

KPI Driving under the influence of drugs

Methodological guidelines

September 2025



KPI Driving under the influence of drugs. Methodological guidelines.

Please refer to this report as follows:

Kint, S.T. v d, Stelling, A., Van den Berghe, W., Mondésir, H., Corado, A., Areal, A., Torbay, A., Forsman, Å, Klipp, S., Folla; K. (2025). KPI Driving under the influence of drugs. Methodological guidelines. Report produced as part the Trendline project, supported by the European Union.

Project details:

Project start date:	15/10/2022
Duration:	36 months
Project name:	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:	1.1
Dissemination level:	Public
Date:	04/09/2025

Report Author(s):

Sander van der Kint; Agnieszka Stelling; Wouter Van den Berghe (SWOV), The Netherlands
Helga Mondésir (ONSR), France
Ana Coroado, Augusto Torbay (ANSR), Alain Aréal (PRP), Portugal
Åsa Forman (VTI), Sweden,
Simone Klipp (BAST), Germany
Katerina Folla (NTUA), Greece

Revision history

Date	Version	Description
26-06-2025	1.0	First version
04-09-2025	1.1	Final version published on the Trendline website

Legal Disclaimer

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

© 2023 by Trendline Consortium

Table of contents

Terms and definitions	iv
1. Introduction	1
1.1. Context	1
1.2. Purpose and background of this document	1
2. Scope	2
2.1. General principles	2
2.1.1. Random saliva testing by the police	2
2.1.2. Random saliva testing without help from the police	3
2.1.3. Self-reported behaviour	4
2.2. Vehicle types	5
2.3. Road types	6
2.3.1. Random saliva testing with or without help from the police	6
2.3.2. Self-reported behaviour	7
2.4. Week periods	7
2.4.1. Random saliva testing	8
2.4.2. Self-reported behaviour	8
2.5. Driver and trip characteristics	9
2.5.1. Random saliva testing	9
2.5.2. Self-reported behaviour	9
2.6. Region/state	10
2.6.1. Random saliva testing	10
2.6.2. Self-reported behaviour	10
2.7. Drug detection in saliva	10
3. Measurement procedure	12
3.1. Random saliva testing	12
3.1.1. First phase of sampling: sessions	12
3.1.2. Second phase of sampling: individuals	14
3.1.3. Minimum driver sample	15
3.1.4. Fieldwork set-up and procedure	17
3.1.5. Counting of traffic volumes	18
3.2. Self-reported behaviour	19
3.2.1. Sampling	19
3.2.2. Minimum sample sizes	21
3.2.3. Questions	21
3.3. Temporal considerations	23
4. Data analyses	25
4.1. Data coding	25
4.1.1. Random breath testing	25
4.1.2. Self-reported behaviour	26
4.2. Post stratification weights and statistical analysis	26
4.3. Expected results and data delivery	26
4.4. Metadata	28
References	29
Appendix 1 Rationale behind the minimum sample requirements	31

About Trendline

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

At the core of the Trendline project are the following 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, methodologies for some new experimental and complementary indicators have been developed and tested within Trendline:

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

For each of the original eight KPIs and the experimental KPIs, a 'KPI Expert Group' (KEG) was established, consisting of European experts. The main role of the KEGs was to draft the common methodological guidelines, to answer methodological questions, and to supervise the pilot tests of the new methodologies.

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 of 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 drugs. It describes three methods of measuring this KPI, defined as:

Percentage of drivers not driving under the influence of drugs

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. This document will closely align with the Methodological Guidelines for Alcohol (Boets et al., 2023) because the methodologies and data collection requirements for the KPIs on drugs and alcohol are highly similar. Given that driving under the influence encompasses both alcohol and drug use - and that their combination worsens impairment - it is advisable to combine the methodological guidelines for both; they are already largely consistent. For example, when a driver is stopped for an alcohol test, a drug test can be performed at the same time using the same procedures. Thus, to facilitate comparability the stratification variables used in the KPI alcohol have been maintained through this guideline.

2. Scope

2.1. General principles

For the KPI on driving under the influence of drugs, several approved methods of data collection are available, each allowing for a different balance of accuracy, feasibility and time investment. These methods are designed to ensure sufficient comparability between countries. The three proposed approaches are:

1. Random roadside saliva testing by the police;
2. Random roadside saliva testing without police help;
3. Self-reported behaviour.

