Programme Support Action (PSA)
to support Member States
in collecting Key Performance Indicators (KPIs)
for road safety

Grant Agreement No. MOVE/C2/SUB/2019-558/CEF/PSA/SI2.835753
KPIs collection for Road Safety

Baseline conclusions and recommendations

January 2023

Belgium | Austria | Bulgaria | Cyprus | Czech Republic | Finland | Germany | Greece | Ireland | Latvia | Lithuania | Luxembourg | Malta | Netherlands | Poland | Portugal | Spain | Sweden

Baseline.vias.be

Co-financed by the Connecting Europe Facility of the European Union

Project:	This document has been prepared in the framework of the Baseline project, for which a grant has been awarded by the European Commission. Information on this project can be found on the website www.Baseline.vias.be
References:	(1) Grant agreement under the Connecting Europe Facility (CEF) No MOVE/C2/SUB/2019-558/CEF/PSA/SI2.835753 collection of Key Performance Indicators (KPIs) for road safety (2) Consortium agreement among the 18 partners of the Baseline project
Authors:	Silverans, P. , & Vanhove, S. (Vias institute, Belgium). Baseline conclusions and recommendations. Baseline project, Brussels: Vias institute.
Publisher:	Vias institute, Brussels, Belgium (www.vias.be)

Please refer to this document as follows:

"*Silverans, P., & Vanhove, Sophie (2023).* Baseline conclusions and recommendations. *Baseline project, Brussels: Vias institute.*"

Disclaimer:	This report only reflects only the author's view. The Commission is not responsible for any use that may be made of the information it contains.

Any comments or feedback regarding this report should be sent to Baseline@vias.be.

Contents

1 Introduction

In 2019 the European Commission elaborated a list of Key Performance Indicators for road safety, further called 'KPIs'. These KPIs are defined and discussed in the Staff Working Document 'EU Road Safety Policy Framework 2021-2030' (SWD (2019) 283 final, Annex 1). The KPIs relate to the main road safety challenges to be tackled and are listed in *Table 1* below. The aim of using the KPIs is to monitor the trends in factors that contribute to reaching the EC targets in road safety.

Table 1. List of KPIs and their definition

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

In the report introduction to Baseline KPI reports (Silverans & Vanhove, 2023), the purpose and methodological approaches to each of the KPIs are described. The detailed methodology, results and discussion for each of the KPIs are included in 8 dedicated Baseline KPI reports.

In this report, the overall conclusions and recommendations for future KPI measurements are discussed. Conclusions and recommendations for the use of the KPI results in road safety policy are discussed in a separate Baseline policy report (Vanhove, 2023).

2 Conclusions on baseline KPI results

2.1 Benchmarking and international comparisons

2.1.1 Strong variations

For all KPIs, the KPI estimates at both aggregate and semi aggregate level show large international variations across countries. For example, for speeding, estimates range from only 40 percent of car drivers on weekdays respecting the speed limit on motorways in one country to 90 percent of car drivers driving within the limit in similar circumstances in another country (van den Broek, 2023, Figure 1). Another example is that the percentage of drivers of passenger cars not holding an electronic device ranged from 90.6 percent in one country to 98.3 percent in the best performing country (Boets, 2022, Figure 2). Percentages of drivers in passenger cars correctly fasted with a seatbelt range from 71 percent to 98 percent.

Since the main objective of the Baseline project was to measure road safety indicators in a harmonized, internationally comparable way, all resources in the project were invested in developing an adequate methodology that facilitated international comparability. Consequently, no supplementary contextual data were collected in a systematic way to interpret the observed international variation and to look into factors that might contribute to the observed differences. In the next phase (the Trendline project, cf. infra), an attempt should be made to identify potential explaining factors. Comparing performance of the least and best performing countries would be a good starting point. Exploiting expert advice on the different measures should also be considered.

2.1.2 Deviations from the standard methodology

The Baseline methodology was built upon expertise and methods used in countries with large historical series of KPI measurements. However, such methods differed significantly between countries. In order to take into account such international variations in methodologies and to increase the feasibility of implementing the methodologies in countries with little or no experience, some freedom was left to Member States in applying the methodologies at national level. Although no systematic relation could be identified between specific methodological choices and the resulting KPI estimates, in the future comparability across countries would increase if the methodological options at country level would be more limited. Options to consider in this regard are discussed in detail in the discussion of the methodological guidelines.

