KPI Driving under the influence of alcohol

Methodological guidelines

Version 1.3, 21 November, 2023

Co-funded by the European Union

KPI Driving under the influence of alcohol. Methodological guidelines.

Please refer to this report as follows:
Boets, S., Folla, K., Houwing, S., Forsman, Å., Klipp, S., Areal, A., Jankowska-Karpa, D. & Meesmann, U. (2023). KPI Driving under the Influence of Alcohol. Methodological Guidelines. 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:**	1.3 Public 21/11/2023

Report Author(s):

Boets, S., Meesmann, U. (Vias institute), Belgium
Folla, K. (NTUA - National Technical University of Athens), Greece
Houwing, S. (CBR - Central Office of Driving Certification), the Netherlands
Forsman, Å. (VTI - Swedish National Road and Transport Research Institute), Sweden
Klipp, S. (BASt - Federal Highway Research Institute), Germany
Areal, A. (PRP - Portuguese Road Safety Association), Portugal
Jankowska-Karpa, D. (ITS - Motor Transport Institute), Poland

Revision history

Date	Version	Description
25/05/2023	1.0	Version for final KEG Alcohol approval
08/06/2023	1.1	Version approved by KEG Alcohol
18/09/2023	1.2	Version after reviews by the coordination team
16/11/2023	1.3	Version published at the website

Legal Disclaimer

Table of contents

About Trendline

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

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

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

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

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

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

Terms and definitions

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 - EC, 2021).

Motorcycle
Motorcycle 50cc up to 125cc: Two or three wheeled motor vehicle, with engine size 50cc up to 125 cc [and] maximum speed exceeding 45km/h (28 mph).
Motorcycle over 125cc: Two or three wheeled motor vehicle, with engine size more than 125 cc. (CARE - EC, 2021).

Light goods vehicle
Goods vehicle under 3.5 tonnes Maximum Gross Weight (3.5t mgw). Smaller motor vehicle used only for the transport of goods (CARE - EC, 2021).

Heavy goods vehicle
Goods vehicle over 3.5 tonnes Maximum Gross Weight (3.5t mgw). Larger motor vehicle used only for the transport of goods (CARE - EC, 2021).

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

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

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

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

Urban road (or road inside urban areas)
Public road inside urban boundary signs.

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

Week – nighttime
Monday 10 p.m. to Tuesday 5.59 a.m., Tuesday 10 p.m. to Wednesday 5.59 a.m., Wednesday 10 p.m. to Thursday 5.59 a.m., Thursday 10 p.m. to Friday 5.59 a.m.

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

Weekend – nighttime
Friday 10 p.m. to Saturday 5.59 a.m., Saturday 10 p.m. to Sunday 5.59 a.m., Sunday 10 p.m. to Monday 5.59 a.m.

1. Introduction

1.1. Context

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

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

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

1.2. Purpose and background of this document

This document presents the methodological guidelines for the KPI Driving under the influence of Alcohol. It describes the minimum methodological requirements to qualify for this KPI, defined as:

Percentage of drivers driving within the legal limit for blood alcohol concentration (BAC)

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, further referred to as "SWD (EC, 2019)" (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 Alcohol and the analysis of the following documents:

- Hakkert, A.S. & Gitelman, V. (Eds.) (2007) Road Safety Performance Indicators: Manual. Deliverable D3.8 of the EU FP6 project SafetyNet. http://www.dacota-project.eu/Links/erso/safetynet/fixed/WP3/sn_wp3_d3p8_spi_manual.pdf
- Hakkert, A.S., Gitelman, V. & Vis, M.A. (Eds.) (2007) Road Safety Performance Indicators: Theory. Deliverable D3.6 of the EU FP6 project SafetyNet. https://www.dacota-project.eu/Links/erso/safetynet/fixed/WP3/sn_wp3_d3p6_spi_theory.pdf See also https://www.dacota-project.eu/Links/erso/safetynet/content/wp_3_safety_performance_indicators_1.html
- Assum, T. et al. (2007) Druid Working paper "Uniform design and protocols for carrying out case-control studies". Deliverable D 2.1.2. of the EU FPD project DRUID. https://www.bast.de/Druid/EN/deliverales-list/downloads/deliverable_2_1_2.pdf;jsessionid=2F12FF52575D8C1057FCCCE0FFA509A9.live21304?__blob=publicationFile&v=1
- Yannis, G. & Folla K. (2022). Baseline report on the KPI Driving under the Influence of Alcohol. Baseline project, Brussels: Vias institute. https://baseline.vias.be/storage/minisites/baseline-kpi-alcohol.pdf

In addition to the specification of the **minimum requirements** (always marked bold) to deliver the national and disaggregated KPIs, this document also includes recommendations for optional, additional activities. Countries 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

Within Trendline two main measurement methods are allowed for the driving under the influence of alcohol KPI. The most appropriate and recommended method, if legally allowed, is **random breath testing** (RBT), i.e., roadside breath testing of randomly selected drivers by police on a representative sample of locations. The EC SWD (2019) expresses a clear preference for this KPI based on RBT, as this method is generally considered to deliver the most accurate picture of the situation. However, random testing is costly and not allowed in some countries. If it is not possible for objective reasons to use RBT, data on **self-reported drink-driving behaviour** can be collected **through representative anonymous surveys**. It is up to the countries to determine which method is feasible for them.

This limitation to two main methods is a deviation from the SWD (EC, 2019) and Baseline guidelines (Boets et al., 2021) where also breath testing results from enforcement actions (even if not random) were allowed. Within Baseline one country used alcohol test results from drivers after a crash. This country's KPI was not comparable to the other delivered KPIs.

KPIs calculated from different methodologies cannot be mixed. They have to be delivered and presented separately. Per method used, the aim is for all participating countries to have comparable indicators, at least for the minimum required stratifications.

2.1.1. Random breath testing (RBT)

Random breath testing data is gathered during roadside alcohol checks by the police. During these alcohol checks, drivers are randomly selected and stopped. The alcohol level of each stopped driver is assessed by means of a breath test. The objective is to estimate the KPI as defined by the SWD (EC, 2019): **percentage of drivers driving within the legal limit for BAC**.

The theoretical population (100%) refers to the total of all journeys (at least from the vehicle types being surveyed) over the national territory. In other words, this reflects the total number of kilometres driven. Hence, by weighting the results by number of kilometres driven on the different stratification variables used, the percentage of drivers respecting the legal limit will also reflect the percentage of kilometres driven with an alcohol concentration below the legal limit (guidelines on weighting will be provided separately by the Statistical Advisory Group).

For drink driving the main strata that are known to contribute to prevalence are time of day (day vs. night) and day of week (week vs. weekend) and interactions of both. Furthermore, different road types need to be considered**. Road type (3 strata) and week period (4 strata)** are the minimum required sampling stratifications for this KPI (see sections 2.3 and 2.4).

2.1.2. Self-reported behaviour

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

- Period-prevalence surveys measure the prevalence of driving below the legal alcohol limit over a specific period (e.g., last 30 days)
- Trip-based prevalence surveys measure the prevalence of driving below the legal alcohol limit during a specific randomly chosen recent trip.

Whereas random breath testing aims at measuring the 'point prevalence' or the proportion of the population driving within the legal blood alcohol level at any point in time, 'trip-based prevalence' aims at measuring the proportion of the population under the legal BAC limit during one of their last journeys, based on which point prevalence estimates can be deduced. The distinction is clear with the 'period-prevalence' where the self-reported frequency over a longer period is gathered.

Within Trendline it is determined that the period-prevalence measurement should **at minimum be in line with the ESRA method for sampling, questioning/filtering and weighting** (Meesmann et al., 2019). The question from ESRA3 (data collection in 2023) to be used, measures the frequency of '*driving when you may have been over the legal limit for drinking and driving over the last 30 days'* as a car driver. Responses are given by using a 5 points rating scale where 1 = never and 5= (almost) always. The respective **KPI refers to the percentage of car drivers indicating never to have driven when they may have been over the legal limit for drinking and driving' in the last 30 days**. The **minimum required KPI is the national aggregate KPI.** Furthermore it is highly recommended to also provide indicators by age (3), gender (2), road type (3) and week period (4) (see sections 2.3 to 2.5).