Random roadside saliva testing done by the police provides the most accurate way to measure the prevalence of drug-impaired driving and is therefore the recommended method or “gold standard”. However, in some Trendline countries, this approach may not be feasible due to the demand of police resources (and possible associated costs), and legal constraints that prohibit random roadside testing without reasonable suspicion. In such cases, **random roadside saliva testing without police involvement** may be used. This method allows researchers to collect data independently, although it may suffer from limitations in terms of enforcement context, legal backing and driver compliance/cooperation. Additionally, drivers can be invited to complete a survey (recommended to use the self-reported behaviour survey discussed 2.1.3).

If neither of the above methods is feasible, the **minimum acceptable approach is self-reported behaviour**. This method is less accurate but still provides relevant insights in those Trendline countries where random roadside testing is not an option while having lower costs. The self-reported behaviour method will be very closely aligned with the method of self-report presented in the KPI Alcohol guidelines.

It is important to understand that these methodologies are not to be mixed within a single KPI. Each method must be reported separately to maintain comparability across data sources to ensure consistent and method-specific KPI's.

The following sections will describe each method in more detail, including their rationale, advantages and limitations.

2.1.1. Random saliva testing by the police

Random saliva testing is conducted during roadside drug checks by police officers, often in conjunction with alcohol testing. During routine patrols, officers randomly select drivers to stop and administer a saliva test. These tests indicate whether a drug is present in the driver's saliva, but do not measure the

drugs' concentration levels in the blood. The primary objective of this method is to estimate the KPI as: **Percentage of drivers not driving under the influence of drugs.**

Similar to the KPI alcohol, the theoretical population (100%) refers to all to all vehicle kilometres driven across the entire country. To make the study results representative, the data should be weighted based on traffic volume estimates under different conditions, known as stratification. For drug driving, the main strata (or factors) that are known to influence drug driving (and drink driving) prevalence include:

- Time of day;
- Day of the week;
- The interaction of Time of day and Day of the week;
- Road type.

At minimum, the sampling strata are road type (3 categories) and week period (4 categories), as will be outlined in sections 2.3 and 2.4.

By performing stratification the results can be weighted, ensuring that the final percentage reflects not just how many drivers follow the legal limit, but also how many kilometres are driven legally. Thus, the goal is to estimate what percentage of total driven kilometres happens within the legal drug limit.

Detailed instructions on how to apply this are presented in the Trendline statistical guidelines:

<https://trendlineproject.eu/media/pages/publications/32e9d29d7c-1719921954/statistical-analysis-of-kpi-data.pdf>

2.1.2. Random saliva testing without help from the police

Random roadside saliva testing **without** help from the police is done by researchers or research staff. Drivers can be randomly selected at gas stations and invited to participate in a study involving a short private questionnaire accompanied with a saliva test. The driver is then allowed to continue their journey. Using a quota sample is recommended to achieve a balanced sample based on age group and gender. Additionally, during this procedure, the driver can also be invited to complete a short survey. It is recommended to use the self-reported behaviour method outlined in the next paragraph (2.1.3) for this.

Using this method, the main strata are similar to those of random roadside saliva testing by the police:

- Time of day;
- Day of the week;
- The interaction of Time of day and Day of the week;
- Road type.

At minimum, the sampling strata are road type (3 categories) and week period (4 categories), as will be outlined in sections 2.3 and 2.4.

This method does suffer from the limitation that it involves voluntary participation which introduces bias. Drivers who have used drugs may decline participation. Therefore the prevalence rates should be interpreted as minimum estimates.

2.1.3. Self-reported behaviour

Data on self-reported drugged-driving behaviour can be collected using two types of questionnaire-based surveys: period-prevalence and trip-based prevalence surveys. These methodologies are extensively described in the *Methodological Guidelines for Alcohol* (Boets et al, 2023), which server as a key reference in the development of this KPI. The two survey methods are:

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

Period-prevalence surveys measure prevalence over a longer period, whereas trip-based prevalence measures the proportion of the population that has driven while under the influence of drugs during (one of) their last journeys.

