KPI Distraction

Methodological guidelines

Version 1.4, 21 November, 2023

Co-funded by the European Union

KPI Distraction. Methodological guidelines

Please refer to this report as follows:
Stelling, A., Boets, S., González Hernández, B., Jankowska, D., Larsson, P., Schumacher, M., Vieira, S., Ziakopoulos, A., (2023). KPI Distraction. 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:**	1.4 Public 21/11/2023

Report Author(s):

Stelling, A., (SWOV Institute for Road Safety Research), the Netherlands
Boets, S., (VIAS Institute), Belgium
González Hernández, B., (CTL), Italy
Jankowska, D., (ITS), Poland
Larsson, P., (Trafikverket), Sweden
Schumacher, M., (BASt), Germany
Vieira, S., (LNEC), Portuga)
Ziakopoulos, A., (NTUA), Greece

Revision history

Date	Version	Description
10/03/2023	1.0	First draft version
02/06/2023	1.1	Final draft
09/06/2023	1.2	Version approved by the KEG Distraction
05/09/2023	1.3	Version after reviews by coordination team
21/11/2023	1.4	Version published at the website

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.

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

Passenger car
Motor vehicle with 3 or 4 wheels, mainly used to transport people, seating for no more than 8 occupants. Motor vehicles with these characteristics used as taxis as well as motor caravans are also included (CARE, 2021).

Heavy goods vehicle
Goods vehicle over 3.5t mgw. Larger motor vehicle used only for the transport of goods (CARE, 2021).

Light goods vehicle
Goods vehicle under 3.5t mgw. Smaller motor vehicle used only for the transport of goods (CARE, 2021).

Bus
Passenger-carrying vehicle, most commonly used for public transport, having more than 16 seats for passengers (CARE, 2021).

Coach
Passenger-carrying vehicle, having more than 16 seats for passengers. Most commonly used for interurban movements and touristic trips. To differentiate from other types of bus, a coach has a luggage hold separate from the passenger cabin (CARE, 2021).

Urban road
Public road inside urban boundary signs.

Rural road
Public road outside urban boundary signs, excluding motorways and expressways.

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.

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.

Week – daytime
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 imple-mentation 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 background of this document

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

Percentage of drivers not using a handheld mobile device

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 (European Commission, 2019), further referred to as "SWD" (see Appendix 1). Most of those minimal requirements are incorporated in this guideline document. The requirements are quantified and specified for each of the parameters. This document is based on a review of the methodological guidelines that were developed within the Baseline project (Boets et al., 2021), expert consultations within the Trendline Key Expert Group Distraction and the analysis of FERSI guidelines (Vollrath et et al., 2019); see also Appendix 3; SafetyNet Manual (Hakkert & Gitelman, 2007) and Baseline report on the KPI Distraction (Boets, 2023). In addition to the specification of the **minimum requirements** (always marked bold) to deliver the main KPI and the disaggregated indicators, this document also includes recommendations for optional additional activities. Member States can decide whether to follow the minimum requirements only or to extend (part of) their methodology, depending on available means and their own research questions.

2 Scope

2.1 General principles

SWD allows "***direct observation** by trained observers on the roadside or from moving vehicles. Other alternatives could be used if available, e.g. automatic detection. To be decided by Member States."* The main method proposed is observational roadside studies, in which all (relevant) drivers or a random selection of (the relevant) drivers are observed. The use of a handheld device is directly observed and coded by trained observers, possibly together with some optional supplementary basic information about the driver (e.g. age, gender).

The objective of the roadside observation study is to estimate the percentage of drivers NOT using a handheld mobile device while driving. The theoretical population refers to the total of all journeys (or at least from the vehicle types targeted) over the national territory. In other words, this reflects the total number of kilometres driven. Hence, the percentage of drivers NOT using a handheld mobile device refers to the percentage of kilometres driven without using a handheld mobile device.

The basic aim is for all participating Member States to have comparable indicators for the minimum required stratifications. Optional disaggregated indicators will only be compared for countries that are able to deliver those.

Self-report methods (e.g. roadside interviews or self-report surveys) are outside the scope of this document, but will be discussed in the context of guidelines for using indicators based on self-report methods.

2.2 Type of distraction to be observed

The KPI uses the term "handheld mobile device use". The use of 'device' instead of 'phone' makes this KPI futureproof. A mobile device can be defined as "a computer small enough to hold and operate in the hand" (e.g. https://en.wikipedia.org/wiki/Mobile_device), such as: mobile phones (e.g. smartphones), mobile computers (e.g. tablets), personal navigation devices, digital cameras.

Most Member States have a ban on mobile phone use while driving, while in some countries this has meanwhile been extended to mobile electronic 'devices'. Participating Member States are expected to provide metadata on the applied regulations and procedures related to this.

As an absolute minimum, two clearly visible distraction categories, excluding each other, should be recorded in each observation:

- **Having a mobile device in the hand (driver is holding a mobile device in the hand, which can be held at the ear, at the steering wheel or anywhere else)**
- **Not having a mobile device in the hand (rest category).**

Although the KPI refers to 'use of a handheld mobile device', this categorization is based on what is visibly detectable during an on-road observation study. This allows a clear and uniform observation procedure, even though handheld mobile device use will be underestimated because drivers often hide their mobile device under the dashboard or on their laps.

Optionally, as a function of their own research questions, Member States can decide to collect additional information on different basic tasks related to using a mobile device in the hand (e.g. phoning or texting), and/or to distinguish mobile phones from other mobile electronic devices. This latter distinction can be especially interesting for Member States with legislation which so far refers to mobile phones only.

The following categories are based on FERSI (Vollrath et al., 2019) and can be used:

- Having a mobile phone in the hand:
  - Handheld phoning: the driver is visibly holding a mobile phone in the hand and is pressing it at his/her ear or is holding it in front of the mouth. He/she is either talking or listening.
  - Texting/keying numbers handheld (mobile phone): the driver is visibly holding a mobile phone in the hand and is operating it.
  - Handheld reading/watching without operating (mobile phone): the driver is visibly holding a mobile phone in the hand and is looking at the phone without operating or handling it.
- Having another mobile device in the hand:
  - Operating another mobile electronic device in the hand: the driver is operating an electronic device other than a mobile phone (e.g., tablet, navigation system) and is holding this device in the hand.
  - No mobile phone or device in the hand (rest category).

Optionally, even more distraction categories could be collected, such as operating in-vehicle systems (see e.g., recommended categories by FERSI – Vollrath et al., 2019). When defining more (differentiated) distraction categories, it should always remain possible to derive the minimum distraction category for the KPI (handheld mobile device use vs. NO handheld mobile device use) from the data.