The trip-based prevalence method is closest to the aim of RBT since the drink driving state at a randomly selected specific recent point in time is questioned and concrete information on the trip is collected. The guidelines for this method are largely based on previous studies using this method (Diependaele & Silverans (2017), Diependaele (2015), Vollrath et al. (2019) and the German KPI alcohol study in Baseline (Schrauth & Funk, in press). A trip is randomly chosen, within the last 24 hours to up to within the last 7 days (to be specified in the metadata). The aim is that week period is taken into account in the random trip selection procedure, if possible also road type. **The KPI to deliver is the percentage of car drivers indicating not to have driven while being over the legal limit for drinking and driving in a recent trip. The minimum required KPIs are the national aggregate KPI and KPIs by road type (3 strata) and week period (4 strata)** (see sections 2.3 and 2.4). It is highly recommended to also provide indicators by age (3) and gender (2) (see section 2.5).

It is important that the self-reported KPIs are based on data from a representative sample.

For the period-prevalence survey based on ESRA the theoretical population is the national general population aged 18 and older. **The sample is representative for the national general population aged 18 and older in terms of age * gender (interlaced) and region/state**. In ESRA the sample for the calculation of the KPI alcohol estimates is a subsample of the representative general population sample, namely the persons that indicated to have driven a car in the last 30 days (or filter in ESRA: drove a car at least a few days a month).

For comparability and uniformity this is also the recommended way of sampling for the trip-based prevalence survey. Countries familiar with or aiming at using other sources for determining sample representativity (e.g., specific driver population statistics instead of general population statistics or aiming at including educational level) can also do so. The method should be specified in the metadata. In this survey method, persons are selected when they indicate to have driven in a recent specific period (in the last 24 hours or up to the last 7 days).

Countries are free to choose either of the two self-reported survey methods. If the trip-based prevalence method is used though, it is recommended to also include the period-prevalence question and provide this KPI as well.

Self-report methods can be self-administered (paper, online) or interviews (face-to-face - e.g., roadside, telephone) or any combination of these. The focus in this document is on the main recommended data collection method, namely representative anonymous online surveys. Main advantages of online surveys are: cost-efficiency (fast data collection, from a large sample), anonymity, no interviewer effect, less socially desirable answers.

2.2. Vehicle types

The EC SWD (2019) requires the inclusion of 'passenger cars as a minimum and goods vehicles, buses and motorcycles if possible (results disaggregated by vehicle type)'. In Baseline only KPIs for car drivers were delivered. Also in Trendline the **minimum requirement is to include passenger car drivers** for RBT as well as for self-reported data collection. Goods vehicles, buses/coaches and motorcycles are optional additional vehicle categories, for which separate KPIs (disaggregated by vehicle type) can be delivered. It should be acknowledged that including other vehicle types requires a larger sample size to get sufficient data for meaningful results per vehicle category (see section 3.1.3). If goods vehicles are included, light goods vehicles (LGVs) and heavy goods vehicles (HGV) are ideally differentiated.

If different vehicle types are included, the data collection should include a variable 'vehicle type' with the included categories. The different vehicle types should be clearly defined for the fieldwork and in the metadata: for example, some cars and LGVs share the same brand/model like Renault Kangoo (a passenger car has a backseat windows and passenger seats; a LGV has no backseat windows and no rear passenger seats).

For RBT the CARE definitions (2021) of vehicle types should be used (see *Terms and definitions,* page v). Passenger cars are defined as: '***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***'.

For the self-reported data collection the ESRA3 definition (or filter) of a car driver should be used, namely '***drive a car at least a few days a month'***. If other vehicle types are included, the same definition '*drive a xxx at least a few days a month'* should be used. If possible, the CARE definitions (2021) of vehicle types are also considered in the self-reported data collection.

2.3. Road types

The EC SWD (2019) specifies that three road types should be covered: (see definition in *Terms and definitions*, page v)

- **Urban roads**
- **Rural roads**
- **Motorways** (if expressways are included in the data collection, the results for expressways and motorways should be merged under the category *motorways*.)

Road type sampling stratification and provision of KPIs by road type is minimum required for RBT. For the period-prevalence survey provision of KPIs by road type is recommended. For the self-reported trip-based prevalence survey provision of KPIs by road type is minimum required.

The results should also be presented aggregated (after weighting) for the whole road network.

When a Member State's road network does not contain motorways, the overall results are calculated using the remaining road types. When a Member State's road network does contain all required road types, but not all road types are included in the study, the results for the remaining road types cannot be aggregated and remain disaggregated for each road type.

2.3.1. Random breath testing

RBT should provide a representative sample of all traffic in the country. This covers mostly three main road types. The **three minimum required road type strata are: motorways** (including expressways if this road type is also included – to be indicated in the metadata)**, rural roads and urban roads** (see *Terms and definitions*, page v). These are sampling strata for which also separate KPIs should be delivered. A deviation from this requirement is only possible in the exceptional case that a specific road type is non-existent in a country (e.g., no motorways in Latvia and Malta).

If Member States historically use a different road categorization, an attempt should be made to infer the minimum required road types.

Countries that used another definition for the road types in Baseline might be asked to recalculate their Baseline KPIs according to the Trendline definition (if feasible) in order to evaluate the impact of changed definitions on the key estimates.

RBT on motorways can be organised on the motorway but also at entrances or exits of motorways for feasibility reasons. This should be indicated in the metadata.

Main characteristics of the included road types should be described in the metadata (e.g., signs, speed regimes, number of lanes, lane separation, allowed vehicles) which allows to assess general correspondence of the road types between the countries (background/contextual information).

2.3.2. Self-reported behaviour

For the period-prevalence measurement, it is recommended to use the ESRA3 question on the different road types for respondents that answered they may have been driving at least once over the legal limit for drinking and driving in the last 30 days:

'You said that you have driven a car when you may have been over the legal limit for drinking and driving. Was this …? (multiple answers possible)

- *on motorways*
- *on urban roads*
- *on rural roads*

This allows providing additional KPIs by road type (e.g., *% car drivers indicating never to have driven when they may have been over the legal limit for drinking and driving on motorways in the last 30 days*). Ideally the ERSO/CARE definitions for road types (see above or in the *Terms and definitions*, page v) are also considered in the period-prevalence survey.

For the trip-based prevalence measurement, data on the road type of the trip should be collected: **motorways (including expressways), rural non-motorway roads and urban roads** (see definitions above or in the *Terms and definitions*, page v) and KPIs per road type should be estimated. The road type in this method can be defined as the main road type, i.e., on which the longest distance was driven.

2.4. Week periods

Week period sampling stratification and provision of KPIs by week period is minimum required for RBT. For self-reported trip-based prevalence the week periods are additional sampling strata (aiming at reaching a sufficient sample for each period) and provision of KPIs by week period is required. For period-based prevalence provision of KPIs by week period is a recommendation.

The Trendline week periods are based on ERSO (EC, 2022) defined as (see also *Terms and definitions*, page v):

- Working week – daytime
  - Monday to Friday 6.00 a.m. to 9.59 p.m.
- Working week – night
  - Monday 10 p.m. to Tuesday 5.59 a.m.
  - Tuesday 10 p.m. to Wednesday 5.59 a.m.
  - Wednesday 10 p.m. to Thursday 5.59 a.m.
  - Thursday 10 p.m. to Friday 5.59 a.m.
- Weekend – daytime
  - Saturday to Sunday 6.00 a.m. to 9.59 p.m.
- Weekend – night
  - Friday 10 p.m. to Saturday 5.59 a.m.
  - Saturday 10 p.m. to Sunday 5.59 a.m.
  - Sunday 10 p.m. to Monday 5.59 a.m.

2.4.1. Random breath testing

RBT should cover four week periods: weekday, weeknight, weekend day and weekend night. In order to harmonize definitions of the week periods, the definitions adopted in the ERSO project (*Terms and definitions*, page v) should be used. The week periods ideally include a good spread of all respective days and hours. When the alcohol checks are limited to specific days of the week or hours of the day (or a combination of those), this should be indicated in the metadata.

In Baseline no formal definitions were provided for the week periods but reference was made to the DRUID definitions (Houwing et al., 2011). In the end different definitions were used by the Member States. Possible impact of applying the required definitions in Trendline on the Baseline results is expected to be minimal, hence no recalculations of the Baseline results will be requested, even when in Baseline a different definition was adopted.