Contrary to the Methodological guidelines Alcohol (Boets et al, 2023), currently the ESRA questionnaire does not provide follow-up questions other than the two questions presented in this paragraph. It is recommended to supplement these questions with some follow-up questions. Some examples are discussed in 2.3.2.

Period prevalence

Within Trendline it is determined that the period-prevalence measurement should **at minimum be in line with the ESRA₃ method for sampling, questioning/filtering and weighting** (Meesmann & Wardenier, 2024). The question from ESRA₃ measures the frequency of ‘*over the last 30 days, how often have you driven within 1 hour after taking drugs (other than prescribed or over the counter medication)*’ as a car driver. Responses are given on a 5 point rating scale where 1 = never and 5 = (almost) always. The KPI then refers to **“the percentage of car drivers indicating never to have driven in the last 30 days within 1 hour after taking drugs other than prescribed or over the counter medication”**. The **minimum required KPI is the national aggregate KPI**. Furthermore, it is highly recommended to also provide KPIs by age (3; 18-24; 25-64; 65+)) and gender (2). If the data allows, road type (3) and week period (4) would be very interesting to have (see sections 2.3 to 2.5).

Trip-based prevalence

The trip-based prevalence method is more alike random saliva testing due to the selection of the random trip. The guidelines for this method are based on previous studies (Diependaele & Silverans (2017), Diependaele (2015), Vollrath et al. (2019) and the German KPI alcohol study in Baseline (Schrauth & Funk, in press). The specific method described here is also adapted from the Methodological Guidelines for Alcohol (Boets et al., 2023).

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 under the influence of drugs during 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 information by age (3) and gender (2) (see section 2.5).

Representative sampling

For the self-reported methods described above, the data should be based on a selection of recent car drivers from the ESRA representative sample, similar to KPI alcohol (Boets et al., 2023). Without a representative sample, the studies are limited on generalization and suffer from selection bias. Both period-prevalence and trip-based prevalence have similar requirements to ensure representativeness.

For both period-prevalence and trip-based prevalence, the required sample has to be representative for the national population on age (18 or older), gender and region/state. ESRA currently calculates the KPI drug based on a subgroup of recent car drivers (last 30 days) from the representative sample.

Similar to the methodological guidelines for the KPI Alcohol, countries that are familiar to or that aim to use 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 sampling/weighting method and all information on representativity should be specified in the metadata.

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. Additionally, since this KPI is closely related to the KPI alcohol, using the same survey method for both KPI's is highly recommended.

Self-report methods can be administered using self-completing formats (on paper, online) or using interviews (face-to-face - e.g., roadside, telephone). This document focuses on the recommended approach: representative, anonymous online surveys.

The following paragraphs will outline the stratification requirements for the methods discussed above.

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 random saliva testing 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. Meaningful distinctions for goods vehicles are light goods vehicles (LGVs) and heavy goods vehicles (HGVs). Be mindful that including other vehicle types though requires a larger sample size for meaningful results per vehicle category (see section 3.1.3). Additionally, if different vehicle types are included, the data collection method should include 'vehicle type' as a new variable and have clear definitions. 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 random saliva testing 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 random saliva testing. For the period-prevalence survey provision of KPIs by road type is recommended, but is not included in ESRA₃. For the self-reported trip-based prevalence survey provision of KPIs by road type is minimum requirement.

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 saliva testing with or without help from the police

Random saliva testing (with or without help from the police) should be based on a random sample of locations on the 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.

Random saliva testing 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 for the KPI drugs, it is recommended to use similar follow-up questions like the ESRA₃ question for alcohol use 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. Currently, ESRA₃ does not contain follow-up questions for driving under the influence of drugs. This could hopefully be added in future ESRA surveys. If this is not the case, inspiration can be taken from the follow-up questions of alcohol:

'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

The question may be adapted into:

'You said that you have driven a car within 1 hour after taking drugs (other than prescribed or over the counter medication). 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 within 1 hour after taking drugs 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 random saliva testing. For self-reported trip-based prevalence the week periods can be 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 saliva testing

Random saliva testing either with or without police help should cover four-week periods: weekday, weeknight, weekend day and weekend night. 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 drug 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.