2.3 Vehicle types to be included

SWD requires the inclusion of "Cars, light goods vehicles, and buses/coaches as a minimum. Other user types if possible (disaggregated by user type)". However, in Baseline not sufficient sample size could be reached for buses/coaches (see Boets et al., 2022). Therefore, in Trendline this category has been replaced by heavy goods vehicles - HGVs for which sufficient sample sizes are expected to be feasible (based on the experience with the countries that collected data for HGVs in Baseline).

The target groups to include at a minimum are (see *Terms and definitions*, page v):

- **passenger cars**
- **light goods vehicles** (LGV; often from companies)
- **heavy goods vehicles** (HGV).

The data collection should include a variable "vehicle type" with these three categories.

The minimum requirement is to provide aggregated results for these three different vehicle types 'together'. This means that the data from the three vehicle types can be combined to provide the KPI. No separate KPI per vehicle type is required. The further specified minimum sample sizes consider the three vehicle types 'together'. It should be noted that replacing buses/coaches with HGVs makes the aggregated KPI for Distraction not fully comparable with the one obtained in Baseline.

Although providing disaggregated results is not requested, it is recommended to also provide differentiated results by vehicle type if the respective sample sizes are large enough to allow this (see Section 3.2).

The different vehicle types and their specific categorization should be clearly defined and illustrated for the observers (training) and in the methodological report: for example, some cars and vans share the same brand/model like Renault Kangoo (a passenger car is a vehicle with backseat windows and passenger seats; a van has no backseat windows and no rear passenger seats).

Bus/coaches (including mini-buses and public transport buses, see also *Terms and definitions*, page v) may be considered as an optional vehicle type category for the countries which expect that a sufficient number of observations can be reached. Of course, including this extra vehicle type should be feasible (see Section 3) and a sufficiently large sample for this extra vehicle type should be reached to provide sufficiently accurate separate results (min. 2,000; see section 3.2).

2.4 Driver characteristics (optional)

Member States with an interest in additional information on risk factors or predictors of distraction while driving, are encouraged to optionally record some easily observable extra variables such as:

- gender of the driver
- estimated driver age category (e.g. Vollrath et al. (2019) FERSI recommendation: young (18-24 years), medium (25 to 65 years), older (> 65 years))
- private vs. professional vehicle or driver (e.g. taxi)
- presence of passengers (yes/no).

Such additional variables can provide valuable input for evidence-based and risk group-oriented countermeasures (e.g. education and awareness building activities such as campaigns).

3 Measurement procedure

3.1 Sampling individuals

Sampling of drivers (of the relevant vehicle categories) **should be random**. Target drivers should always be randomly selected from all the possible drivers at the location where the observation is made. The easiest way to guarantee random sampling is that after finalisation of the coding of one observation, the first next passing target driver (on the specified road lane and direction) should be observed.

Most of the observed drivers will be car drivers as this is the most frequent vehicle type in motorized traffic. **While there are generally fewer light goods vehicles and heavy goods vehicles, the observer should give no specific priority to them in the measurement**. Only if the first next passing vehicle in the observation lane is a LGV or a HGV this driver should be coded.

Observations should be made in flowing traffic only, so of drivers while driving, since distraction behaviour is different when stationary, e.g. waiting at traffic lights. **No observation should be made of stationary drivers** (see also Section 3.4).

3.2 Minimum total sample size

Defining a minimum required sample size is by definition arbitrary since it depends on the level of accuracy that is considered adequate. Assuming an overall prevalence percentage of 5% to 10 % for handheld mobile device use while driving, accuracy in the order of 5% ±1 to 10% ±1.3 for this KPI can be considered acceptable (see Table 1).

$$\text{CI} = \text{prevalence} \pm z * \sqrt{\frac{prevalence\,(100 - prevalence)}{n}}$$

Prevalence	Lower bound CI, n=2000	Upper bound CI, n=2000	Lower bound CI, n=500	Upper bound CI, n=500
5%	4.04%	5.96%	3.09 %	6.91%
10%	8.69%	11.31%	7.37 %	12.63%

Table 1: Assuming simple random sampling and depending on prevalence levels between 5% and 10% for handheld mobile device use, the 95% confidence intervals (CI) for n=2,000 and n=500 are estimated using the formula above (z value 1.960 for 95% CI): upper and lower bound of the CI for the point estimates.

A sample size of about 2,000 observations should therefore be sufficient to provide frequency estimations (percentages) of the order of 1-1.3% with a 95% confidence interval. **Thus, as an absolute minimum 2,000 observations overall** (for the three minimally required vehicle types together) **is**

required. This minimum refers to valid datapoints in the study dataset in order to be considered for the national KPIs. No minimum sample size for the different vehicle types is defined because the minimum requested KPI is the aggregated result for the three types. However, given the high number of observations that have been reached in Baseline (see Boets et al., 2022) countries willing and able to collect more data than the minimum of 2,000 in total are encouraged to collect data of 2,000 drivers per vehicle type (i.e. 2,000 for passenger cars, 2,000 for LGVs and 2,000 for HGVs).

Member States aiming at having higher accuracy can calculate the required sample size to gather results with a specified accuracy level and confidence interval, using this formula: (FERSI - Vollrath et al., 2019):

When planning the study, the sample size required for the respective purpose should be considered. For example, to estimate the prevalence of texting on the mobile phone with a precision of 1% (width of 95% confidence interval) and it is assumed that the percentage lies around 5%, one can use the following formula to determine the required number of observations.

$$N_{required} = \frac{1.96^2 * \left(P * (100 - P)\right)}{Precision^2}$$

When using a P = 5% and a Precision of 1%, this gives:

$$N_{required} = \frac{1.96^2 * (5 * (100 - 5))}{1^2} = 1825$$

Accuracy for different subgroups or stratifications, such as the three road types, will by definition be lower. If higher accuracy levels are expected for particular strata (road type, regions) , it is strongly recommended to increase the total sample size. Ideally, a multiple of the minimum sample size can be obtained, which increases the accuracy of the estimates, and optionally can allow delivery of reliable estimates for separate categories of vehicle types or for further (crossed) stratifications (e.g. per road type x time period, per region).

Appendix 2 gives an overview of the argumentation behind the minimum driver sample. If, optionally, Member States aim at having disaggregated results by vehicle type, then the minimum sample size of 2,000 drivers should be applied to each vehicle type.

If similar accuracy levels are expected for particular stratifications/subgroups, it is strongly recommended to increase the total sample size. Member States optionally willing to have reliable KPI estimates for different possible combinations of stratifications (e.g. road type x time period; region x road type; region x road type x time period) should have a design with minimum 500 observations for the different relevant crossed strata (e.g. 3 regions x 3 road types x 3 time periods = 27 strata x 500 observations = needed sample of 13,500 drivers).