2.4.2. Self-reported behaviour

For the period-prevalence measurement, it is recommended to use the ESRA3 question on the different week periods for respondents that answered they may have been driving at least once over the legal limit for drinking and driving in the last 30 days:

'You said that you have driven a car when you may have been over the legal limit for drinking and driving. Was this ...? (multiple answers possible)

- *in the week during daytime*
- *in the week during night-time*
- *in the weekend during daytime*
- *in the weekend during night-time'*

This allow providing additional KPIs by week period (e.g., *% car drivers indicating never to have driven when they may have been over the legal limit for drinking and driving in the week during daytime in the last 30 days*). Ideally the ERSO definitions (see above and in the *Terms and definitions*, page v) for week periods are also considered in the period-prevalence survey.

For the trip-based prevalence measurement, the week periods are ideally part of the random trip sampling procedure (see section 3.2.1). Separate KPIs per week period - **weekday, weeknight, weekend day, weekend night** (see *Terms and definitions*, page v) should be estimated.

2.5. Driver and trip characteristics

Sampling stratification based on age * gender and region (for region see section 2.6) is required when using the self-report surveys and provision of KPIs by age and gender is highly recommended. For RBT it is recommended to collect data on age and gender of the driver and to provide separate KPIs.

2.5.1. Random breath testing

Countries with an interest in additional information on risk factors or predictors of driving under the influence (DUI) of alcohol can choose to record additional variables during the RBT.

The main variables which allow to identify the target groups that are of higher risk and which were already proposed in Baseline are driver gender (male, female) and driver age category (18-24; 25-64; 65+ - based on Vollrath et al., 2019). Within Trendline it is highly recommended to collect this data and to provide KPIs by driver gender and age group.

Furthermore, if different BAC limits are in force for specific driver groups it is recommended to additionally provide differentiated KPIs (KPIs for private vs. professional drivers, for novice drivers).

Other possible driver and trip related variables can be included in the data collection if countries are interested in exploring the problem of DUI of alcohol more in-depth, for example place of departure (e.g., café/bar, restaurant, feast/disco, family/friends, sport, home), motives for the trip (e.g., work, leisure, shopping, visiting family-friends), number of passengers in vehicle, number of minor (age < 18) / young (18 < age < 25) passengers in vehicle, (planned) duration of the trip, estimated kilometres of the trip. This data is not considered in Trendline but can be valuable input for evidence-based countermeasures (e.g., awareness campaigns) in a country.

2.5.2. Self-reported behaviour

For the self-report surveys the **sampling should be stratified by age * gender (interlaced) and region/state** (where the respondent is domiciliated) and any disproportionality should be corrected by using post-stratification weighting to be representative for the national population according to the used stratifications (guidelines on weighting will be provided separately by the Statistical Advisory Group). Optionally also educational level and/or other population characteristics are considered.

For the period-based prevalence survey the stratifications used in ESRA3 should be used (Meesmann et al., 2019): interlaced (crossed) quota for gender (male, female) * 6 age groups (18-24, 25-34, 35-44, 45-54, 55-64, 65-74) next to regional quota (not interlaced) based on national population statistics. It is highly recommended to provide KPIs by gender and by the same three age categories that are also recommended for the RBT (18-24; 25-64; 65+) to be uniform and in line with the Baseline indicators.

For the trip-based prevalence survey it is recommended to use the same ESRA3 sampling strata as for the period-based prevalence method (ESRA) to be representative for the national general population. It is highly recommended to provide KPIs by gender and by the same three age categories that are also recommended for the RBT (18-24; 25-64; 65+) to be uniform and in line with the Baseline indicators.

2.6. Region/state

2.6.1. Random breath testing

Disaggregation of results by region/state (where the driver is tested) is no requirement but countries are free to choose supplementary stratifications according to country regions (e.g., NUTS 1 regions). In such cases countries can consider collecting data from each region or from a representative selection of regions – see also section 3.1.1 (to be explained in the metadata). Countries aiming to have meaningful KPIs at regional level, including disaggregated regional KPI estimates according to road type and week

period, will need to multiply their location and driver sample to have a sufficient sample for meaningful results. If stratification in regions is used, ideally results are also weighted according to traffic volumes (by road type and week period) by region (guidelines on weighting will be provided separately by the Statistical Advisory Group). KPIs by region are not considered within Trendline.

2.6.2. Self-reported behaviour

The sample of the self-report surveys should be representative for age * gender (interlaced) and region/state (not interlaced), where region/state refers to where the respondent is domiciliated. KPIs by region are not considered within Trendline. Any disproportionality should be corrected by using post-stratification weighting to be representative for the national population according to the used stratifications (guidelines on weighting will be provided separately by the Statistical Advisory Group).

2.7. Blood alcohol concentration level

As the legal BAC limit ranges from 0.0 g/l to 0.8 g/l in Europe, the general KPI for alcohol ('below the legal limit') is not comparable between countries. Furthermore, legal limits can vary over time due to legal changes, which also influences the comparability of the results within a country.

When using RBT it is therefore highly recommended to provide complementary indicators to the KPIs, such as indicators on intoxication levels which would allow to better assess the DUI problem and compare the performance of countries with different maximum BAC limits. Such indicators are not subject to legal changes and differences within and between countries.

Such BAC categories reflect the risk that they pose for crashes and injuries based on, for example, the DRUID project (Schulze et al., 2012) – see Table 1 – and studies by Borkenstein et al. (1964) and Blomberg et al. (2005). Low BAC offenders cause relatively few road casualties compared to the high BAC offenders. An estimate on high BAC offenders (above 0,8 g/l) is thus of particular interest however for practical and strategic reasons ideally also indicators on lower BAC offenders are provided.

Table 1 Relative risk level of getting seriously injured or killed in an accident for various levels of blood alcohol concentration (Druid project, Schulze et al., 2012)

Risk level	Risk	Substance group
Slightly increased risk	1-3	0.1 g/l ≤ alcohol in blood < 0.5 g/l
Medium increased risk	2-10	0.5 g/l ≤ alcohol in blood < 0.8 g/l
Highly increased risk	5-30	0.8 g/l ≤ alcohol in blood < 1.2 g/l
Extremely increased risk	20-200	Alcohol in blood ≥ 1.2 g/l

To reinforce countries to also provide these indicators, these will be included as additional optional indicators in the Trendline datafiles:

- % drivers with a BAC ≥ 0.5 g/l
- % drivers within BAC levels: 0.5 - 0.79 g/l ; 0.8 - 1.19 g/l ; 1.2 or more g/l

3. Measurement procedure

3.1. Random breath testing

3.1.1. First phase of sampling: sessions

Random breath tests are typically conducted by setting up police alcohol checks at particular locations. The **selection of locations should be random, covering the entire geographical area of the country.** There are different options for random location selection: e.g., simple random or stratified random (e.g., random sampling in different regions, in different police jurisdictions or on different road types).

Bigger countries may consider in a prior stage the selection of one or more regions/states which are considered to be representative for the country with regard to driving under the influence of alcohol (e.g., Houwing et al., 2011). This can add to the fieldwork feasibility. If this is done, it should be explained in the metadata.

The basic procedure to randomly select locations consists of three steps:

1. The required number of locations (for the country or per region) is determined.
2. The number of locations is randomly selected on a map using the entire area in question (e.g., country or region), taking sufficient geographical spread into account. The specific requirements for each location (e.g., feasibility to set-up an alcohol check) do not have to be considered at this point yet. 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 or in cooperation with the police unit responsible for the respective location. Care should be taken to ensure that the locations for the different road types are sufficiently spread geographically.

A convenient way of selecting locations randomly (i.e. 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 no GIS software is available, step 2 can also be done manually using a national geographic map, e.g., Google Maps/Google Earth**. For the selection of final locations** (step 3) **practical arguments related to setting up roadside alcohol checks should be considered**: the ability to stop vehicles in a safe way and the ability to park the vehicles of the police, (researchers) and drivers who have been drinking above the legal limit. On motorways, controls can be set up at entrances and exits or transfer of drivers to rest and parking areas can be considered for safety reasons. Control sessions on high-speed roads should always be carried out in accordance with applicable (road) safety regulations. **Location bias must be avoided**: no specific

selection of locations based on proximity to places known for a higher DUI prevalence (e.g., near bars, discotheques…). Since random selection of locations will also include lower volume roads, it is expected that several low volume locations will be available for each stratum. If however traffic flow is too low (less than 10 cars passing per hour), it is acceptable not to include them. As mentioned before, sufficient geographical spread should be considered, so if random selection leads to locations on a same road, then it is suggested to randomly select a different location.