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 within 1 hour after taking drugs (*other than prescribed or over the counter medication*) in the last 30 days:

'You said that you have driven a car within 1 hour after taking drugs (other than prescribed or over the counter medication). 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 within 1 hour after taking drugs (other than prescribed or over the counter medication)*). 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 (and/or post-stratification weighting) based on population statistics for 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 the minimum requirement. For random saliva testing it is recommended to collect data on age and gender of the driver and to provide separate KPIs.

2.5.1. Random saliva testing

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

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 drug limits are in force for specific driver groups it is recommended to additionally provide differentiated KPIs (KPIs for private vs. professional drivers, 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 drugs 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 living) 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 random saliva testing (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 random saliva testing (18-24; 25-64; 65+) to be uniform and in line with the Baseline indicators.

2.6. Region/state

2.6.1. Random saliva testing

Disaggregation of results by region/state (where the driver is tested) is not required 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. Drug detection in saliva

This KPI is based on oral saliva testing. The KPI Alcohol in Baseline (and Trendline), is tied to legally defined thresholds and reflects illegal behaviour. Applying the same principle to the KPI Drugs would require blood tests or advanced saliva tests to determine whether drug use exceeds legal limits. However, such procedures demand significant police resources that may not be available in all Trendline countries. Oral fluid tests thus offer a practical alternative: They are easy to administer and provide results within a couple of minutes, making them more practical for data collection compared to blood sample testing or advanced saliva analysis. However, interpreting the results of oral fluid tests can be ambiguous. Countries use different types of drug tests, each with its own cut-off levels. Therefore, it's essential to document the specific test used in the accompanying metadata. Additionally, to ensure accurate comparisons and consistency for the KPI drugs, it is recommended to monitor developments in saliva testing technology. Additionally, unlike alcohol testers, saliva tests can measure multiple drugs. Drivers can test positive on more than one substance at the same time, which should be reflected in the meta-data.

For future research and refinement of these guidelines, more insight is needed in the specific saliva testers used by Trendline countries. Drug tests vary between countries, and some countries use multiple types of drugs tests. Currently there is no comprehensive list detailing all the drugs detected by saliva tests for all Trendline countries.

However, a list of drugs that are most likely being measured by all drug tests across Trendline countries (and associated risks) is presented in Table 1. These categories reflect the risk they pose for crashes and injuries. In all cases, higher doses of drugs reflect an increase in crash risk. It is highly recommended to provide the results for these commonly found drugs separately.

Table 1: Risk increase caused by driving under the influence: results from meta-analyses. Adapted from SWOV factsheet drugs and medicines (SWOV, 2020).

Drug	Crash Severity	Risk Increase (Odds Ratio)	95% Confidence Interval	Significance
Amphetamines	Fatal [Elvik, 2013]	5.2	(2.6 – 10.4)	
	Injuries [Elvik, 2013]	6.2	(3.5 – 11.1)	
Cannabis	Fatal [Elvik, 2013]	1.3	(0.9 – 1.8)	Not significant
	Injuries [Elvik, 2013]	1.1	(0.9 – 1.4)	Not significant
	Fatal & Injuries [Rogeberg & Elvik, 2016]	1.4	(1.1 – 1.6)	
	Fatal & Injuries [Rogeberg & Elvik, 2016]	2.5	(1.7 – 3.7)	
	Fatal & Injuries [Rogeberg, 2019]	1.3	(1.2 – 1.4)	
Cocaine	Fatal [Elvik, 2013]	3.0	(1.2 – 7.4)	
	Injuries [Elvik, 2013]	1.7	(0.9 – 3.0)	
Opiates	Fatal [Elvik, 2013]	1.7	(1.0 – 2.8)	
	Injuries [Elvik, 2013]	1.9	(1.5 – 2.4)	
Multiple Drugs	Crashes [Hels et al., 2011]	5 – 30	–	
Alcohol + Drugs	Crashes [Hels et al., 2011]	20 – 200	–	

3. Measurement procedure

3.1. Random saliva testing

3.1.1. First phase of sampling: sessions

Random saliva 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 drugs (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. Determine the required number of locations for the country or per region.
2. Randomly select a number of locations on a map using the entire area in determined in the previous step, without them being too close to each other (i.e. spatially constrained random sampling). The feasibility to set-up a drugs check is not of concern at this point yet. This step is to ensure a reasonable geographical spread of the randomly selected locations.
3. Manually select the observations of the locations randomly selected in the previous check to perform the random saliva testing. This selection must be based on the location requirements (different road types), inclusion/exclusion criteria (if applicable) and practical considerations of performing random saliva testing. This final selection can be made using Google Street View or in cooperation with the police unit responsible for the respective location.