3.3 Sample size per road type

On-road observation studies should provide a representative sample of all traffic in the considered study area. For distraction **the minimum stratification to take into account is road type**. This covers three main road types (see also *Terms and definitions*, page v):

- **motorways** (public road with dual carriageways and at least two lanes each way). If expressways are included in the data collection, the results for expressways and motorways should be merged under the category *motorways*.
- **rural roads** (public roads outside urban boundary signs, excluding motorways),
- **urban roads** (public road inside urban boundary signs, assuming exclusion of motorways).

This is the minimum required categorization.

If Member States historically use a different road categorization, an attempt should be made to infer the minimum required road types. The road types considered and any deviation from the minimum requirements should be explained in the methodology (general characteristics like traffic signs to define inside/outside built-up area, possible speed regimes and number of lanes…).

In order to ensure a minimum number of observations for each road type, even if this would imply disproportionate sampling, at least 500 observations for each category of road type are required, thus:

- **minimum 500 drivers on urban roads**
- **minimum 500 drivers on rural roads**
- **minimum 500 drivers on motorways** (this requirement does not apply to Member States with no motorways or where the network of motorways is very limited).

It should be noted that this leads to bigger error margins for the point estimate for each of these roads. Given an overall prevalence of distraction of 5% to 10% this would give the following 95% confidence intervals for this level of aggregation: 5% ±1.9 to 10% ±2.6 (see Table 1).

3.4 Sampling and selection of locations

The **selection of locations should be as random as possible**, covering the geographical area of the country. There are different options for random location selections, such as simple random and stratified random (e.g. random sampling in different regions). The basic process for the random selection 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 can be made 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 automatically selects location points within defined areas randomly (e.g. https://desktop.arcgis.com/en/arcmap/latest/extensions/geostatistical-analyst/an-introduction-to-sampling- monitoring-networks.htm). If Member States have no GIS software, step 2 can also be done manually using a national geographic map, for example. OpenStreetMap/Google Maps/Google Earth.

Since a random selection of locations will also include low volume roads, it is expected that several low volume locations will be available for each stratum. If, however, traffic flow is too low, it is also acceptable not to include them. It is acceptable not to include locations with less than 10 relevant vehicles passing per hour.

Bigger countries may consider in a prior stage the selection of one or more regions/states which are considered to be representative of the country with regard to distracted driving. This can add to the fieldwork feasibility. If this is done, it should be explained in the meta-data.

Pragmatical considerations related to the locations should be taken into account: the observers should have a good view of the passing traffic while also ensuring that the observations can be conducted safely and inconspicuously (see also Sections 3.6.2 and 3.6.3).

Furthermore, typical criteria for observation studies on distraction are related to the free flow of the traffic, so ideally no locations should be chosen in front of traffic lights. Observation can take place near intersections but **only drivers who are driving should be observed, not drivers who are stationary**.

It is recommended to sample locations for the three road types proportionally to traffic volume on the road types (or proportionally to the kilometres driven on each road type in a country), assuming that each of the three road types represent a share of traffic volume above 20%, with this based on available national data (e.g., traffic/mobility data by road type from national traffic surveys). If traffic volume data is not available, or if the traffic volume share of a road type is less than 20%, then an **absolute minimum of 10 different locations per road type** should be selected in order to ensure representative results for the entire road network (see Appendix 2 for the argumentation behind the minimum location sample of 10 locations per road type):

- **Minimum 10 locations on urban roads**
- **Minimum 10 locations on rural roads**
- **Minimum 10 locations (or sections) on motorways**.

Taking into account the other criteria (Sections 3.2 and 3.3), this comes down to a mean minimum of 67 observations per location, if 30 locations are chosen. It is allowed to re-use the same sampling location for different times of day or days of week (different sessions).

When, optionally, stratification according to time period is used too, a minimum of 2 different locations for each combination of strata should be used (e.g. 3 road types x 3 time periods = 9 crossed strata). For more information on random sampling of locations and for determination of the minimal sample size, reference can be made to the SafetyNet general recommendations for SPI (safety performance indicators):
http://www.dacota project.eu/Links/erso/safetynet/fixed/WP3/sn_wp3_d3p8_spi_manual.pdf

To summarize, the minimum required sample sizes to provide the KPIs are:

- **Minimum 2,000 observations in total (aggregated vehicle types)**
- **Minimum 500 observations per road type (3)**
- **Minimum 10 locations per road type (3) = min. 30 locations in total**.

3.5 Optional further stratifications

3.5.1 Stratification by time period

SWD only requires the observations "during daylight" and no differentiation regarding week-weekend is requested. The **minimum requirement is to plan the observation sessions at mixed time intervals during daylight hours on normal working days. The mix of possible moments should be balanced over the three road types (i.e. to have a similar variation of considered day hours for the three road types)**.

Optionally, time period can also be considered as an additional stratification for Member States willing to have results for different relevant time periods, i.e. week day versus weekend day.

Additionally, even more optional time periods can be delivered following the FERSI recommendation of using three time periods (weekday peak, weekday off-peak and weekend day) can be considered, cf. FERSI (Vollrath et al., 2019):

> *"It is recommended that observations cover the whole daytime and different working days. This can be achieved by doing observations at least at two time intervals: peak hours (commuters) and off-peak hours and from Mondays to Saturdays to be able to differentiate between week-weekend. This allows to work with three time intervals: week-peak (e.g. 7-9, 16-18), week-off-peak (e.g. 10-15), weekend (e.g. 10-18).*
>
> *If different time intervals are selected, these should be randomly allocated to the different (stratified) locations within each location type selection (either one location is assigned a specific time interval, or different (time interval) sessions are organized at one location). It should be checked that the distribution of road types and time intervals is proportional to traffic volumes x time intervals OR that it is balanced with a minimal number of sessions in each combination for proper data analysis (and application of weights afterwards).*
>
> *Observation sessions within a specific time interval should start and end within this time interval."*

For Trendline the following definitions of week vs weekend daytime, should be used (adopted from the ERSO project, European Commission, 2022):

- Week – daytime: 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.

If stratification according to different time periods is also aimed for, then the minimum of 500 observations and 10 locations should be used per time period also. To ensure a balanced sampling for road types and time periods, a minimum of 2 locations for each combination of road type and time period should be used.

3.5.2 Stratification by region

Disaggregation by region is not a requirement. Member States are free to choose supplementary stratifications according to country regions (e.g. NUTS 1 regions). In that case countries can consider collecting data from each region or from a representative selection of regions.

