadsafety/files/pdf/ersosynthesis2017-detail-performanceindicators15_en.pdfperformanceindicators15_en.pdf

European Commission (2019). Commission staff working document EU road Safety Policy Framework 2021-2030 – Next steps towards "Vision Zero". SWD (2019) 283 final. Retrieved from https://transport.ec.europa.eu/system/files/2021-10/SWD2190283.pdf

European Commission (2022) Facts and Figures Urban areas. European Road Safety Observatory. Brussels, European Commission, Directorate General for Transport. https://road-safety.transport.ec.europa.eu/system/files/2022-07/ff_roads_inside_urban_areas_20220707.pdf

European Commission (2022) Facts and Figures Roads outside urban areas. European Road Safety Observatory. Brussels, European Commission, Directorate General for Transport. https://road-safety.transport.ec.europa.eu/system/files/2022-07/ff_roads_outside_urban_areas_20220707.pdf

European Commission (2021) CARE DATABASE. CaDaS Common Accident Data Set. 04/06/2021 Version 3.8. Care Team. Directorate-General for Mobility and Transport.https://road-safety.transport.ec.europa.eu/system/files/2021-07/cadas_glossary_v_3_8.pdf

Hakkert, A.S and V. Gitelman (Eds.) (2007). Road Safety Performance Indicators: Manual. Deliverable D3.8 of the EU FP6 project SafetyNet. https://www.dacota-project.eu/Links/erso/safetynet/fixed/WP3/sn_wp3_d3p8_spi_manual.pdf

Hakkert, A.S, Gitelman, V. and Vis, M.A. (Eds.) (2007). Road Safety Performance Indicators: Theory. Deliverable D3.6 of the EU FP6 project SafetyNet. https://www.dacota-project.eu/Links/erso/safetynet/fixed/WP3/sn_wp3_d3p6_spi_theory.pdf See also https://www.dacotahttps://www.dacota-project.eu/Links/erso/safetynet/content/wp_3_safety_performance_indicators_1.htmlproject.eu/Links/erso/safetynet/content/wp_3_safety_performance_indicators_1.html

Helmets - A road safety manual for decision-makers and practitioners, WHO, 2006 (p 35) – https://www.who.int/violence_injury_prevention/publications/road_traffic/helmet_manual.pdf

Moreau, N. et al. (2021). Methodological guidelines – KPI Helmet use of Cyclists and Powered Two-wheelers (PTWs). Baseline project, Brussels: Vias institute

Yannis, G., Folla K. (2022). Baseline report on the KPI Helmet use among Cyclists and Powered two-wheelers (PTWs). Baseline project, Brussels: Vias institute

Appendix 1. Extracts SWD document

COMMISSION STAFF WORKING DOCUMENT
EU Road Safety Policy Framework 2021-2030 – Next steps towards "Vision Zero", SWD (2019) 283 final https://transport.ec.europa.eu/system/files/2021-10/SWD2190283.pdf

Rationale

The use of a protective helmet is often cited to be an essential passive safety measure for powered two-wheeler riders (for whom it is mandatory) and for cyclists.

Definition of the KPI

Percentage of riders of powered two-wheelers and bicycles wearing a protective helmet.

Minimum methodological requirements

Data collection method	Direct observation (if appropriate, using cameras).
Road type coverage	The indicator should cover motorways, outside the built-up area, and built-up area. The results could be presented separately for the three different road types if available.
Vehicle type	The indicator should include riders (also passengers) of powered two-wheelers (motorcycles and mopeds) and cyclists (including those riding power-assisted bicycles). Results should be disaggregated for driver and passengers. The results for bicycles should be presented separately. Where available, data for children should be shown separately, to take into account any legal requirements.
Location	Random sample (methodology for Member States to decide).
Time of day	Observations to take place during daylight.
Day of week	Separate observations for weekdays and weekend and data to be shown separately.
Month	Late spring, early autumn.

NB: A note should accompany the results for cyclists stating the existing state of the legal requirements (or the absence of requirements) concerning helmet use.

The additional requirements established within Trendline are:

Vehicle type	The indicator should include cyclist riders and PTW riders and passengers (motorcycles and mopeds). Results for PTWs should be disaggregated for drivers and passengers. Results for bicycles and PTWs should be presented separately.
Sex	Men, women.
Age	The basic age distribution for a more accurate comparison is 0-14 (children) and 15 and over (adults).