In the next step the **alcohol checks (sessions) are determined by attributing a week period (4) to each location in a balanced way over the three road types and geographical spread.**

The number of locations for each road type can be proportional to the actual traffic volume on each road type in the country (or region), assuming that each of the three road types represent a share of traffic volume above 20% of the total traffic volume. To do so, national traffic data (e.g., representative traffic/mobility surveys) is needed though. However, proportional sampling is not suitable for the different week periods since traffic volumes during weekend nights are generally very low, and strictly proportionate sampling according to traffic volume data would lead to much wider confidence intervals (less accurate estimates) for weekend night drivers than for higher volume time periods. For this reason, the **night-time periods should be oversampled** (and thus not to sampled proportionately to traffic volume), in order to guarantee sufficient numbers of observations.

If locations per road type and week period are not sampled proportionally (which is the case when for each stratum a same number of locations is selected), stratification weighting is needed to estimate nationally representative KPIs (specific guidelines will be provided by the Statistical Advisory Group).

As an absolute **minimum 10 different locations per level of stratification** variable are required with the aim of getting sufficient data for the entire road network and all week periods for meaningful KPI estimates. The required number of different locations is (see Appendix 2 for the argumentation behind the minimum location sample of 10 locations per stratification variable):

- **Minimum 10 locations on urban roads**
- **Minimum 10 locations on rural roads**
- **Minimum 10 locations on motorways** (including expressways if considered)

- **Minimum 10 locations on weekdays**
- **Minimum 10 locations on weeknights**
- **Minimum 10 locations on weekend days**
- **Minimum 10 locations on weekend nights**

The absolute **minimum is 30 different locations**. It is allowed to re-use a same location for sessions on different week periods. To ensure a balanced sampling for each combination of road type (3) and week period (4), a **minimum of 2 different locations for each combination of strata** (i.e., 12 crossed strata) should be used:

- **Urban roads x weekdays: minimum 2 locations**
- **Urban roads x weeknights: minimum 2 locations**
- **Urban roads x weekend days: minimum 2 locations**
- **Urban roads x weekend nights: minimum 2 locations**

- **Rural roads x weekdays: minimum 2 locations**
- **Rural roads x weeknights: minimum 2 locations**
- **Rural roads x weekend days: minimum 2 locations**
- **Rural roads x weekend nights: minimum 2 locations**

- **Motorways x weekdays: minimum 2 locations**
- **Motorways x weeknights: minimum 2 locations**
- **Motorways x weekend days: minimum 2 locations**
- **Motorways x weekend nights: minimum 2 locations**

The requirements concerning motorways (including expressways if considered) do not apply to countries with no motorways or where the network of motorways is very limited.

Ideally more than 10 locations for the different strata and more than 2 locations for the crossed strata are used for sampling, especially for (crossed) strata with generally low traffic volumes (e.g., nights). The recommendation is to boost the sample (especially on low volume strata) to allow a more accurate estimation of disaggregated indicators.

The crossed strata sampling design should be explained in the metadata.

If countries optionally want to have regional KPIs (e.g., NUTS 1) including all road and week stratifications per region, then it is recommended to use the sample size guidelines (for each stratum) per region. Regional KPIs are not considered within Trendline.

For more information on random sampling of locations and determining minimum sample sizes, we can refer to the SafetyNet general recommendations for SPIs (safety performance indicators): (Hakkert & Gitelman, 2007) http://www.dacota-project.eu/Links/erso/safetynet/fixed/WP3/sn_wp3_d3p8_spi_manual.pdf

3.1.2. Second phase of sampling: individuals

Drivers should be sampled randomly, irrespective of any suspicion for DUI of alcohol. Any selectivity, either in the locations chosen (e.g., risk areas for drink driving) or in who is being checked and who is not, leads to a bias and decreases the representativeness of the data. RBT should therefore be done in collaboration with police forces, as in many countries they (and probably only they) have a legal basis for stopping drivers and testing all drivers stopped.

Researchers generally do not have nor can acquire the legal right to stop drivers randomly on the road and test all drivers stopped. In some countries RBT is also not allowed for the police – they can only stop drivers in case of certain suspicious signs. In that case RBT cannot be used. Breath testing based on voluntary participation cannot not be considered as an alternative as it is methodologically difficult to get a representative sample based on this. Voluntary participation leads to a self-selection or participation bias. It is therefore required to collaborate with police forces which can randomly stop drivers and test each stopped driver.

RBT on a random sample of locations can mean for police forces that they must adapt their normal way of working. Often alcohol enforcement is done in a selective way (for the selection of locations, periods and drivers). It is therefore necessary to verify that the sampling is done randomly as expected. If researchers always assist the fieldwork, they can assure that testing is done randomly. This is the preferred situation. If the police do the sampling without the presence of a researcher, results may not be fully random. In that case it is recommended to evaluate if there are indications at session level of selectivity (e.g., outliers in terms of the proportion of drivers with a BAC above the limit).

3.1.3. Minimum driver sample

Defining a minimum required sample size is by definition arbitrary since it depends on the level of accuracy that is considered adequate. With the national KPI estimates in Baseline in the range of 99 to 96% (Yannis & Folla, 2022), the accuracy (width of the 95% confidence interval) ranges between 0.4 to 0.9 percent points with a sample of 2,000 drivers (see Table 2 – calculations assuming simple random sampling[1]), which can be considered a sufficient accuracy. **Minimum 2,000 car drivers is the required sample for the national KPI for DUI of alcohol (for 3 road types, 4 week periods, passenger cars).** This minimum refers to valid datapoints in the dataset in order to be considered for the national KPI.

Table 2 Width of the 95% Confidence Interval depending on KPI levels and sample sizes, assuming simple random sampling, based on the formula. [2]

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

KPI estimate	Lower bound CI n=2,000	Upper bound CI n=2,000	Lower bound CI n=500	Upper bound CI n=500	Lower bound CI n=250	Upper bound CI n=250
99	98.56%	99.44%	98.13%	99.87%	97.77%	100%
96	95.14%	96.86%	94.28%	97.72%	93.57%	98.43%
88	86.58%	89.42%	85.15%	90.85%	83.97%	92.03%

*National KPI alcohol levels in Baseline ranged between 96 and 99% while weekend nights yielded percentages down to 88% (Yannis & Folla, 2022), (z value 1.96 for 95% CI)

Since coverage of the three road types and four week periods is required (see sections 2.3 and 2.4), it is highly recommended that each of these strata cover at least 20% of the total data collection, even if this requires disproportionate sampling (e.g., oversampling during the nights), to ensure a sufficient number of drivers for each stratum and allow the calculation of sufficiently accurate KPIs per stratum.

It is highly recommended to have at least 500 drivers for each road (3) and week (4) stratum. With week period KPIs down to 88% (i.e., lowest % for weekend nights in: Yannis & Folla, 2022), a sample of 500 drivers leads to an estimate of 88% with a 95%-confidence interval of 85.2% - 90.9%, so ± 2.9 percent points (see Table 1).

[1] Following the recommendations of Baseline, the information on calculating confidence intervals may be updated in separate guidelines from the Trendline Statistical Advisory Group.
[2] Calculations can also be done via: https://sample-size.net/confidence-interval-proportion/

Based on the feasibility issues in Baseline to reach the original minimum sample per stratum (500 drivers), it is decided within Trendline to decrease the minimum driver sample for 'difficult strata' (based on Baseline: for some countries one or both night periods, for other countries one or both weekend periods) to 250 drivers. It should be noted that this leads to bigger error margins for the point estimate though, e.g., (see Table 2) for a KPI estimate of 88% based on a sample of 250 drivers the 95%-confidence interval is 84.97% - 92.03%, so ± 4 percent points (instead of ± 2.9 percent points with a sample of 500 drivers).

This indicates the importance of presenting the 95%-confidence intervals together with the KPI estimates.

In summary, the **required minimum sample sizes for each road (3) and week (4) stratum are:**

- **minimum 500 car drivers on urban roads**
- **minimum 500 car drivers on rural roads**
- **minimum 500 car drivers on motorways** (including expressways if these are considered)

- **minimum 500 car drivers on weekdays**
- **minimum 250 car drivers on weeknights** (highly recommended: min. 500)
- **minimum 250 car drivers on weekend days** (highly recommended: min. 500)
- **minimum 250 car drivers on weekend nights** (highly recommended: min. 500)

KPIs based on smaller stratum samples than 250 drivers may still be reported within the Trendline results but with an indication of 'deviation to the minimum requirements. KPIs based on stratum samples between 250 and 499 drivers may also be indicated with 'small samples' in the reporting.