Member States aiming at having meaningful KPIs at the regional level, including road type differentiation per region, will need a multiplication of the required minimum location sample and driver sample. The minimum location and driver sample requirements are then required for each region surveyed (see Sections 3.2, 3.3 and 3.4).

3.6 Practical organisation of the observations

3.6.1 Fieldwork set-up and procedure

A uniform fieldwork procedure should be chosen. Member States can estimate how many sessions and observation hours will be needed in order to reach the required or aimed at driver sample size, taking also the minimum location requirements into account. **One observation session should last at least 30 minutes**. Ideally and for practical reasons, however, 1 hour or longer (e.g. up to 3h) sessions are recommended. Furthermore, different sessions can be spread over mixed hours (or, optionally, over different time periods) at one location (e.g. spreading and balancing time per location) or each location can be used for one session (i.e. balancing time over locations within road types; this is the minimum requirement). **When planning the fieldwork sessions, one should ensure a balanced combination of the 3 road types and the time periods considered**, to avoid a systematic sampling bias (e.g. all motorway sessions in the morning and all urban sessions in the afternoon).

Prerequisites for carrying out observations are generally: good enough weather conditions (no heavy rain, no storm, no snow), good visibility (no darkness, no fog), good road conditions (no ice), flowing traffic (no accident or construction site).

Observation of drivers in HGVs may be more difficult than observation of car drivers due to their high seat position and windows, even as compared to bus/coach drivers which generally have more extended lower windows. For observing 'higher' positioned drivers, observers should have a high enough observation position or viewpoint. Suggestions are to use a safe and stable device to stand upon; taller observers will also have an advantage. When observations from a moving vehicle are used (e.g. on high speed roads, see Section 3.6.3) then ideally a vehicle with a higher seat position (e.g. a bu/coach) is used. **The observations should be made by well-trained observers along the road or from moving vehicles**. As indicated by FERSI (Vollrath et al., 2019),

> *"…this requires a thorough training of the observers, ideally both theoretically (e.g. a briefing explaining aim, variables and definitions, coding tool, complete procedure) and practically (e.g. exercises on the road with a trainer), and ideally also including a performance test to ensure a high inter-rater reliability between the observers. This is ideally checked from time to time during the fieldwork in order to ensure a high data quality.*
>
> *Regarding the number of observers for one observation session, one well-trained observer can be used. This has the advantage of being unobtrusive and efficient. At*

very busy sections it may be advisable to have two observers, e.g. one doing the observation and telling the results to the second observer, who is recording them. However, when using a limited number of variables even single observers are well able to observe and record at the same time."

For the on-site coding, paper sheets or tablet/ computers/ smartphones can be used. Using a tablet or a smartphone can have some advantages (e.g. direct coding, real-time central data collection, automatic coding of meta-data like the exact location, date and time of each coding, which also could serve for quality assessment), but the tool should be tested beforehand (user friendliness, speed, correction possibilities…) and be evaluated to be better than paper coding. For Trendline a dedicated software for observation measurements, including distraction, called 'SPIN' and developed by CDV, can be used free of charge.

Screenshots of the 'SPIN' software.

3.6.2 Observations at urban and rural roads

Stationary and moving observations on low and high speed roads should always be carried out in accordance with the applicable (road) safety regulations. Observations on urban and rural roads can be made from a **safe place along the road**. It is recalled that observation can take place near intersections but **only drivers who are driving should be observed**, not drivers who are stationary. If the traffic flow is disturbed at a selected location (e.g. due to works or a crash), then the observer should choose a new location on the same lane or nearby (within the same road category). Furthermore, more complex traffic situations requiring the full driver's attention are also best avoided.

3.6.3 Observations on motorways

Observations on motorways (or high speed roads) are possible from locations along the motorway that are easily reachable for observers (e.g. on rest/parking areas) and where observers can stand behind a safety barrier to observe oncoming and passing vehicles on the motorway lanes. It is important that these locations allow observation of traffic travelling undisturbed (not therefore locations where drivers have to stop or pay special attention to circumstances). This observation location should at least be usable for observation of vehicles on the lane closest to the observer (right lane) and for vehicles driving generally slower (e.g. HGVs). Observing vehicles on the lanes further away or vehicles at high speed may be more difficult.

A complementary or alternative method on motorways is to make observations from a moving vehicle in real traffic, with a driver and an observer on the backseat, which allows observing overtaking and overtaken vehicles on different lanes and also observing vehicles riding at different (also high) speeds (e.g. Riguelle & Roynard, 2014). Using this method, the geographical location is rather determined as a section (from location x to location y) than as one specific location on a certain motorway. These sections should reflect as far as possible the required min. 30 minutes duration of driving/observation. The lanes and speeds of the observation vehicle should be varied in a systematic way in order to carry out the observations in a representative way (e.g. 15 min. driving on the right lane at 90km/h and observing overtaking vehicles on the middle lane, then 15 min. driving on the middle lane at 120 km/h and observing overtaking and overtaken vehicles) within one observation session. In order to carry out the required traffic counts, the observer can stop at a safe location along the motorway section (e.g. behind a barrier overlooking the motorway at a rest/parking area).

The method of observation from moving vehicles (e.g. on the middle and left lanes) can be combined with stationary observation of vehicles on the right lane. Ideally, vehicles with a higher seat position (e.g. buses or coaches) should be used for a better view of the drivers, especially when observing HGVs. If this is not possible, observations from overpasses or bridges can also be considered, as long as these are not too high and provide a good viewpoint on the lanes; but a possible drawback is that observers in that position may be more noticeable by drivers which makes inconspicuous observation more difficult.

Camera observation may also be considered for safety reasons on higher speed roads, even though this method also presents some disadvantages (see Section 5).

3.6.4 Counting of traffic volumes

Traffic volumes should be counted during each observation session, even when national traffic volume statistics are available. This information is needed for the calculation of the percentage of drivers not holding a mobile device for each observation session and for correct calculation of the confidence intervals (weighting).

Counting of traffic during a session is ideally done by counting all (including the observed) passing relevant vehicles (i.e. the types that are considered in the study; this can be combined for the three vehicle types but if separate KPIs per vehicle type are aimed at (optional), this should be done also separately per vehicle type), in the same lane(s) and in the same direction as the observation. In the ideal situation where each passing relevant vehicle can be observed in a session, the total number of observed vehicles corresponds to the total session count.