2.1.3 Contextual data

Since several KPIs refer to the percentage of drivers / road users respecting the applicable traffic rules, it is obvious there is a relation between the severity or level of tolerance allowed by internationally varying traffic rules and KPI performance. In the case of speeding, it is clear that countries setting maximum speed limits of the same type of roads at a higher level, will make it easier to comply, and hence result in better performance. Put simply, when a similar road has a speed limit of 70 km/h in one country and 100 km/h in another, the percentage of drivers complying will very likely be higher in the 100 km/h country. In this case, better KPI performance might correlate with worse overall road safety performance. A similar issue applies to countries with very lenient rules for child restraint system use or - mutatis mutandis - to countries applying a higher alcohol limit than others.

2.1.4 Relation between KPI performance and road safety outcomes

The relationship between KPI performance and overall road safety performance (e.g., the number of fatalities per million population) is still to be analysed in depth. At behavioural level, enough studies are available on population attributable risk, allowing to assess the theoretical impact of specific at risk behaviours on road safety outcomes (e.g. Martensen & Daniels, 2020). With the international rankings per KPI now available, it should be verified if the expected relationship based on attributable risk is reflected in the correlation between KPI performance and road safety performance. Contextual differences in traffic regulations should be included as a mediating factor.

2.1.5 Relations with other methods for obtaining performance indicators

The focus of the Baseline methodologies for the behaviour related KPIs is on direct observation and/or measurement in moving traffic. Other approaches to estimating the prevalence of infractions of specific traffic rules are possible. The most commonly used methodology to estimate the percentage of road users that "sometimes" (c.q. often, very often, etc...) commit traffic infractions are questionnaire surveys. For international comparisons, the ESRA survey series is an interesting framework on self-reported infractions (and a multitude of road user attitudes, opinion, beliefs and norms). Several KPI reports directly refer to self-reported ESRA data, but a further exploration of the correlations between baseline KPI indicators and the self-reported ESRA indicators can shed light on the relationships between both and on possible factors mediating the relationship between objectively measured performance and self-reported data.

2.2 Assessment of safety levels and future target setting

The Commission staff working document mentions that the purpose of measuring KPIs in different areas is to consider target setting with the aim of attaining the objectives of halving the number of fatalities and injuries. From a theoretical point of view, deliberate unsafe behaviour (like drink driving, not wearing seatbelts, excess speeding etc...) should be entirely banned from traffic in a vision zero policy.

Although sometimes relatively small percentages of non-compliant drivers are observed (e.g. seatbelt wearing in the best performing countries, drink driving on weekdays in the best performing countries etc...) this often looks better than the underlying reality is. Since the aim of the KPI estimates is to be representative for the entire traffic volume in each member state, KPI results reflect the average percentage (or share) of kilometres driven per country with or without compliance to traffic rules. Realizing that the total kilometres driven are hundreds of billions of kilometres per year, this implies that with a 1 percent share of infractions, that yearly one or more billions of kilometres are driven while being infraction. Therefore, efforts are always warranted, even with relatively small shares of infractions. When building a strategy, efforts should take both safety impact and overall KPI performance into account.

2.3 Monitoring and policy relevance

Apart from allowing international benchmarking, the purpose of estimating key performance indicators is to monitor changes in performance over time. This explains why the project was called 'Baseline”: it had the intention to make available baseline measurement to evaluate trends over time. In many Member States monitoring KPI performance over time has been going on for several decades. For some KPIs and target groups this showed improvements over time (e.g. percentage correctly wearing seat belts). For other KPIs relatively little progress (or stagnation) of performance over many years can be observed (e.g. drink driving).

Benchmarking and historical monitoring are key tools in road safety management. Therefore, a dedicated report on the use of key performance indicators at policy level is an integral part of the baseline project. A detailed discussion of use of KPI measurements in road safety policy is included in a separate Baseline policy report (cf . Vanhove et al., 2023).