A convenient way of selecting locations randomly (i.e. step 2) is to use a GIS system (e.g., cartographic software like ARCVIEW/ARCGIS or Qgis which is open and free of charge) 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 drug 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 used drugs. On motorways, checks 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, sport clubs etc). 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 traffic flow proves to be 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 reselect a different location.

In the next step the **drug 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 not sampled proportionately to traffic volume), 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 1 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 randomly selected, irrespective of any suspicion for DUI of drugs. Any selectivity, either in the locations chosen (e.g., risk areas for drug driving) or in who is being checked and who is not, leads to a bias and decreases the representativeness of the data. Random saliva testing 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.

In some countries random saliva testing is also not allowed for the police and they can only stop drivers in case of certain suspicious signs. In some countries random saliva testing by researchers can be done. However, saliva testing based on voluntary participation is methodologically difficult to get a representative sample. Voluntary participation leads to a self-selection or participation bias. It is therefore highly recommended to collaborate with police forces which can randomly stop drivers and test each stopped driver.

Random saliva testing on a random sample of locations can mean for police forces that they must adapt their normal way of working. Often drug (and 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 sampling¹), which can be considered a sufficient accuracy. **Minimum 2,000 car drivers² is the required sample for the national KPI for DUI of illicit drugs (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.³

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

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

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 If it proves difficult to undertake the minimum required number of tests (2000 in total) over a period of 4 weeks, It can be considered to spread the measurements over a longer period (e.g. 8 consecutive weeks, or an additional 4 weeks in the next year) provided one can assume that the context and environment has not changed considerably between the measurement sessions.

³ Calculations can also be done via: <https://sample-size.net/confidence-interval-proportion/>

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).

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 exhibit notable low compliance with legal drug-driving regulations, (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 $3 \times 4 = 12$ 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 drug checks. If drivers are stopped randomly (also irrespective of the vehicle type) during the drug 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 1 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 drug 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) random saliva testing measuring devices (e.g., Dräger, Honac, Securetec). After a positive test the normal legal/sanctioning procedures follow.

Each location corresponds to one or more (e.g., different week periods) drug 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 random saliva testing, 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 cannot 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.

Police checks on high-speed roads should always be carried out in accordance with applicable (road) safety regulations. For safety and feasibility reasons random saliva testing 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 drug check.

The required data is coded as an additional task by the police or by a research worker assisting the police check. 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 saliva 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 drugs (or 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. The counting can be done by a research worker or by one of the policemen. Ideally all passing (including stopped) cars at the drug 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 random saliva testing (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 this KPI.

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) * 3 age groups (18-24, 25-64, 65+) 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 random saliva testing 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 compared to random saliva testing because trips take place over a number of roads. 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 drugs. 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). It is recommended that the question on driving over the legal limit for drug 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 within 1 hour after taking drugs (other than prescribed or over the counter medication)
(recommended)

If the answer was at least once on the first question, it is followed by:

You said that you have driven a car within 1 hour after taking drugs (other than prescribed or over the counter medication). 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*

You said that you have driven a car within 1 hour after taking drugs (other than prescribed or over the counter medication). Was this ...? You can indicate multiple answers:

Cannabis - cocaine - amphetamines (e.g., speed, ecstasy) - illicit opiates (e.g., morphine, codeine; not prescribed as medication) - other

As the ESRA3 question referring to "driven within 1 hour after taking drugs (other than prescribed or over the counter medication)" is closest to the KPI definition this is currently recommended.