Minimum manual traffic counts are made by counting all the passing relevant vehicles in the same lane(s) and in the same direction as the observation, during a 10 minute break in the middle of the session, or 5 minutes before and 5 minutes after the session. This break is in addition to the minimum 30 minutes (ideally for practicality min. 1h) observation session. If disaggregated results for different vehicle types are aimed for (optionally) then the vehicle types should be counted separately. Additional counting can also be done with an automatic counter during the whole session (e.g. loop on the road) so as to have an indication of the general traffic volume (optional).

3.6.5 Time of the year

SWD does not set a specification for time of the year (months). **Holiday periods (bank and school holidays) and hard winter conditions should however be avoided**, as these disturb normal traffic patterns. All months are allowed except for December-February to avoid a higher risk of (very) adverse weather conditions which may influence driver behaviour and can complicate the observational work (e.g., due to the weather conditions and shorter daylight periods), as well as July-August (in some Member States June too) to avoid typical holiday periods in the interests of representativeness. For the other months, sessions during official holidays should therefore also be avoided.

When Member States have historical series of measurements it is recommended to use the same periods of the year as for the earlier measurements. Member States intending to organise more than one roadside observation study to deliver the KPIs (e.g., one in Spring and one in Autumn) need to apply the minimum sample size requirements to the combination of both measurements. The data from both measures can be combined to deliver the indicators.

Regarding the impact of the COVID-19 pandemic on timing, it is recalled that the observation studies are ideally done in as normal driving situations as possible. Studies should not take place when a country or region is in a severe lockdown, with e.g., restrictions on journeys, closure of schools, and/or closure of non-essential shops. When less or less severe restrictions apply and there is a sufficiently normalised traffic flow (e.g. 75% of the normal flow), observation studies for distraction can be conducted. A night curfew is less relevant for the distraction study as daylight measures only are required.

4 Data analyses

4.1 Data coding

Detailed specifications for the data delivery and data matrix for the Trendline dataset will be provided at a later stage. As a first guideline, it is suggested to include for each datapoint (i.e. each observation or each driver) in the dataset, the following variables:

- vehicle type (3)
- distraction: use or no use of a handheld mobile device (2)
- road type (3)
- date
- start hour
- end hour
- total observation duration
- unique location code (to know which observations belong to the same session)
- unique session code (only needed if the same location is used for different sessions)
- observation session duration
- traffic count duration
- traffic count total (at a minimum all relevant vehicle types together, ideally per considered type)

Variables such as road type, time period, location code, session code, day and time of a session, traffic counts can be coded once per session by the observers. These variables should then be added in the dataset to each datapoint (each observed driver) in the same observation session.

The following list gives some additional variables which can optionally be coded and included in the dataset:

- Coded per vehicle observation:
  - driver characteristics: age category, gender
  - presence of passengers
- Coded per observation session (once per session) and included in the dataset for each observation line from one session:
  - region
  - time period category (e.g. week off-peak, weekend, week peak)
  - code of the observer(s)
  - weather condition
  - road condition
  - flow of traffic
  - number of lanes
  - observation lane(s)
  - observation direction.

4.2 Post stratification weights and statistical analysis

Specifications on calculating weights and confidence intervals are provided in Appendix 4 *Suggested approach for weighting sample data and calculation of statistics*.

4.3 Expected results and data delivery

For each indicator defined below, a point estimate as well as a 95% confidence interval is expected. Results should also include the unweighted number of drivers the result is based on.

The main indicator is the percentage of drivers not using a handheld mobile device across all day times and road types (locations). When optional vehicle types are included in the observations (e.g. buses/coaches, motorcycles or bicycles), the main KPI should only include the three required vehicle types. **Furthermore, KPI values (point estimates and confidence intervals) are also required for each of the three road types**.

It is optional to also provide estimates for specific categories of road users and for additional stratifications, if sample sizes are sufficiently large:

- by vehicle type (cars; possibly also light goods vehicles, heavy goods vehicles)
- by time period (e.g. FERSI: week peak, week off-peak, weekend)
- by region (if applicable)
- by age group (e.g. FERSI: young (18-24 years), medium (25 to 65 years), older (> 65 years))
- by gender
- by private vs. professional vehicle or driver (e.g. taxi).

It is also recommended to provide estimates for combinations of these, if sample sizes allow this.

For the **data delivery** to the Trendline consortium (inclusion in the Trendline database), two possible levels of aggregation are possible (further instructions on dataset structure and variables will be provided later):

1. **Minimum level requirement: point estimates (%) for all categories of the minimum required levels of disaggregation (combination of the 3 vehicle types; road type (3)), and for any optional recommended additionally considered level of disaggregation, including confidence interval (CI) estimates**. The minimum output includes main effects of specific variables. Interactions are not mandatory.
2. Semi-aggregate level: crossed-level matrix of all considered levels of disaggregation (crossed point estimates) + Confidence Intervals

4.4 Metadata

Member States should deliver the data together with metadata including at least the following information:

- the study design, including the vehicle types considered
- the method used and rationale for choosing the locations (sampling method, inclusion/exclusion criteria, minimal traffic flow considered)
- the fieldwork procedure (planning of session/hours, method to record the observations, considered prerequisites for a session, days of the week and hours of the day, crossed designs, variables collected, months on which the observations took place)
- the statistical techniques used to weight the data, to calculate the CIs, and to analyse the results
- metadata on the applied regulations and procedures related to this KPI (e.g. legislation on mobile phone and/or device use).

5 Requirements for automatic detection via roadside cameras

SWD also allows other observation methods if available, e.g. automatic detection. Smart cameras could automatically detect whether drivers have a mobile phone or device in the hand. This technology seems promising and could have clear advantages as compared to using observers in terms of e.g., reliability, data collection duration, night time use. Possible drawbacks should however be evaluated (e.g., lacking variables). This is new technology on the market and should therefore have been tested and validated before use. For privacy considerations, faces should not be caught on camera.

Example pictures:

The experience with such smart cameras for detecting mobile device use, in enforcement and certainly for research purposes, is still very scarce. Stelling-Kończak et al. (2020) performed a study into various enforcement methods for mobile device detection including camera-based enforcement. Some insights and conclusions from their study are:

> *"Cameras can be fixed (unmanned; mostly installed for weeks or months) or mobile (manned, easily movable from location to location; e.g. placed device on the ground), as well as have different levels of intelligence or smartness:*
>
> - *not intelligent: camera makes images of all passing vehicles and these have to be manually checked*
> - *partly intelligent: camera makes images of drivers that presumably are using their mobile device (based on intelligent image recognition software) and these have to be manually confirmed*
> - *fully intelligent: camera fully automatically identifies drivers using a mobile device (based on intelligent image recognition software) without a need for a manual confirmation.*

Such cameras are still rarely used (for enforcement), and if used, responses indicated that slightly more often mobile cameras are used than fixed, and so far only not or party intelligent cameras. As yet, 'smart', partly automated cameras are only used in a few countries, among which Australia, Saudi Arabia and (on a small scale) the Netherlands. [...] The most important reasons mentioned for not using these cameras are: technical and legal barriers and for mobile cameras the high costs. Technical issues such as polarizer filter and infrared light for night and bad weather observations are often present. Viewing angle positions have to be changed in order to observe either lower vehicles (cars, vans) or higher vehicles (trucks). Not all cameras can be placed on all road types (motorways, urban and rural roads). Mostly they are placed at a height. The steeper the viewing angle, the deeper the view inside the vehicle can be.