If specific strata show particularly low levels of legal BAC limit compliance (such as weekend nights), countries can consider increasing the sample size for highly problematic strata in order to obtain a more detailed view on predictor variables (more accurate disaggregated indicators).

It is difficult to guarantee a minimum number of observations for all possible combinations of levels of stratification. For the combination of road types and week periods, this would lead to 3 x 4 = 12 levels. If countries optionally want to have accurate KPI estimates for all possible combinations of stratifications, then it is recommended to consider the sample size guidelines indicated above (per stratum) for each crossed stratum, thus boosting their sample from what is minimum required. KPIs for each combination of week and road stratum are considered within Trendline.

If countries optionally want to have accurate indicators for all stratifications for other vehicle types than cars, then it is recommended to consider the same sample size guidelines (indicated above) as for cars. In practice, data collection will often be done for all considered vehicle types together during the alcohol checks. If drivers are stopped randomly (also irrespective of the vehicle type) during the alcohol checks, then the frequency of the vehicle types in the sample will correspond to the proportion of the vehicle types in the actual traffic. If 2,000 cars drivers are sampled this way, then the number of other vehicle types is expected to be (much) lower. It is also possible to oversample the other vehicle types though. The minimum required KPI estimates (national, by week period, by road type) only refer to car drivers. If optional KPIs for other vehicle types will be considered in Trendline, these will be presented separately.

If countries optionally want to have regional KPIs (e.g., NUTS 1) including all stratifications per region, then it is recommended to consider the sample size guidelines indicated above (i.e. total and per stratum on the national level) per region. Regional KPIs are not considered within Trendline.

It can be summarized that multiplying the minimum sample sizes can increase the accuracy of the estimates and allow delivery of additional disaggregated KPIs. It is up to the country to decide on this.

Appendix 2 gives an overview of the argumentation behind the minimum driver sample.

3.1.4. Fieldwork set-up and procedure

A uniform data collection procedure should be decided upon at the start. The main elements should be mentioned in the metadata.

Typically, drivers are sampled in police alcohol checks at a particular location where several drivers are being checked for the duration of the control session. The police use their standard (approved and calibrated) RBT measuring devices (e.g., Dräger, Honac). After a positive test the normal legal/sanctioning procedures follow.

Each location corresponds to one or more (e.g., different week periods) alcohol checks (sessions). When planning the checks, it should be ensured that all combinations of road types and week periods are sampled with a minimum of 2 sessions/different locations. Each week period should include a balanced spread of all respective days and hours.

There are no weather, visibility or road quality prerequisites for organizing the RBT, only practical feasibility considerations for setting up an alcohol check: ability to stop vehicles in a safe way and to park the vehicles of the police, (researchers) and drivers above the legal limit, sufficient traffic flow (it is acceptable not to do sessions on places/times where it is expected that less than 10 cars pass per hour). Care should be taken that drivers can not take escape routes.

The minimum sample requirements refer to the number of locations/sessions (see section 3.1.1) and drivers (see section 3.1.2). Countries can estimate and follow up the required number of sessions and control hours to reach at least the minimum requirements.

Controls on high-speed roads should always be carried out in accordance with applicable (road) safety regulations. For safety and feasibility reasons RBT on motorways (incl. expressways) can be organised at exit or entrance lanes of the motorway.

Each session should last minimum 30 minutes (traffic count – if done separately – not included, see section 3.1.5). Control sessions can last longer, but 30 minutes to 1 hour is recommended because the longer a location is in use, the bigger the risk of drivers becoming aware of the alcohol check (e.g., through alerts on social media) and subsequently avoiding it.

Ideally all drivers are stopped and tested but if this is not possible (traffic volume, police capacity), the main procedure is that after finalising the procedure with one driver, the first arriving next driver should be stopped and tested.
On locations and times where a high traffic volume is expected it is recommended to increase the police capacity at the alcohol check.

The required data is coded as an additional task by the police or by a research worker assisting the control session. Ideally research workers assist all alcohol checks. They can assist the police in the data coding and meanwhile verify that the methodology is fully complied with (correct location and week period, breath tests done in a random way).

Suggestions to decrease the risk of non-response to optional additional questions to the driver (see section 2.5 on collecting optional additional driver/trip characteristics) are: let the police ask the questions, have the breath test after the questions, provide small incentives and limit the number of questions.

For the on-site coding, paper sheets or tablet computers/smartphones can be used. Using a tablet or smartphone can have some advantages (e.g., direct coding, real-time central data collection, automatic coding of metadata like the exact location, date and hour 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 as better than paper. For Trendline a dedicated software for behavioural measurements called 'SPIN' was developed by CDV, nevertheless not for driving under the influence of alcohol so far.

Screenshots of the 'SPIN' software for seatbelt use (not available so far for KPI alcohol).

3.1.5. Counting of traffic volumes

Traffic volume should be counted during each control session, even when national traffic volume statistics are available. This information is needed for correct calculation of the confidence intervals (weighting) (add a reference to the document on weighting and CI).

The counting can be done by a research worker of by one of the policemen.

Ideally all passing (including stopped) cars at the alcohol check are counted during the entire duration of the alcohol check.

If this is not possible, **as a minimum traffic counts of passing (including the stopped) cars should be done during 10-minutes in the middle of the session or during 5 minutes before and 5 minutes after the session**. The counting time is not part of the minimum 30 minutes alcohol check duration unless it is done in parallel with the RBT (which is possible if there is enough police capacity or if traffic volume is very low). **It is required that the number of counted cars (including stopped ones) and the duration of the count are always coded together to be able to correctly calculate the number of passed vehicles per minute** (i.e., the traffic volume during the session) and so to avoid mistakes in that calculation.

If different vehicle types are considered, these are ideally counted separately. This allows the calculation of specific weight factors per vehicle type (more info on this will be provided by the Statistical Advisory Group).

3.2. Self-reported behaviour

3.2.1. Sampling

Random sampling is required. Random sampling from available online research panels is allowed (e.g., from market research agencies like iVOX, Ipsos). Convenience samples are too biased to generate a representative estimate.

For estimation of period-prevalence, the sampling method should be **in line with the ESRA method** (Meesmann et al., 2019). The same minimum sampling criteria as in ESRA3 should be used: **interlaced (crossed) quota for gender (male, female) * 6 age groups (18-24, 25-34, 35-44, 45-54, 55-64, 65-74) next to regional quota (not interlaced) based on national population statistics** (proportionate stratified sampling). Any disproportionality is corrected using post-stratification weighting in order to be representative for the national population according to age category * gender and region. The use of the same data source for weighting in the different countries is important for international comparability. In ESRA UN population statistics are used, but for Trendline it can be suggested to use Eurostat which also provides information on regions.
This way the total sample is representative for the national population (aged 18 and more) with regard to age * gender and region/state. The sample for the calculation of the KPI estimates is a subsample of this: persons who drove a car at least a few days a month, i.e., car drivers.

For estimation of trip-based prevalence, the sampling is ideally also done using quota for gender (2) * age (3 groups – 18-24, 25-64, 65+ – or 6 groups like in ESRA) (interlaced) and region (not interlaced), like for period-based prevalence, with disproportionalities being corrected afterwards using stratification weighting. The **sample for calculating the KPIs should be representative in terms of age * gender and region for the theoretical population** (general population 17+ like in ESRA, or more specifically the car driver population 17+). For international comparability the ideal is that comparable sources are used as weighting base, like the UN or Eurostat general population statistics. On the other hand, countries may use other sources for historical comparability (e.g., German KPI alcohol study in

Baseline used more specific driver population statistics). A first sampling filter is on the car driving status: respondents should have been driving at least once a car in the considered recent period for the trip selection. The period of consideration for the recent trip selection can be within the last 24 hours (e.g., Diependaele & Silverans, 2017) to up to the last 7 days (e.g., Schrauth & Funk, in press). The random selection of a trip is ideally done in a stratified way taking the four week periods (see section 2.4) into account, and ideally even also the three road types (see section 2.3), with the aim of sampling sufficient data for each road and week stratum and allow the calculation of more accurate indicators. It is also possible to sample a recent trip completely randomly. For reaching a sufficient sample for difficult strata like the nights though, it is recommended to oversample trips at night, which should be corrected afterwards with stratification weighting according to the actual share of the nights in a week. Like for RBT it is recommended to weight results of trip-based prevalence surveys also according to traffic volumes by week period and road type. The determination of road type in this method is less straightforward as in the RBT as during one trip more than one road type can be driven on. The proposal is to ask for the main road type of the trip, i.e., on which the longest part of the trip (distance) took place.