3 Recommendations for the adaptation of the methodological guidelines

3.1 KPI definitions and specifications

The KPI specifications of the EU road safety policy framework 2021-2030 are described in general terms. To be able to collect data in an internationally comparable way, the specifications were translated into operational requirements. Examples of this are the concepts of "free flowing traffic", "in moving traffic", "correctly wearing seatbelts" etc... Such terms may seem to be specific enough for everyday use, but from a scientific point of view these concepts still need further clarification. Following good scientific practice, operations should be described in such a way that exact replication of the research activity becomes possible, also by new researchers. A good example of this is the concept of free flow speed, for which SWD 283 states *"... Member States should elaborate the indicator for day hours in free-flow traffic (...) Measurements should not take place near speed cameras, neither fixed nor mobile"*. In order to specify this, the methodological guidelines for speed gave a general guideline: *"All places where vehicles are likely to stop, accelerate or brake should be avoided, since at these locations free flowing traffic cannot be guaranteed"* and provided specific operational details on all aspects to be taken into account. The same principles were followed in the methodological guidelines on distraction (Boets et al., 2021), which refer in general terms to *"Observations should be made in flowing traffic only, so of drivers while driving, since distraction behaviour is different when stationary, e.g. waiting at traffic lights. No observation should be made of stationary drivers"* and specify *"ideally no locations should be chosen in front of traffic lights. Observation can take place near intersections but only drivers who are driving should be observed, not drivers who are stationary."* Similarly, the KPI on seat belt refers to the percentage of vehicle occupants wearing seatbelt or child restraint systems correctly, specifying correctness of use in terms of *"Possible misuses of safety belts are (non-exhaustive): Belt behind the back, Belt under arm, Incorrect height setting of seat belt's top guidance, and Use of 'foreign objects' such as clothespins to deviate the seat belt or reduce its tension".*

Detailed methodological guidelines were published to give detailed specifications on as many issues as possible. The activities undertaken throughout the Baseline project learned however, that despite the attempts to make definitions and specifications sufficiently clear, they weren't always interpreted in the same way everywhere. Examples of this are "front seat car occupants", which was limited to front seat passengers in some measurements and to the driver only in other countries. Another example are references to the period of observation in terms of day of the week or time of day. Although "in day hours" seems specific, this does not specify how to code and analyse twilight or dawn hours. Similarly, weekend nights might or might not include Friday evenings. Sometimes early Monday morning hours are also included in the definition of weekend. The differences in definitions of weekday and weekend resulted in ambiguity for interpreting post-crash intervention times. Terminological confusion also arose in relation to the concepts of urban and rural roads, which is used as a framework in SWD 283. For practical considerations urban versus rural was specified as within built-up area versus outside built-up area in several guidelines. Other examples of interpretation problems can be found in the conclusions on data quality and comparability in all KPI reports.

Most KPI definitions in SWD 283 are specific with regard to the final estimates concerned. An exception to this is the KPI on road infrastructure, which only refers to "*Percentage of distance driven over roads with a safety rating above an agreed threshold (still to be defined) leaving the rating methodology to the choice of Member States until an agreement on the threshold is reached*." The experimental work on the infrastructure KPI in Baseline took this into account by allowing 4 different KPI estimates, referring to either road network length or traffic volumes and either to safety assessments or safety characteristics of the road design (cf. Van den Berghe, 2023). In the meantime, within the framework of the EU Directive 2019/1936 EGRIS - DG Move (Study on a EU Methodology for Network-wide Road Infrastructure Safety Assessment), a new methodology was defined in order to assess road infrastructure safety. Directive 2008/96/EC on road infrastructure safety management, as amended by Directive (EU) 2019/1936, provides that, by the end of 2024, the Member States should do the first Network Wide Road Safety Assessment, that will reveal what are the most dangerous road sections to be addressed with targeted inspections and remedial measures. The model developed considers both the specific features of the road, that can predict the risk of accident, as well as the accidents' statistics (European Commission, 2022). Future work on KPIs for road infrastructure will build upon this framework to propose a common definition and methodology for assessing international variations in road safety.

3.2 Degree of risk

Within the baseline project the main focus was on estimating the percentage of drivers complying with road traffic regulations (or other specific risk behaviour, not necessarily imposed by a legal framework, such as in countries not mandating helmets for pedal cyclists). Only in the context of speed, indicators on the severity or degree of infraction was defined (c.q. averages speed and V85 speed, cf. *van den Broek B. & Aarts, L.* (2022). For other KPIs (e.g. drink driving, cf. Yannis & Folla et al, 2022), it would also be interesting to include additional KPIs taking severity or risk into account (e.g. alcohol intoxication levels). Severity also plays a role in seatbelt and helmet use KPIs. The seatbelt directly refers to correct use (which could theoretically be classified into different levels of severity). Helmet used by both cyclists and motorcyclists might also be of different make and model and not always be fastened correctly. It could be considered to take these factors into account as well.