In the Netherlands different legislations specify that police are allowed to use these cameras. Based on the first trials with their mobile camera, they conclude that improvements of the technology and legal interpretations are possible (image not always sufficiently clear, not always sufficiently visible if there is a device in the hand).

A general concern about the use of such cameras (mainly in the USA and Australia) is that they are a violation of privacy because an image is taken from the driver (and passengers). Generally, this violation of privacy is [or can be] minimised by erasing pictures without an offence immediately. In the Netherlands furthermore passengers are automatically detected and if so, that part of the image is automatically 'masked', so not visible during the manual check/confirmation.

Experiences [with partly automated cameras] are positive, but new technological developments are expected to offer more application possibilities. Thus, artificial intelligence will presumably make it easier to recognise offenders, reducing the time needed to manually check and confirm the images. ... The difficulty with ... camera- based enforcement is that drivers often try to hold their phones in such a way, for example on their laps or close to the car door, that they are hard to detect from the outside. ... For camera-based enforcement a good view inside passing cars is also important. To achieve this, cameras should be directed downwards at an angle that is as straight as possible. In addition, further improvements are possible in preventing light reflection from windscreens and in the ability to simply combine monitoring of car drivers on the one hand, and truck and bus drivers on the other hand."

Proposed further reading by the authors:
https://roadsafety.transport.nsw.gov.au/stayingsafe/mobilephones/technology.html

When smart cameras are used, in general, the same minimum requirements, expected results and data delivery as for roadside studies with observers apply (see Sections 2 to 4): with regard to vehicle types (inclusion of 3 vehicle types: cars, vans and HGVs), road types (motorway, urban and rural roads) and locations (as random as possible), time of the observation (mixed time intervals at daytime hours on weekdays), sampling (random) and sample sizes (min. 10 locations per road type, min. 2,000 drivers (combined for the 3 vehicle types) and min. 500 drivers per road type).

Member States aiming to use cameras should first evaluate the feasibility of these minimal requirements for delivering the KPI for distraction in a pilot study. Some issues are for instance:

- National (regional, local) regulations (admission, procedures...) which apply to using this method
- GDPR constraints which apply
- Reliability of the camera (false negatives, false positives)

- Technical feasibility (e.g. number of cameras available, restrictions concerning the choice of locations where a camera can be placed) and camera quality (resolution, a proper angle, dynamic range and lightning conditions, preventing light reflection)
- Arduousness of the method: labor, time and effort needed to reliably determine the use of a handheld mobile device
- Because data collection is not only required of drivers using a handheld device, minimally the number of all passing relevant vehicles during the observation should be counted. Ideally, data collection (images) includes both drivers with and without a handheld mobile device. This would allow a manual check, although time consuming, and may allow also coding additional variables, such as driver variables.
- Vehicle type determination by the image (car, van, HGV). The data collection should include these three vehicle types at a minimum; if other vehicle types are also included the type of vehicle should be coded, because disaggregated results are then needed.
- If cameras made for deployment on overpasses are used, this restricts the random location sampling procedure and may also complicate the inclusion of the three road types.

Member States aiming at using this technology should provide detailed information in the methodological report on the technical aspects of the camera, sampling procedures (locations and drivers; vehicle types included), camera accuracy (false positive/negative ratio), data-collection/coding procedures, data quality and correction procedures, data treatment, and data analysis including weighting procedures (see Appendix 6 for an example of a report delivered by Finland in Baseline). As for the roadside studies with observers, the results should be weighted according to traffic volumes by type of road (and other considered stratification variables). The dataset should minimally include datapoints for handheld mobile device users and non-handheld mobile device users, including the minimum measurement session variables in which the observations are nested (location code, road type, date, start and stop time, see Section 4.1).

References

Boets, S. et al. (2021). Methodological guidelines – KPI Distraction. Baseline project, Brussels: Vias institute

Boets, S. (2023). Baseline report on the KPI Distraction. Baseline project, Brussels: Vias institute

European Commission (2021) CARE DATABASE. CaDaS Common Accident Data Set. Version 3.8 – June 2021. https://road-safety.transport.ec.europa.eu/system/files/2021-07/cadas_glossary_v_3_8.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

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

Riguelle, F. & Roynard, M. (2014) Conduire sans les mains. Utilisation du GSM et d'autres objets pendant la conduite sur le réseau routier belge. Bruxelles, Belgique: Institut Belge pour la Sécurité Routière – Centre de connaissance Sécurité Routière. https://www.vias.be/publications/Rijden%20zonder%20handen/Conduire%20sans%20les%20mains%20-%20Utilisation%20du%20GSM%20et%20manipulation%20d'autres%20objets%20pendant%20la%20conduite%20s ur%20le%20r%C3%A9seau%20routier%20Belge.pdf

Riguelle, F. & Roynard, M. (2014) Rijden zonder handen Gebruik van de GSM en andere voorwerpen tijdens het rijden op het Belgische wegennet. Brussel, België: Belgisch Instituut voor de Verkeersveiligheid – Kenniscentrum Verkeersveiligheid. Retrieved from: https://www.vias.be/publications/Rijden%20zonder%20handen/Rijden%20zonder%20handen%20-%20Gebruik%20van%20de%20GSM%20en%20andere%20voorwerpen%20tijdens%20het%20rijden%20op%20het% 20Belgische%20wegennet.pdf

Riguelle, F. & Roynard, M. (2014) Summary. Driving without hands Use of mobile phone and other objects while driving on Belgian roads. Brussels, Belgium: Belgian Road Safety Institute – Knowledge Centre Road Safety. Routière. https://www.vias.be/publications/Rijden%20zonder%20handen/Driving%20without%20hands%20-%20Use%20of%20mobile%20phone%20and%20other%20objects%20while%20driving%20on%20Belgian%20roads. pdf