As the method of the trip-based prevalence survey is more dependent on specific country choices it is very important to give more information on the methodology, including the sampling, in the metadata. This is required in order to verify the level of international comparability.

Reference can be made to two published studies using the trip-based prevalence survey, as example:

- Schrauth & Funk (in press) measured the Baseline KPI alcohol by asking information on a randomly selected trip in the last 7 days in a sample that is representative of the national driving population 17+.
  More info: The population for this survey are persons aged 18 and older who have driven a car in the last 30 days before the study. An online access panel was used to recruit the respondents. A random selection of participants was made using quota according to gender, age, region (Federal States of Germany) and education to ensure representativeness with regard to these strata. The final dataset comprised a sample of 4,459 respondents with at least one trip as a car driver in the last seven days in Germany. Respondents were first asked to indicate the days from the past week (last 7 days) on which they were on the road as a car driver. Then one day was selected randomly. For this day, individual trips and their starting times were asked. If several trips were reported, one trip was selected randomly. For this selected trip additional questions were asked, including the one for KPI alcohol. Results were weighted by gender, age and region.

- Diependaele & Silverans (2017) measured the prevalence of sleepiness in car drivers by asking information on a pseudo-randomly selected trip in the last 24 hours in a survey sent to a sample that is representative of the national population 17+.
  More info: The survey is distributed to a large number of potential respondents (more than 130.000 to reach in the end 2,500 valid respondents), representative of the national population above age 17. Persons indicating to have driven a car in the last 24 hours were included. Maximum heterogeneity in the timing of the trips is aimed at by sending out the invitations to online panel members in small batches each hour of the day (day and night). Half of the invitations are sent during the week and the other half during the weekend. This is done to facilitate a balanced statistical comparison (i.e., similar precision in estimates for) between the behaviour during the week and during the weekend. Survey invitations are distributed evenly across each week/weekend day. Each panel member that agrees to participate is only granted access to the survey once. The sampling takes different time slots into account (6-12am, 12am-6pm, 6pm-12pm, 12pm-6am). Respondents are asked if they have driven a car in one or more of these slots (in the last 24 hours). Then the trip of interest is selected randomly as either

*the first or the last trip within the given week period. If more than one period is checked, one of these is sampled. This is done in a pseudo-random way, i.e., a random choice is made, except when a 'night-time' period is checked. In that case, the night-time slot is always chosen. This is done in order to avoid low precision (i.e., large confidence intervals) for prevalence estimates at night due to the lower traffic volume at that time. In order to correct for the deliberate oversampling of night-time driving, the night trips receive a lower weight. Specifically, trips receive a weight proportional to the ratio of (a) the number of times the time slot has been selected by the pseudo-random algorithm and (b) the overall number of times the slot was checked by the respondents. The underlying assumption is that the latter distribution reflects the natural distribution of departure times (in terms of the current week periods) in the population of car drivers. Furthermore, weights are applied proportional to the ratio of (a) the age * gender and region distribution in the sample and (b) the age * gender and region distribution of eligible drivers in the population.*

3.2.2. Minimum sample sizes

Sample size guidelines depend on the type of questionnaire survey:

- For KPI estimates of period-prevalence, **the minimum total sample is 750 car drivers**. Ideally though data from at least 1,000 car drivers is sampled for this KPI. The reduction to 750 car drivers compared to the 1,000 car drivers in Baseline is to be more in accordance with ESRA3 for delivering the period-prevalence KPI for alcohol. The ESRA3 country sample generally consists of 1,000 road users, not car drivers. The car drivers are always a selection of the total and this ranged in ESRA2 between 49% and 91% of the total. The EU mean is 75,9% car drivers in the sample.
  - This is the minimum for the required national KPI.
  - Disaggregated indicators can be delivered by age, gender as well as by week period and road type (recommended). The results are always presented with the 95% confidence intervals and small subsamples are marked (<250).

- For estimates of trip-based prevalence, countries can define the required sample size. The minimum sample of car drivers will have to be a multiple of 1,000. A **minimum total sample of 2,500 car drivers** is required (cf. Diependaele, 2015 for an example).
  - This is the minimum for the required national KPI.
  - Disaggregated indicators by week period and road type are also required.
  - Disaggregated indicators by age and gender are recommended.
  - The results are always presented with the 95% confidence intervals and small subsamples are marked (<250).

These numbers refer to completed and validated surveys, so the final sample for calculating the indicators.

3.2.3. Questions

The period-prevalence survey should be in line with the ESRA method (Meesmann et al., 2021). As a minimum requirement the question on driving over the legal limit for drinking and driving used in the ESRA3 survey (https://www.esranet.eu/) should be used in order to ensure international and historical comparability (see box below). This question is asked to respondents that indicated at the start of the survey to drive a car at least a few days a month. A complete overview of the ESRA2 methodology can be found at: https://www.esranet.eu/storage/minisites/esra2-methodology-report-updatewave2-def.pdf.

ESRA 3 questionnaire: Self-declared safe and unsafe behaviour in traffic :

Over the last 30 days, how often did you as a CAR DRIVER …?
(car driver defined as: drive a car at least a few days a month)
You can indicate your answer on a scale from 1 to 5, where 1 is "never" and 5 is "(almost) always". The numbers in between can be used to refine your response.
Binary variable for all items: at least once (2-5) - never (1)
Items (random):
• ***drive when you may have been over the legal limit for drinking and driving*** (minimum)

If the answer was at least once on the first question, it is followed by:
You said that you have driven a car when you may have been over the legal limit for drinking and driving. Was this …? (multiple answers possible)
• *in the week during daytime*
• *in the week during night-time*
• *in the weekend during daytime*
• *in the weekend during night-time*
• *on motorways*
• *on urban roads*
• *on rural roads*

As the ESRA3 question referring to '*may have been over the legal limit for drinking and driving*' is closest to the KPI definition this is the minimum required question.

The minimum overall KPI thus reflects the **percentage of car drivers indicating 'never' to have driven when they may have been over the legal limit for drinking and driving' in the last 30 days.** The filter question allows to also provide this indicator for the 4 week period strata and 3 road type strata, which is highly recommended.

In the trip-based prevalence survey respondents are selected if they indicated to have driven a car at least once in a specific recent period (e.g., last 24h, up to last 7 days). It is important to indicate that the question concerns driving on public roads between two geographically distinct locations, excluding short breaks (e.g., at gas stations). Then a trip is randomly selected in the questionnaire – for this selection there are different options, e.g., (based on the methods used in previous studies like the German KPI alcohol study in Baseline (Schrauth & Funk, in press) and Diependaele & Silverans (2017):

- Recent period = last 7 days:
  - Respondent is asked on which days of the last 7 days he/she drove (Monday to Sunday)
  - Random selection of one of the indicated days unless only one day was ticked. It can be recommended to oversample weekend days in order to have enough data for weekends though (stratified sampling)
  - Question on how many trips he/she made as a car driver on that day and specification per trip of the departing hour (1st trip: hh:mm, 2nd etc until 7th)
  - Random selection of one of the indicated trips unless only one was indicated. It can be recommended to oversample night trips in order to have enough data for the nights though (pseudo-random selection: if a night trips are indicated then one of these is selected)

- Recent period = last 24 hours: (by sending out the survey in small batches each hour of the day, half in the week and half in the weekend, the expectancy is that 'the last 24h' will comprise all week periods in the final sample)
  - Respondent is asked in which time spans he/she drove in the last 24 hours (cfr. time spans in hours in the EC definitions of week periods, see section 2.4).
  - Random selection of one of the time spans unless only one was ticked. It can be recommended to oversample the night-time span in order to have enough data for nights (pseudo-random selection: if night-time is ticked then this time span is selected)
  - Random selection of the first or last trip in the chosen time span.

Road type (longest distance; see section 2.3) can be added as sampling stratum (like week period) – allowing stratified sampling with the aim of gathering sufficient data per stratum – or can be asked as a variable on the selected trip.