The minimum requirements may be extended by:

- other age categories,
- division into moped and motorcycle,
- addition of the e-bike category,
- addition of other means of personal mobility, including their electrical modifications,
- other categories of roads,
- option to use „not determined" for sex (if it is not clear).

Examples to consider for further data collection:

- type of helmet (full face, open face, half helmet, tropical, ...),
- correct use of the helmet (properly fastened, not properly fastened, not fastened at all, and wearing the helmet backwards) (WHO, 2006),
- use of other protective equipment (gloves, jacket, trousers, shoes, ...) (Hakkert & Gitelman, 2007),
- the colour of the helmet (dark, with lights, reflective, ...),
- wearing of reflective clothing,
- private or a public/shared vehicle,
- professional/non-professional rider,
- age category,
- wearing earphones (only for cyclists),
- using mobile phone.

Appendix 2. Suggested approach for weighting sample data and calculation of statistics

A. Introduction

Within Trendline, several of the "KPIs" (Key Performance Indicators) refer to the relative number of vehicles or road users that respect certain legal limits and rules. These are sometimes called the "behavioural" KPIs. They refer to speeding, driving under the influence of alcohol, use of protective equipment, wearing a seatbelt or distraction.

In general, it is impossible to measure the performance of all vehicles at all times. Therefore, the KPI values are actually estimates based on a sample of vehicles and/or road users observed or surveyed. The main aim of these estimates is to estimate the percentage of kilometres driven on the entire road network (over a period of time, which one could be set to one year for instance) by vehicles respecting the legal limits and rules.

In term of sampling this means that the statistical population to be considered is the total traffic volume (typically expressed in kilometres driven) of moving vehicles over a certain area (i.e. country or region) over a certain period of time (e.g. one year). Estimates are made by sampling individual vehicles (or road user) at particular locations and moments in time. Hence the question arises as to how each of these individual observations have to be weighed in order for the overall average or percentage to reflect the overall percentage of vehicles complying with the rules in the total population.

For many KPIs within the Trendline project, data is being collected during observations (e.g., for distraction by mobile phone) or surveys (e.g., for driving under the influence of alcohol) at different locations. For all behavioural KPIs sampling on three different road types is required (motorways, rural roads, urban roads). For some KPIs sampling of different time periods and/or vehicle types is also required (for other KPIs only one type is considered).

Sampling is done in 2 steps:

1) Random selection of locations. Most beneficiaries use a disproportionately stratified random sample of locations, e.g., a same amount of locations per considered road type.
2) Random selection of vehicles/road users (nested) in each session.

The minimum number of locations for observations or surveys in Trendline is 10 per road type. At a given location, there may be several observation sessions. If different time periods are required in the sampling, then time periods should be linked to locations in a balanced way and also a minimum of 10 locations per time period is required as well as minimum of 2 locations for each combination of road type and time period. These constitute the sessions.

The data collected during these sessions allows to calculate a KPI value for that session, and, if sufficient data are available, also for subcategories (e.g., male/female; position in the car, type of vehicle). Moreover, for every session at least the road type is coded:

- Motorways,
- Urban roads,
- Rural roads.

These are the generally required minimum sampling strata for the behavioural KPIs.

For most behavioural KPIs also a time period is coded for the observation session, specifically:

- Weekday,
- Weekend day.

For drink driving, four time periods are considered (Weekday daytime, Weekday nighttime, Weekend daytime, Weekend nighttime). For some KPIs (e.g., distraction) only one time period is considered (weekday daytime).

Each combination of road time and time period should be considered as a separate stratum: a combination of 3 road types and 2 time periods would lead to 3x2 or 6 strata.

Calculating KPIs for crossed strata of road type x time period is generally not minimum required but recommended, in particular if these categories have been a part of a sampling strategy. For such strata to include sufficient and sufficiently reliable data, a minimum requirement is that for each stratum (combination of road type and time period) minimum 2 different locations are used (but more are recommended).

There is a need to **weight** the results at the observation locations **within** the stratum (to arrive at the best estimate for the KPI value within the stratum) but also **across** the strata (to obtain, for example, a value for all considered time periods or for all roads together).