Stelling-Kończak, A., Goldenbeld, Ch. & van Schagen, I.N.L.G. (2020) Handhaving van het verbod op handheld telefoongebruik. Een kijkje in de keuken van Nederland en andere landen. [Enforcement of the ban on handheld phone use: a look behind the scenes in the Netherlands and abroad] Resume R-2020-23, SWOV Den Haag, 2020. Retrieved from: https://www.swov.nl/publicatie/handhaving-van-het-verbod-op-handheld- telefoongebruik

Vollrath, M., Schumacher, M., Boets, S. & Meesmann, U. (2019) Guidelines for assessing the prevalence of mobile phone use in traffic. FERSI technical paper. Retrieved from https://fersi.org/wp-content/uploads/2019/11/Guidelines-prevalence-mobile-phone-use.pdf

Appendix 1 SWD KPI 5 for driver distraction by handheld devices

Ref: Commission Staff Working Document - EU Road Safety Policy Framework 2021-2030 - Next steps towards "Vision Zero, SWD (2019) 238, https://ec.europa.eu/transport/sites/transport/files/legislation/swd20190283- roadsafety-vision-zero.pdf

Rationale

Driver distraction is considered as a collision factor of growing importance due to the increased use of mobile devices - mainly smartphones - during the past years, and the widespread use of texting applications has aggravated the existing problem of phone calls. This is why the use of a handheld mobile device while driving is proposed as a proxy to assess the driver distraction problem.

Definition of the KPI

Percentage of drivers NOT using a handheld mobile device.

Minimum methodological requirements

Data collection method	Direct observation by trained observers on roadside or from moving vehicles. Other alternatives could be used if available, e.g. automatic detection. To be decided by Member States.
Road type coverage	The indicator should cover motorways, rural non-motorway roads, and urban areas. The results may be presented separately for these three different road types.
Vehicle/user type	Cars, light goods vehicles, buses/coaches as a minimum. Other user types if possible (disaggregated by user type).
Location	Random sample (methodology for Member States to decide).
Time of day	Observations to take place during daylight.

Appendix 2 Rationale behind the minimum sample requirements

The methodological guidelines for all KPIs are designed to ensure international comparability between KPI values while taking into account 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. Giving priority to feasibility and affordability, as a rule of thumb the minimum total number of observations was set at 2,000, the minimum number of observations for different strata at 500. It was agreed that 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.

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. Once again, there is no statistical rationale for determining the required minimum number of locations to ensure representativeness of the observations for the entire country. This mainly depends on the amount of variance between locations and within a country. Giving priority to affordability, a rule of thumb was also used to define the minimum number of locations at 10 per stratum. In order to ensure representativeness for the entire country larger numbers of locations might be required for larger countries. Taking field work costs into account, it was however decided to only identify the minimum requirements and leave decisions on the final number of locations to the discretion of the member states. Equally importantly, in order to ensure representativeness of the measurement locations these should be randomly selected as far as possible.

The main objective in defining the minimum methodological requirements is to keep a balance between affordability of the field work and the requirements to make meaningful international and historical comparisons. Therefore, the emphasis is placed on the minimum requirements that can also be taken into account by smaller countries. It is however of interest to any member state to increase the accuracy of the KPI estimates by boosting the number of locations and the number of observations.

Appendix 3 Main FERSI recommendations

Ref: Vollrath, M., Schumacher, M., Boets, S., & Meesman, U. (2019) Guidelines for assessing the prevalence of mobile phone use in traffic. FERSI technical paper. Retrieved from https://fersi.org/wp-content/uploads/2019/11/Guidelines- prevalence-mobile-phone-use.pdf

5. Summary and overview of recommendations

The next three Tables present an overview of the most relevant recommendations related to the estimation of point prevalence of mobile phone use in traffic, distinguishing between BAsic (BA) recommendations for epidemiological studies (Table 2), recommendations for Roadside Observational Studies (ROS; Table 3), and recommendations for Self-Report Studies (SRS) - roadside interview, telephone interview, and online survey(Table 4).The list of recommendations is not aimed to be exhaustive.

Table 2: Summary of BAsic (BA) recommendations for epidemiological studies

BAsic (BA) recommendations for epidemiological studies	
BA1	In all epidemiological studies a thorough control and documentation of possible influencing context factors is needed (e.g., region, road type, traffic density, day of week, time of day, weather).
BA2	In all epidemiological studies a uniform definition of basic tasks should be used.
BA3	The main distraction categories should be assessed in all epidemiological studies. When reporting the results, the full definition of the categories should be given.
BA4	In each epidemiological study, a core set of subject characteristics should be included.

Table 3: Summary of recommendations for Roadside Observational Studies (ROS)

Recommendations for Roadside Observational Studies (ROS)	
ROS 1	To estimate the extent of the problem of mobile phone use in traffic, or to investigate the effect of a campaign or legal measure, observational studies deliver most direct and valid data, and are therefore the preferred methodological option.
ROS 2	To better understand the subjective background of mobile phone use in traffic or to describe the traffic participants' behaviours in more detail, self-report studies are more suitable.
ROS 3	To get a full understanding of the problem of mobile phone use, a combination of ROS and self-report studies is recommended, taking the advantages of each method while compensating for their shortcomings.
ROS 4	When planning an observational study a set of basic methodological aspects should be considered. These should also be documented in the report.
ROS 5	The report should include an exhaustive definition of all variables which are observed or are recorded.
ROS 6	Each observational study requires a thorough training of the observers including a performance test to ensure a high inter-rater reliability.
ROS 7	The rationale for choosing the locations of the observations should be provided.
ROS 8	The basic characteristics of the observation sites should be recorded and included in the report.
ROS 9	Observations should cover the whole daytime and different working days. This can be achieved by selecting at least three time-intervals per day and by doing observations at least from Mondays to Saturdays.
ROS 10	In order to ensure a high quality of the data, observations should only be done if some basic requirements are met (e.g., weather, light, road condition).
ROS 11	External validity of the results requires that the target group is clearly defined.
ROS 12	For car drivers, cyclists and pedestrians the observation study should always include private traffic participants (including those on their way to/from work) as they represent the largest part of traffic.
ROS 13	Target objects should always be randomly selected from all the possible objects at the location where the observation is done.
ROS 14	For the observers, there has to be a clear definition which of the traffic participants belongs to the target group with regard to their characteristics, but also their location.
ROS 15	It is recommended to use tablet computers for the observations.
ROS 16	At the beginning of each observation session, a key set of variables describing the location should be recorded.
ROS 17	When planning the study, the appropriate sample size should be estimated and used to determine the required number of observational sessions, taking the different types of traffic participants (car drivers, cyclists, pedestrians) into account.
ROS 18	When selecting the locations of the observations, it must be ensured that the observations can be conducted safely and inconspicuously.
ROS 19	Each data set (one observation) should include the information about the location and observation time as well as the individual information about the traffic object observed.
ROS 20	Data could be weighted according to traffic volumes at the different locations of observation.
ROS 21	Even more importantly, when the results aim to be representative, a weighing with regard to relevant characteristics of the traffic participants and their mobility should be done.