The KPI would thus reflect the **percentage of car drivers indicating 'never' to have driven within 1 hour after taking drugs (other than prescribed or over the counter medication) 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 drugs is **'Do you think that you were driving within 1 hour of having used drugs (other than prescription or over the counter medication) during this trip?'** If a respondent answers affirmatively, it is strongly recommended to follow up with a question to identify the specific drug(s) used. This question can mirror the approach used in the period prevalence method: "You said that you have driven a car within 1 hour after taking drugs (other than prescribed or over the counter medication). Was this ...? cannabis - cocaine - amphetamines (e.g., speed, ecstasy) - illicit opiates (e.g., morphine, codeine; not prescribed as medication) - other

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 overall KPI thus reflects the **percentage of car drivers indicating not to have driven when they may have been under the influence of an illicit drug during a random recent trip.**

Ultimately, it is understandable that some countries already have their own preferred and established method for setting the parameters of these questions. For example, using a recall period of one year instead of 30 days, or defining the time window as two hours instead of one. In that case it is possible to differ from the proposed parameters. However, while it is not mandatory to adhere strictly to the proposed parameters, it is essential that the core intent of the questions remains comparable across countries. That would insure the validity of the KPI.

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 drugs 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 drugs.

These guidelines and restriction apply to random saliva testing 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:
 - **binary drug test result (2) (drug detected yes/no)**
 - binary drug test result for multiple drugs (multiple drugs detected yes/no)
 - type of drug(s) detected, separately for the most commonly found drugs (see 2.7)
 - driver type (private, professional)
 - driver experience (novice or not)
- Session level:
 - **road type (3)**
 - **week period (4)**
 - **date**
 - start hour
 - end hour
 - total duration of the drug 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:
 - driver age category (highly recommended)
 - driver gender (highly recommended)
 - vehicle type (if other than cars are considered)
- Session level:
 - region/state
 - police zone
 - weather condition

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:

- **age**
- **gender**
- **region/state (domicile)**
- response to the KPI drug question for car drivers: 5 point rating scale (never to (almost) always)
- **dichotomized response on the KPI drug 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 drug 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 the Trendline statistical guidelines: <https://trendlineproject.eu/media/pages/publications/32egd29d7c-1719921954/statistical-analysis-of-kpi-data.pdf>

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 random saliva testing the **main indicator is the percentage of car driver not under the influence of illicit drugs 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 single drug use vs multiple drug use (2 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 not to have driving within 1 hour after using drugs during 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 within 1 hour of using drugs of 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 drug 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 (random saliva testing, self-report: trip-based, period) including also:
 - random saliva testing: collaboration with police or not
 - random saliva testing: type of test used
 - 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:
 - random saliva testing: 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
 - random saliva testing: 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 post-stratification 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 drug limits 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/deliverables-list/downloads/deliverable_2_1_2.pdf;jsessionid=2F12FF52575D8C1057FCCCEoFFA509A9.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. (2017) 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/SIaperig%20achter%20het%20stuur%202017/Sleepy_at_the_wheel.pdf
- Elvik, R. (2013). Risk of road accident associated with the use of drugs: A systematic review and meta-analysis of evidence from epidemiological studies. In: Accident Analysis & Prevention, vol. 60, p. 254-267.
- 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/fixe/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
- Hels, T., Bernhoft, I.M., Lyckegaard, A., Houwing, S., et al. (2011). Risk of injury by driving with alcohol and other drugs. Deliverable D2.3.5 of the EU FP6 project DRUID. European Commission, Brussels.
- 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/deliverables-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>
- Meesmann, U., & Wardenier, N. (2024). ESRA3 methodology. ESRA3 Thematic report Nr. 1. Version 1.0. ESRA project (E-Survey of Road users' Attitudes). (2024 – R – 09 – EN). Vias institute. <https://www.esranet.eu/storage/minisites/esra3-methodology-report.pdf>
- Rogeberg, O. (2019). A meta-analysis of the crash risk of cannabis-positive drivers in culpability studies—Avoiding interpretational bias. In: Accident Analysis & Prevention, vol. 123, p. 69-78.
- Rogeberg, O. & Elvik, R. (2016). The effects of cannabis intoxication on motor vehicle collision revisited and revised. In: Addiction, vol. 111, nr. 8, p. 1348-1359.
- 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
- SWOV (2020). Drugs and medicines. SWOV fact sheet, March 2020 SWOV, The Hague.
- 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 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.