For certain KPIs other breakdowns are also possible (or even required), such as region, vehicle type/road user or sex. In such cases the number of strata that can be considered will be higher. However, in general strata with less than 500 data points should not be considered for calculating KPIs (unless specified differently in the minimum requirements of the methodological guidelines for the KPI), because the number of different observations and/or observation locations is too small and/or confidence intervals will be too wide. When strata with less than 500 observations are obtained and delivered to the Trendline coordination team, they will be treated differently in the tables and graphs of Trendline reports (e.g., shown in another colour or marked with an asterisk). However, such strata could be combined with or added to other strata to achieve this minimum. For instance, "weekday daytime" and "weekday nighttime" could be combined to "weekday".

B. First step: processing the data of each stratum individually

For each stratum (in the example above each of the 6) the following steps should be followed. Suppose you have *K* survey sessions in that stratum. For instance, you may have 6 observation sessions for observations on urban roads during weekdays. In that case, *K = 6* for that stratum.

For each survey session *k* (with *k* varying between 1 and *K*) the traffic count(s) need to be determined. The traffic count obtained may concern all vehicles (or vehicles of a certain type) that passed by during the entire observation session, or for a fraction of the period (e.g. for 10 minutes in the middle of the session or for 5 minutes before and 5 minutes after the session). The duration of the counting is important. Please register both the actual count of the number of relevant vehicles and the time used to count. In case you have grounds to believe that the traffic density during the observation/survey session is quite different from the density during the counting session (e.g. because there was a sudden traffic

jam causing much less vehicles to pass by during the observation, or because there was a bridge opening during the counting session), it is also useful to make an estimate of the number of relevant vehicles that passed by during the survey session. This estimate is somewhat redundant but would allow for unique unexpected situations.

Often it is planned that all observation or survey sessions have the same length of time (e.g., 60 minutes). This can be considered as the "standard duration" of a session. However, in practice, the duration of a session may deviate from the standard value, and this variation has to be accounted for when weighting the results.

So, for the session *k* in the stratum the following data is recorded:

Duration of the period used to count passing vehicles	$t_p(k)$
Number of passing (relevant) vehicles counted during the counting period	$N_p(k)$
Duration of the observation session	$T(k)$
Relative duration of the observation session = $\frac{T(k)}{Standard\ duration}$	$d(k)$
Estimated total number of (relevant) passing vehicles during the observation session, usually[2] this is equal to $N_p(k)\ x\ T(k)\ /\ t_p(k)$	$N(k)$
Number of (relevant) vehicles/individuals surveyed during the observation session	$n(k)$

It is important to have a good estimate of the total number of vehicles that passed this survey location during a session (this is $N(k)$). Otherwise, we do not know what share the individual survey sessions have within the stratum.

It is considered acceptable to assume that what is observed amongst the surveyed vehicles – $n(k)$ – is representative for all passing vehicles. Therefore, each surveyed vehicle represents $N(k)/n(k)$ vehicles in a session[3]. If the observation session took (a little) longer or shorter than the standard duration of the observation session (often the standard duration is 1 hour or 60 minutes), we can correct for that too (this is d(k)), yielding an observation weight for this vehicle type in this session in this stratum of:

$$Weight\ of\ observations\ in\ session\ k = W(k) = \frac{N(k)}{n(k) \times d(k)} \qquad (1)$$

When these weights are applied to all individual survey observations, the weights should add up to the number of vehicles that passed on all sessions in the stratum, had they been identical in duration.

[2] In exceptional cases where the traffic during the counting session is not representative for the traffic during the observation session, use the best estimate $N_h(k)$ (i.e. estimate of the total number of (relevant) passing vehicles 'per hour' during the observation session).

[3] If an observed vehicle represents 4 vehicles in the session, we have just one observation, not four, but it 'weights' for four vehicles

C. Calculate the KPI value per stratum

Now it is possible to create a database table or a spreadsheet with columns: this weight $W(k)$ and the actual observed values (surveyed vehicles – if required also vehicle type) and results noted as V(k), possibly augmented with administrative information (where, when, etc.) and further breakdowns (e.g., gender, position, …) but keeping an eye on privacy of sensitive data. For instance, the observations of using a seatbelt in a survey could be ordered in the way as indicated in *Table 1* below (the other variables would concern the position of the person, whether he/she is driver or not, sex, …).