The respondent is then asked to reflect on this trip. The proposed question for the KPI alcohol is **'Do you think that you were driving above the legal limit for [alcohol in your blood OR drinking and driving] during this trip?'** Member States are free to add more questions on the trip, e.g., trip duration, estimated kms, presence of passengers (see section 2.5).

In addition to the trip-based question it is recommended to also include the ESRA3 period-prevalence question in this survey.

The minimum overall KPI thus reflects the **percentage of car drivers indicating not to have driven when they may have been over the legal limit for drinking and driving in a random recent trip.**

3.3. Temporal considerations

Ideally the data collection is done in late Spring or early Autumn (EC SWD, 2019). In practice, all months are allowed except December, January, July and August. **Holiday periods (bank / school holidays) and hard winter conditions should be avoided,** as these disturb normal traffic patterns.

When countries have historical series of measurements it is recommended to use the same period(s) of the year as for the earlier measurements.

Countries willing to organise more than one data collection period to deliver the KPIs (e.g., one in Spring and one in Autumn) need to comply with the minimum sample size requirements for both measurements combined. The data from both measurements can be combined to deliver the overall and disaggregated indicators.

The COVID-19 pandemic and related national policies had implications on DUI of alcohol behaviour. It is recommended not to plan data collection in case the traffic situation and mobility patterns in a (large part of the) country are very different from the normal situation (low representativity), as well as when (temporary) restrictive policy measures are in force such as a lock-down, a night curfew, closed bars/restaurants, limitations of social contacts etc., because these relate to typical risk factors for DUI of alcohol.

These guidelines and restriction apply to RBT as well as to the self-report methods. In order to get representative KPIs, it is important to collect data in a sufficiently representative context.

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.

4.1.1. Random breath testing

As a first guideline, it is suggested to include/process for each driver in the dataset, the following variables:

- Driver level:
  - exact alcohol test result (in BrAC or BAC – to be specified)
  - **binary alcohol test result (2) (below or above legal BAC limit)**
  - driver type (private, professional)
  - driver experience (novice or not)
  - applicable legal limit in BAC g/l
- Session level:
  - **road type (3)**
  - **week period (4)**
  - **date**
  - start hour
  - end hour
  - **total duration of the alcohol check** (for weighting)
  - **unique location code** (to know which breath test results belong to a same session, e.g., geo-coordinate or a qualitative code)
  - unique session code (needed if a same location is used for different sessions)
  - traffic count duration (minutes),
  - total counted of (passed by including stopped) cars (minimum) (if more vehicles are considered: this is per vehicle type)
  - **traffic count per minute** (= total count of passing cars / count duration in minutes)

Other variables can be interesting to code/process too (optionally in addition; see section 2.5): e.g.,

- Driver level:
  - category of BAC in g/l (see section 2.7) (highly recommended)
  - driver age category (highly recommended)
  - driver gender (highly recommended)
  - exact BAC in g/l
  - exact breath alcohol concentration (BrAC) in mg/l
  - category of BrAC in mg/l
  - vehicle type (if other than cars are considered)
- Session level:
  - region/state
  - police zone
  - weather condition

Recent types of alcohol testers can automatically store the data on the breath test outcomes which can be exported later. This allows automatic recording of date and time for each breath test result, and this could also be used for setting up the dataset, as long as the additionally required session variables are also collected and correctly combined with this data. If countries optionally wish to collect additional driver characteristics, then automatically stored data is not sufficient because the breath test result should be coded in combination with the additional variables so as to ensure the data link.

4.1.2. Self-reported behaviour

The main driver variables to be coded/processed for the self-report methods are: (see also sections 2.5 and 3.2)

Period- prevalence: (for most, if not all countries, this data collection is done within ESRA3)

- **age**
- **gender**
- **region/state (domicile)**
- response to the KPI alcohol question for car drivers: 5 point rating scale (never to (almost) always)
- **dichotomized response on the KPI alcohol question for car drivers: never or at least once** (KPI refers to % never)
- response to the additional filtered questions
- for each of the 4 week periods: yes or no (KPI refers to % no)
- for each of the 3 road types: yes or no (KPI refers to % no)

Trip-based prevalence:

- **age**
- **gender**
- **region/state (domicile)**
- trip info:
  - date
  - **week period**
  - **main road type**
  - day of week
  - departure time
  - estimated kilometres
  - duration
- **response to the KPI alcohol question for car drivers: yes, no (KPI refers to % no)**

4.2. Post stratification weights and statistical analysis

Specifications on calculating weights and confidence intervals are provided in Appendix 3 *Suggested approach for weighting sample data and calculation of statistics.*

4.3. Expected results and data delivery

For each provided indicator, a **point estimate as well as a 95% confidence interval** is expected. Results should also include the **unweighted number of drivers** and **number of locations** the results are based on. The final information on the expected results and data delivery will be provided later together with the datafile info. The current information is definite only with regard to the minimum requirements.

For RBT the **main indicator is the percentage of car drivers below the legal BAC limit over all road types (3) and week periods (4), i.e., the weighted national aggregate KPI for car drivers. An estimate is also expected for each level of the following stratification variables:**

- **KPIs by week period (4 levels)**
- **KPIs by road type (3 levels)**

Optionally also other indicators can also be provided. It is highly recommended to also provide (see sections 2.5 and 2.7) :

- KPIs for combinations of week periods (4 levels) and road types (3 levels)
- KPIs by age category (3 levels)
- KPIs by gender (2 levels)
- KPIs according to the level of BAC (4 levels)
- KPIs by type of driver (professional, private)
- KPIs for novice drivers (< 2 years driving licence – if differently defined this should be added in the metadata)
- KPIs for other vehicle types than cars

For the self-reported period-prevalence survey **the main indicator is the overall percentage of car drivers indicating never to have driven when they may have been over the legal limit for drinking and driving in the last 30 days, i.e., the weighted national aggregate KPI for car drivers.**
It is highly recommended to also provide: (see section 2.5)

- KPIs by age category (3 levels)
- KPIs by gender (2 levels)
- KPIs by week period (4 levels)
- KPIs by road type (3 levels)

For the self-reported trip-based prevalence survey **the main indicator is the overall percentage of car drivers indicating not to have driven while being over the legal limit for drinking and driving in a recent trip over all road types (3) and week periods (4), i.e., the weighted national aggregate KPI for car drivers. An estimate is also expected for each level of the following stratification variables:** (see sections 2.3 to 2.5).

- **KPIs by week period (4 levels)**
- **KPIs by road type (3 levels)**

It is highly recommended to also provide: (see section 2.5)

- KPIs by age category (3 levels)
- KPIs by gender (2 levels)

For both self-report survey methods, optionally also other indicators can be considered like:

- KPIs by type of driver (professional, private)

- KPIs for novice drivers (< 2 years driving licence – if differently defined this should be added in the metadata)
- KPIs for another period than 'last 30 days'
- KPIs for different vehicle types than cars

For all methods, it is recommended to also provide crossed point estimates for all considered levels of disaggregation.

It should be acknowledged that the required alcohol KPIs are based on the legal limit in a country which can vary. Therefore, the estimates should always be interpreted against the background of the legislative information provided in the metadata.

For the **data delivery** to the Trendline consortium further instructions on dataset structure and variables will be provided later. As a first guidance reference is made to the Baseline datafiles:

- **Aggregate datafile: all minimum required KPI point estimates (%) and 95%-confidence intervals overall and by stratum.** In Baseline there were also optional entries according to driver type (private vs. professional).
- Semi-aggregate datafile (optionally recommended): crossed matrix of all considered levels of disaggregation (crossed point estimates) and 95%-confidence intervals.
- Variables: in addition to the KPI estimates and 95%-confidence intervals **additional specific stratum related data will have to be coded** like in Baseline (number of locations, number of drivers…).