Although defined at a different level, the KPI on post-crash intervention times did take into account a quantitative estimate of average response times by defining the key KPI in terms of a 95th percentile of the intervention time (cf. Nuyttens et al., 2022). The same goes for the KPI on vehicle safety, for which both 4- and 5-star Euro NCAP rating KPI estimates were calculated for all newly registered vehicles (cf. Wardenier et al., 2022).

3.3 Reliability of measurement methods

3.3.1 Observation by human observers

The five behavioural KPIs in Baseline rely mostly on coding by human observers by the side of the road. The methodological guidelines define the parameters that should be included in the observations. Firstly at the level of mandatory predictors to be included in all studies (e.g. vehicle type, day of week, etc...). Secondly, optional, interesting parameters that are interesting for that particular KPI (e.g. gender, estimated age group, ...). Within Baseline, the coding of the variables observed during the observations was left at the discretion of the member states. Some countries used computer (c.q. tablet or cell phone) based coding, others paper forms. Provided the definitions for observations are defined clearly and specifically, both methods can yield comparable levels of quality. A point of attention is the data treatment of unclear or uncertain observations (where the observer cannot categorize parameters with certainty). For these cases supplementary procedural definitions are required.

Quality assurance procedures are described for outlier treatment, evaluation of data homogeneity and other quality assurance procedures at the level of database management (cf. Van den Broek et al., 2021). No procedures were defined on the training of observers, evaluation of tenders for subcontracting of the field work, double checking the application of procedures in the field etc... Supplementary quality assurance procedures at this level could increase quality of the field work and the resulting KPI estimates.

3.3.2 Camera observations

Although camera footage was allowed according to the methodological guidelines for behavioural studies, only for the KPI on distraction two countries relied on camera images for the observations. Lessons learned from these experiences are that "*For analysing pictures high quality cameras, directed at a proper angle and preventing light reflection are crucial. Otherwise, it can be difficult to differentiate using a phone with other types of gestures such as touching the hair or glasses. Using pictures is in some MS not possible though due to GDPR issues*". (cf. Boets at al., 2022). The annex of the report describes requirements for automatic detection via "unintelligent", partly intelligent and fully intelligent camera systems. The experience in Finland with camera observations also includes recommendations to optimize the use of camera images and the coding of the images. Using camera observations rather than human observations should be carefully considered given the challenges at the level of encoding and categorizing data (partly related to resolution, dynamic range and lightning conditions, partly related to reliance on human observers to categorize data and on artificial intelligence systems used for pre-coding data).

When license plate information or facial images are part of the camera observations, specific attention is required to assure compliance with general data protection regulations (GDPR). For that reason, cameras can be positioned in order to exclude images of the drivers' faces in some instances. Whether or not facial images are part of the recordings, a full review of the national and international regulations will be necessary. In general, the process will require specific authorisation from responsible authorities for most countries. When the necessary authorisations and protocols are not possible to define/attain, other observations methods will be required.

3.3.3 Measurement tools

Different types of radar and loop detector technology qualify for measuring (spot) speeds of vehicles, including manually controlled speed gun technology. The methodological guidelines describe the respective characteristics of all available technologies (Teuchies, 2020). From Table 3 of the KPI report, it is clear that a wide range of technologies were used across countries. Tolerance margins were explicitly not to be taken into account for the KPI estimate (cf. SWD 283). Although the reliability of the speed measurements is adequate, the methodological requirements acknowledge that the estimates of vehicle length used to distinguish for instance personal cars from large vehicles are less reliable and only allow an approximate categorization. Further research is needed on the possible issues with categorizing vehicle type based on the lengths measured during the observations.

For the sake of completeness, or the KPI on drink driving, breathalyser test and analysers were used in collaboration with police forces. Within this framework measurement reliability is assured. But collaborating with police force to randomly sample drivers from moving traffic poses particular challenges. Since police enforcement activities are often deliberately select and partly based on suspicion it is necessary to evaluate if sampling is indeed random and the check whether there are indications (sometimes at session level) that randomness has not always been assured (cf. Yannis & Folla, 2022).