Table 1. Data to be collected per observation

Date	Time	Location	Road type	Vehicle type	Time period	Within Stratum Weight W(k)	Seatbelt	Other variables
1-May-23	12:15	Site 51	Rural road	Passenger car	Weekend day	4	1	…
1-May-23	12:16	Site 51	Rural road	Passenger car	Weekend day	4	0	…
1-May-23	12:16	Site 51	Rural road	Truck	Weekend day	3	1	…
…								
2-May-23	12:15	Site 52	Urban road	Truck	Weekday	5	1	…
2-May-23	12:16	Site 52	Urban road	Passenger car	Weekday	3	1	…
2-May-23	12:16	Site 52	Urban road	Passenger car	Weekday	3	0	…
…								

Per session the KPI value V(k) can then be calculated as the average value of all observations. If a "positive" observation is given a score of 1 and a negative observation a score of 0, the average value is then a value between 0 and 1, which can be expressed as a percentage. We can then obtain a table with summary data on all the sessions. *Table 2* gives such information for the example of a stratum of passenger cars observed on weekdays on rural roads.

Table 2. Example of summary data of all sessions within a stratum

Session	Road type	Vehicle type	Time period	Observed vehicles n(k)	Within Stratum Weight W(k)	Seatbelt use V(k)
1	Rural	Passenger car	Week day	120	4.4	88.6%
2	Rural	Passenger car	Week day	110	3.8	92.7%
3	Rural	Passenger car	Week day	95	6.1	94.3%
4	Rural	Passenger car	Week day	130	2.6	78.6%
5	Rural	Passenger car	Week day	118	3.7	84.5%
6	Rural	Passenger car	Week day	84	4.1	94.3%
7	Rural	Passenger car	Week day	156	3.3	92.1%
8	Rural	Passenger car	Week day	124	4.0	86.2%
9	Rural	Passenger car	Week day	130	2.8	87.4%
10	Rural	Passenger car	Week day	145	2.7	88.1%

The formula for the KPI value of that stratum with *K* sessions is then:

$$KPI\ Value\ of\ the\ Stratum = \sum_{k=1}^{K} \frac{n(k)*W(k)*V(k)}{\sum_{1}^{k} n(k)*W(k)} \qquad (2)$$

For the example, the KPI value of the stratum would be 89%. For each different stratum, in general a different KPI value will be obtained.

D. The case of several vehicle types, road users or further breakdowns within the stratum

For some KPIs it is desirable or even required to make a distinction between several vehicle types and/or road users. This implies that each of these subgroups should be considered as a separate stratum; the logic discussed above should be applied to each considered vehicle or road user type.

However, this supposes that you can also count these different types during the traffic count in each session. If that is not possible, then you should assume that the distribution of vehicles passing by is the same as that of the vehicles observed/surveyed. This assumption is justified as the general rule during the fieldwork is to observe (or survey) the first arriving vehicle after coding the former one (random sampling - no deliberate over- or under-sampling of a specific vehicle/road user type).

This means that you have to adapt $N(k)$ above accordingly and use a value of *n(k)* per considered vehicle/road user type.

Other variables like age category and sex are generally no specified sampling strata in behavioural measurements on the road but collected variables of the surveyed road users[4]. If you, for instance, also want to make a distinction between male and female drivers, then the same assumption applies that the relative number of females in the set of the observed vehicles is the same in the set of the vehicles passing by.

E. Aggregation of the KPI results of different strata

From a policy perspective it can be useful to aggregate the data, for instance to arrive at a national indicator taking into account all road types, time periods and vehicle types. This is also desirable and often required within Trendline.