4.4. Metadata

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

As a guidance, the metadata includes the main methodological information, like:

- Type of study (RBT, self-report: trip-based, period) including also:
  - RBT: collaboration with police or not
  - Self-report: survey method
  - Self-reported period-prevalence: data collection within ESRA3 scope or not
  - Self-reported trip-based prevalence: also period-prevalence question
- Considered sampling stratifications
- Info on the random sampling, including also:
  - RBT: method used and rationale for choosing locations (sampling method, inclusion/exclusion criteria, prerequisites, minimal traffic flow considered) ; definitions ; method for allocating week periods
  - Self-report surveys: sampling source, sampling quota (strata)
- Considered vehicle types
- Description of the fieldwork procedure, including also
  - RBT: number of locations and number of sessions in total and per (crossed) sampling stratum, considered days of the week and hours of the day for the week periods, coding tool, variables collected, period, session duration average …
  - Self-report: exact question(s) and filters
- Total driver sample and by main strata
- Description of the weighting/representativeness
  - Available national traffic volume data

- Traffic count per session, duration, vehicle types
- Statistical techniques to weight the data and to calculate the CIs
- weight formula

- Applicable regulations and procedures related to this KPI during the fieldwork: legal BAC limit in g/l in general for cars and for different driver and vehicle types if these are considered.

References

Assum, T. et al. (2007) Druid Working paper "Uniform design and protocols for carrying out case-control studies". Deliverable D 2.1.2 https://www.bast.de/Druid/EN/deliverales-list/downloads/deliverable_2_1_2.pdf;jsessionid=2F12FF52575D8C1057FCCCE0FFA509A9.live21304?__blob=publicationFile&v=1

Blomberg, R., Peck, R.C., Moskowitz, H., Burns, M. & Fiorentino, D. (2005) Crash Risk of Alcohol Driving: A Case-Control Study. Stamford, CT: Dunlap & Associates.

Boets, S., Silverans, P., Houwing, S., Forsman, Å., Klipp, S., & Folla, K. (2021). Methodological guidelines – KPI Driving under the Influence of Alcohol. Baseline project, Brussels: Vias institute. https://baseline.vias.be/storage/minisites/methodological-guidelines-kpi-alcohol.pdf

Borkenstein, R.F., Crowther, R.F., Shumate, R.P., Zeil, W.W. & Zylman, R. (1964) The role of the drinking driver in traffic accidents. Bloomington, IN: Department of Police Administration, Indiana University.

Diependaele, K. & Silverans, P. (20175) Monitoring driver sleepiness using single trip survey data. TRB 96th Annual Meeting Compendium of Papers. https://trid.trb.org/view/1437298

Diependaele, K. (2015) Sleepy at the wheel. Analysis of the extent and characteristics of sleepiness among Belgian car drivers. Brussels, Belgium : Belgian Road Safety Institute – Knowledge Centre Road Safety. https://www.vias.be/publications/Slaperig%20achter%20het%20stuur%202017/Sleepy_at_the_wheel.pdf

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

European Commission (2019) Commission staff working document EU road Safety Policy Framework 2021-2030 - Next steps towards "Vision Zero". SWD (2019) 283 final. Retrieved from https://ec.europa.eu/transport/sites/transport/files/legislation/swd20190283-roadsafety-vision-zero.pdf

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

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

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

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

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

https://www.dacota-project.eu/Links/erso/safetynet/content/wp_3_safety_performance_indicators_1.html

Houwing, S., Hagenzieker, M., Mathijssen, R., Bernhoft, I.M. Hels, T., van der Linden, T., Legrand, S.-A. & Verstraete, A. (2011) Prevalence of alcohol and other psychoactive substances in drivers in general traffic; Part I: General results. Deliverable D2.2.3 Part 1 of the EU FP6 project DRUID. https://www.bast.de/Druid/EN/deliverales-list/downloads/Deliverable_2_2_3_Part1.pdf?__blob=publicationFile&v=1

Meesmann, U., Torfs, K. & Van den Berghe, W. (2019) ESRA2 methodology. ESRA2 report Nr. 1. ESRA project (E-Survey of Road users' Attitudes). Brussels, Belgium: Vias institute. https://www.esranet.eu/storage/minisites/esra-methodology-reportno1.pdf

Schrauth, B. & Funk, W. (in press) Key Performance Indicator für Alkohol – Entwicklung einer Methodik und Ersterhebung. Forschungsprogramm Straßenverkehrssicherheit FE 82.0758/2021. Schlussbericht zum 29. Juli 2022. Nürnberg: Institut für empirische Soziologie an der Friedrich-Alexander-Universität Erlangen-Nürnberg.

Schulze, H., Schumacher, M., Urmeew, R. & Auerbach, K. (2012) DRUID Final Report: Work performed, main results and recommendations. DRUID - Driving under the influence of alcohol, illicit drugs and medicines - Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006). Retrieved from: https://www.bast.de/Druid/EN/Dissemination/downloads_and_links/Final_Report.pdf?__blob=publicationFile&v=1

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

Yannis, G. & Folla K. (2022). Baseline report on the KPI Driving under the Influence of Alcohol. Baseline project, Brussels: Vias institute. https://baseline.vias.be/storage/minisites/baseline-kpi-alcohol.pdf

Appendix 1 SWD KPI Driving under the influence of Alcohol

Ref: Commission Staff Working Document - EU Road Safety Policy Framework 2021-2030 - Next steps towards "Vision Zero, SWD (2019) 238, https://transport.ec.europa.eu/system/files/2021-10/SWD2190283.pdf

Rationale

Driving under the effect of alcohol is frequently cited as a major collision causation factor.

Definition of the KPI

Percentage of drivers driving within the legal limit for blood alcohol content (BAC).

Minimum methodological requirements

Data collection method	Random breath testing. If random testing is not possible: Breath testing results from enforcement actions (even if not random) and / or Self-reported behaviour through anonymous surveys.
Road type coverage	Motorways, rural non-motorway roads, and urban roads to be covered.
Vehicle type	Passenger cars as a minimum; goods vehicles, buses and motorcycle if possible (results disaggregated by vehicle type).
Location	Random sample (methodology for Member States to decide).
Time of day	Any time for testing (not relevant for self-reporting).
Day of week	Separate results for weekdays and weekends.
Month	Late spring, early autumn.
Tolerance	Measurement instrument error.
Sampling methods	Random (methodology for Member States to decide).
Sample size	Member States to decide.

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. Based on feasibility issues in Baseline to reach the minimum 500 drivers' sample per stratum though, it is decided within Trendline to decrease the minimum driver sample for 'difficult strata' (based on the Baseline experience: for some countries one or both night periods, for other countries one or both weekend periods) to 250 drivers. It should be noted that this leads to bigger error margins for the point estimate though, e.g., (see also Table 2) for a KPI estimate of 88% based on a sample of 250 drivers the 95%-confidence interval is 84.97% - 92.03%, so ± 4 percent points (instead of ± 2.9 percent points with a sample of 500 drivers).

For many countries, the sampling according to the minimum sample sizes will imply disproportionate sampling of certain strata compared to the distribution of traffic volume 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 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[3] 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[4]. 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.

[3] 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).

[4] 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[5]. 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:

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

[5] 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:

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

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 4 Summary of random breath testing requirements and recommendations

SWD minimum requirements	Trendline minimum requirements for RBT	Trendline recommended options for RBT
KPI: % within legal BAC limit - Method: roadside survey with breath testing of randomly selected drivers - Road type: rural, urban, motorway - Vehicle type: min. cars, other if possible - Location: random - Time: any - Day: week, weekend - Tolerance: instrument error - Month: late spring, early autumn	KPI: % ≤ legal BAC limit + CI - aggregated - per road type - per week period - Collaboration with police (no voluntary participation) - 4 week periods: night/day x week/weekend - Exclusion of motorways if not available in road network - Expressways included in motorways, if considered - Location sampling: random and geographical coverage - Min. 10 different locations per road type (3) - Min. 10 locations per week period (4) - Min. 2 different locations per combination of road type and week period (12 crossed strata) - Min. sample size: 2,000 tested car drivers - Min. 500 drivers per road type (3) and per week period (4) - 1 location = min. 1 control session = min. 30 minutes - Free month choice but not during holidays or heavy winter period - Traffic counts during sessions (10 min) + estimates of road network length (3 types) for weighing data - data should be weighted	- Boost sample size (locations and drivers) for more accurate estimates and further (crossed) stratifications - KPIs according to BAC levels - Driver/ride characteristics: age (3), gender (2) - Complete disaggregated data - Other vehicle types (light, heavy goods vehicles, buses, motorcycles) - Exclusion of locations with <10 cars/hour is allowed - Region as sampling stratification (e.g., NUTS1; all min. sample sizes per region) – no KPIs for Trendline - Use official traffic volume data to sample locations and to weight data according to considered stratifications