Appendix 4 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[1] 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[2]. 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.

[1] 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).

[2] 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[3]. 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:

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

[3] 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.

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 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

Appendix 5 Summary overview of on-road observation study requirements and recommendations

Trendline minimum requirements for on- road observation study	Trendline recommended options for on-road observation study
KPI: % not using a handheld mobile device - % no device in the hand + CI aggregated - % no device in the hand + CI per road type (3)	- Different types of distraction - Driver characteristics
- Direct observation by well-trained observers along the road or from moving vehicles; alternatively roadside cameras may be used for observations (preferably a pilot study first)	
- Three road types: rural roads, urban roads, motorways	
- Locations: selection as random as possible, good view, safe, inconspicuous	- Exclusion of locations with <10 vehicles/hour is allowed - Geographical coverage - Region stratification - A representative set of locations instead of randomly selected locations is allowed
- Three vehicle types: passenger cars, light goods vehicles, heavy goods vehicles	-
- Min. sample size: 2,000 observations for the 3 vehicle types together (it is allowed not to report disaggregate data for the 3 included vehicle types)	- Boost sample size for more accurate estimates and further (crossed) stratifications - Collect data of 2,000 drivers per vehicle type (i.e. 2,000 for passenger cars, 2,000 for LGVs and 2,000 for HGVs).
- Min. 500 observations/road type (3)	
- Min. 10 different locations/road type	
- 1 location = min. 1 observation session of min. 30 minutes	
- Fieldwork organisation: mix of daytime hours: on and off peak on week days, balanced over road types/locations	- Time period stratification: week days versus weekend day (min. 10 locations per time period; min. 2 locations per time period x road type; min. 500 observations/ time period) - Additional time period stratification: week day peak, week day off-peak, weekend day (same requirements as above per time period)
- Not during holidays or heavy winter period	
- Exclude observations of stopped vehicles, include all other	
- Traffic counts during sessions (10 min) for weighing data + estimates of road network length (3 types)	- Use available traffic volume data to sample locations and to weigh data according to included stratifications
	- Complete disaggregated data (crossed strata)

Appendix 6 Example of a methodological report for automatic detection via roadside cameras[4]

The measurements were made with a combination of two cameras: camera 1 detected the licence plates from passing vehicles. Camera 2, which was a higher-resolution device intended to photograph the driver, was triggered whenever camera 1 detected a licence plate in its field of view. A polarisation shield was installed to camera 2 to minimise the reflections from the windshield of the vehicle. The equipment was adjusted so that it would take one picture of each passing vehicle. The reason for automatising the data collection was to enable larger sample sizes to be obtained from the traffic stream with reasonable resources. In addition, the reliability of detecting phone use and other KPIs were assumed to be higher from a static picture than from a live observer at the side of the road. The camera used to recognise the license plates was assessed to function reliably also in situations when the licence plate is dirty or foreign. The equipment was battery-powered with the battery located on the ground next to the equipment with the cameras.

A sign including basic information on the measurements was attached near the equipment. The sign also included an internet link with access to the privacy statement. The pictures taken during the measurements will be deleted after the analysis is finalised.

Some examples of locations where our camera system was in use are presented in Figure 1, Figure 2 and Figure 3.

4. Adopted from Boets, S. (2023). Baseline report on the KPI Distraction. Baseline project, Brussels: Vias institute. Annex 6.

Figure 1. Example of location where the cameras were installed in a tree near the road.

Figure 2. Example 2 of location where cameras were installed in a pole with several other already existing equipment.

Figure 3. Example 3 of location where cameras were installed near road signs and in a pole with some other equipment.

The processing of pictures was done manually. However, an annotation tool was developed to speed up the otherwise labour-intensive process. The tool consisted of a graphic interface with keyboard shortcuts for the required classifications. The tool was created with Python and the open source QT5 software library. The identification of the vehicle type was done by the rater based on the pictures alone. The projection of the camera taking pictures of the cockpit was optimised to take pictures of passenger cars, hence a higher share of pictures of trucks and buses were unusable compared to passenger cars due to the glare from the windshield or interior being too dark.

The total sample included 18,259 pictures. From this sample 2,508 pictures were duplicates or otherwise erroneous and/or unusable (i.e. driver not clearly visible e.g. due to the glare, rain or too dark interior), and 2,056 pictures were too unclear to reliably identify the mobile phone usage. The remaining sample was 13,695. This final sample covered 70.3% of the passing vehicles (calculated over all measuring points (27/30) with nearby loop detector calculating the total traffic count).

One person analysed the whole dataset according to the pre-defined criteria. This analysis was used to calculate the weighted KPI values (reported to the coordinator by using the excel template). In addition, an interrater reliability check was done to a smaller sample of pictures. In total, 1,394 of pictures (slightly over 10 % of final sample) were analysed by two extra persons (three persons in total). The pictures for this additional check were selected randomly from the whole dataset so that all road types were equally covered.

This reliability check showed that identification of phone usage is challenging and not always indisputable. From the analysed sample, all raters agreed that 1.29% of drivers were using mobile phone while driving. The estimated mobile phone usage by rater varied between 1.94–2.87%. In total, 3.80% of drivers were estimated to use a mobile phone based on at least one rater.

We did not employ an "unclear" category for the annotation process, as the amount of images forced us to be economical in the number of classifications or "tags" for each image. In short, the vehicle types were easily seen from the images, whereas phone use had some ambiguity due to the (relatively) low resolution and dynamic range of the camera sensor. Seatbelt usage proved to be impossible for this particular reason – in some images the seatbelt could be seen, but it was impossible to confirm the absence of the seatbelt. For phone use it was easier to recognize that the driver was *not* holding a phone, with some ambiguity in the positive classification due to arm position etc. While we cannot provide a quantitative estimate for the amount of ambiguous classifications, overall we consider that amount to be low for the KPIs reported.

In general, we think that this method was rather suitable for this purpose. Analysis of pictures (1 picture per vehicle) requires significantly less resources to analyse the data compared to video footage. In addition, compared to the onsite observations, pictures allow assessments to be done by several persons and looking at the situation for longer than the time the vehicle is passing by the location. However, we agree that our method was not either perfect. In the future, some development work could be done, for example, to improve the quality of the pictures and to improve the coverage of the passing by traffic stream.