If two (or more) strata need to be aggregated, the relative importance of each stratum within the aggregation (sum) needs to be assessed. Within Trendline, the relative importance is based on the (estimated) volume of traffic in each of the strata. If the first stratum represents (or is representative for) 50% of traffic volume, the second represents 30% and the third 20%, the aggregated value is:

$$\text{Aggregated KPI value} = 0.5 \times KPI \text{ value stratum } 1 + 0.3 \times KPI \text{ value stratum } 2 + 0.2 \times KPI \text{ value stratum } 3.$$

Thus, more general,

- if there are *M* strata to be aggregated
- let *TR(i)* represent the relative traffic volume of stratum *i* (*i* ranging from 1 to *M*)
- let *KPI(i)* be the KPI value of stratum *i*

Then:

$$Aggregated\ KPI\ Value = \sum_{i=1}^{M} TR(i) * KPI(i) \quad (3)$$

If crossed strata are considered, traffic information can come from different sources (e.g., national counts on roads for the proportions on the road types, and online representative mobility survey data for the relative proportions according to time period) which should be combined in a logical way to calculate a traffic volume % for each stratum (all summing up to 100%).

There are two possible ways to account for the relative importance of traffic volume and hence to determine or estimate *TR(i)*:

(1) National data on traffic volume (vehicle kilometres driven) by type of vehicle and type of road and time period. In the ideal situation national traffic volume data is available for all considered crossed strata but possibly this information has to come from combing different

[4] In questionnaire surveys age and sex are sampling strata - so there it makes sense to weight according to population statistics. But this is not the case in roadside surveys.

sources. It is also possible that no data is available for specific strata (e.g., no indication of national traffic volume according to the considered time periods).
Information on traffic volume can come from different sources such as national counts on roads for proportions on the road types. Representative online mobility survey data may be available for the relative proportions according to time period. If traffic volume data are available for each road type and information is available or can be estimated for the distribution of traffic volume over the time periods (e.g. 10 % of traffic at night, 20 % of traffic in the weekend), these proportions should be combined in a logical way to calculate a percentage of the traffic volume for each crossed stratum, all summing up to 100%.

(2) If no traffic volume information is available but a reliable estimate of the length of the roads of each road type is available, one could alternatively use the traffic counts from the sessions in the stratum to make an estimate of the hourly number of vehicles at the survey locations (= Nh(k)). If the locations are randomly selected, this average (time-standardized) vehicle count is an estimate of the average hourly vehicle count of all locations in the stratum. This value, multiplied by the estimate of the length of the roads in the stratum – and, if different time periods are considered, the number of hours in the time period considered – should give some estimate of the traffic volume in the stratum. These values could then be used to weight strata.

Let us develop this second approach which is based on road length:

- if there are *M* strata to be aggregated
- let *Ns(i)* be the average number of vehicles per hour (or any other duration standard) for stratum *i* (*i* ranging from 1 to *M*)
- let *Ps(i)* be the relative proportion of the time periods considered (e.g., 5/7 for weekdays, 2/7 for weekend days)
- let *RL(i)* be the total road length of stratum *i*
- *KPI(i)* be the KPI value of stratum *i*

Then:

$$Aggregated\ KPI\ Value = \sum_{i=1}^{M} \frac{Ns(i) * Ps\ (i) * RL(i) * KPI(i)}{\sum_{1}^{M} Ns(i) * Ps\ (i) * RL(i)} \qquad (4)$$

Note that $Ns(i)$ is the average number of passing vehicles per hour on the road type (e.g., urban roads) and within the time period (e.g., weekdays) the stratum (i) represents. $Ns(i)$ is equal to the mean of all $N_h(k)$ in the stratum *i*.

As an example, consider the following data for six different strata:

Table 3. Example of data for different strata

i	Road type	Time period	Road length (km)	Ns(i)	PS(i)	KPI(i)
1	Urban	Weekday	10 000	100	5/7	87%
2	Urban	Weekend	10 000	80	2/7	92%
3	Rural	Weekday	25 000	50	5/7	82%
4	Rural	Weekend	25 000	30	2/7	79%
5	Motorway	Weekday	3 000	600	5/7	78%
6	Motorway	Weekend	3 000	350	2/7	74%

Application of formula (4) will then yield an aggregated KPI value of 81.4%.

In order to get an idea of how realistic this approach is (this analysis may lead to rejecting this approach rather than accepting it) it can be bootstrapped. The $Aggregated\ KPI\ Value$ value above depends on the average number of vehicles per hour value $Ns(i)$ which is calculated for each stratum. For each stratum $Ns(i)$ is calculated from the $N_h(k)$values obtained from the survey sessions. The purpose of this bootstrapping approach is to see what values the $Aggregated\ KPI\ Value$ could have attained if the $Ns(i)$ values were consistent with the $N_h(k)$ values, but reasonably different.

A way to do this is for each $Ns(i)$ collect the $N_h(k)$ for k = 1, …, K. The "bootstrap" way would be of selecting L values (with L<K) from $N_h(k)$, k = 1, …, *K* and calculate a new value for $Ns(i)$. Do this for each stratum *i* and equation (4) can be applied to obtain a new value of $Aggregated\ KPI\ Value$. When applying this step quite a number of times (with replacing the L values), one gets an idea of how well determined the $Aggregated\ KPI\ Value$ is.

The idea behind this approach is that both $RL(i)$ and $KPI(i)$ are quite accurately known compared to $Ns(i)$ within each stratum. Obviously, $RL(i)$ is constant within the stratum and we ***assume*** $KPI(i)$ is reasonably similar within the stratum (e.g., on motorways at night, you have this percentage of seatbelt use). Assuming this assumption holds, and we took another sample, we would have identical $RL(i)$ and quite similar $KPI(i)$ but only different $N_h(k)$, k = 1, …, *K*. The best guess for the values the $N_h(k)$, *k* = 1, …, *K* are the *K* values that were counted. Therefore, we sample with replacement *K* values from that set to get an estimate of $Ns(i)$. If the range of values for $Aggregated\ KPI\ Value$ obtained this way is too large to be useful (e.g., varying with more than 5%), the whole approach is probably not accurate enough. Unfortunately, if the range is too large to be useful, we still have the assumption that the $KPI(i)$ are reasonably similar within each stratum. This may not hold, so we cannot conclude, but we might tentatively assume the approach is not too bad.

Trendline beneficiaries should also report in their metadata whether bootstrapping has been applied.

Reporting:

When reporting results to the project coordinator, Trendline beneficiaries need to report, for each stratum used in the analysis, an estimate of the traffic volume (or at least percentual share of it), since this is a key element in assuring respect for minimal requirements for weighting and to assure internationally comparable results.

Important: if no vehicle counts or no road length information is available, or no otherwise obtained (actual or estimated) traffic volume information, one should only treat the strata separately, and defer from aggregation. In such cases, some of minimum required KPIs in Trendline cannot be delivered.

F. Calculation of confidence intervals (CI)

Calculation of confidence intervals for the data described above is far from trivial. The statistical reference works considered do not precisely cover the sampling problem considered and the methods discussed that appear to be feasible for implementation. Some Trendline beneficiaries appear to use gaussian approximations to statistics to aggregate over sample sessions within strata and aggregate over strata, although there are also some who are using statistical software taking the complex sampling

design into account. In general, using gaussian approximations in the aggregation process is acceptable for the averages and percentages themselves but may cause serious problems determining confidence intervals thereof.

Weighting factors for observations within a stratum are given in formula (1) and weighting approaches for aggregation of different strata in formulas (3) and (4).

Trendline beneficiaries should use a method for calculating Confidence Intervals that takes the sampling design method into account, in particular the fact that observations are nested in sessions. Trendline beneficiaries need to indicate in the metadata how they calculated the CIs. Since approximations that assume simple random sampling clearly lead to unrealistically small confidence intervals, approximations using simple random sampling are not acceptable.

G. Using appropriate statistical software

It is advised to use dedicated survey software, as readily available in R and other software packages. Table 1 introduced above and all other variables needed for the weighting will serve as input to these procedures.

Packages that can be considered are:

- R Survey Package https://cran.r-project.org/web/packages/survey/index.html
- STATA Analysis of Complex Survey Data in Stata e.g. https://www.stata.com/meeting/mexico10/mex10sug_canette.pdf
- SPSS: https://www.ibm.com/products/spss-statistics/complex-samples
- SAS: https://support.sas.com/documentation/cdl/en/statug/63033/PDF/default/statug.pdf (hefty document including documentation of proc survey means)

Books considered:

Cochran, W. G. (1977). Sampling Techniques. Wiley

Thompson, S. K. (2012). Sampling. Wiley

Wu, C., Thompson, M. E. (2020). Sampling Theory and Practice. Springer International Publishing