3.3.4 Questionnaire surveys

The only KPI for which questionnaire survey data could be used to estimate the KPI was drink driving, for which self-reported data relative to either specific points in time or relative to larger periods of time qualified. Although questionnaire data can be used to follow evolutions in self-reports over time, the resulting KPI estimates are incomparable with objectively measured alcohol concentrations in roadside surveys. As acknowledged in the KPI report, it should be considered to require roadside surveys conducted by the police as a minimal requirement, except for countries in which the police is not allowed to test randomly. In such countries roadside testing based on voluntary participation could be considered.

In parallel to roadside surveys, questionnaire data (such as the ESRA survey series, cf. infra) offer ample opportunities to analyse predictors of specific at-risk behaviours and explore correlations with external data (enforcement, social norms, ...). In some observational KPI studies, questionnaire data were used to complement data collection and analysis. This was for instance the case in in-vehicle inspections of child restraint systems. In some countries, random data collected by the police were also complemented with questionnaire data on trip length and origin.

3.3.5 Database linking and analysis

The KPI estimates for infrastructure, vehicle safety and post-crash care are based on the exploitation of existing databases. Although exploiting pre-existing datasets does not require a specific measurement method, the exploitability of the data largely depend on the structure and details of the data included. In order to assess the Euro NCAP level of newly registered cars, several Member States were faced with difficulties in linking national vehicle databases to the Euro NCAP database because of different data encoding strategies for model and mark of vehicle. Post-crash data contained issues with the unit of measurement of response times (sometimes provided only at the accuracy level of minutes rather than seconds) but also for the description of predictors included in the analysis (e.g. lack of homogeneity in definitions of week/weekend etc...). Specific recommendations to increase comparability in KPI estimates based on different database structure are included in the respective KPI reports.

3.4 Identification of appropriate locations for observations

KPI estimates based on observation and/or measurement are mostly performed at randomly selected locations to obtain a representative estimate for the entire traffic volume in a region or country. The different methodological guidelines took recommendations from earlier EU projects into account (cf. Hakktert & Gitelman, 2007: SafetyNet general recommendations for safety performance indicators for an overview):

The basic procedure for location sampling is described in the guidelines for the KPI on alcohol as (Yannis & Folla et al., 2021 and Boets et al., 2021;):

(1) Define the required number of different locations (for the country or per region)

(2) Random selection of locations on a map, taking sufficient geographical spread into account

(3) The final locations for the observations are chosen in the area surrounding the locations randomly selected based on the location requirements (different road types), inclusion/exclusion criteria (if applicable) and practical considerations.

The sampling scheme also took time period and road type into account for most measurements. Details on the distribution of location over the different strata in each study design are discussed in the next paragraph.

According to the description of the methodologies for selecting locations provided in the meta-data in the baseline database, random sampling procedures were used most of the time, although practical considerations were also taken into account. Specifically for low volume roads, it was agreed that low traffic volume roads were less than 10 relevant vehicles passing by per hour were to be expected were excluded. This requires cautions, since this limitation can impact traffic volume estimates derived from traffic counts during the observations.

3.5 Stratification and minimal sampling size

For most KPIs, a minimal number of locations was required for specific strata as a minimal requirement. A minimal number of observations per type of road or time period was also required for several KPIs (e.g., Yannis & Folla, 2020). In practice, it appeared difficult for Member States to comply with all minimal requirements. As illustrated in the report on the KPI for alcohol, this causes uncertainty on the comparability of results between Member States. For alcohol measurements, it was recommended to sample the locations for the three road types proportionally to traffic volume (and therefore proportionally to the kilometres driven on each road type in the country or region, cf. Yannis & Folla et al., 2021), assuming that each of the three road types represent a share of traffic volume above 20% of the total traffic volume, based on available national traffic data (e.g. nationally representative traffic/mobility surveys). It was also recommended to carry out location sampling proportional to traffic volumes by time period, and ideally crossed with road types.

The Baseline methodological guidelines defined the following minimal sample size for the KPI on speeding (Teuchies, 2020):

The minimal required sample sizes to provide a KPI are:
- *min. 10 locations per road type and stratum*
- *min. 500 observations per road type and stratum*
- *min. 2000 observations in total*
- *the proportion of observations at each of the three road types should be at a minimum of 20%*

These minima were chosen to assure sufficient accuracy of semi aggregate KPI estimates in terms of confidence interval width. This general rule for minimal sample size was also the guiding principle for other KPI estimates based on sampling (e.g. seatbelt wearing, helmet wearing, distraction,...). Apart from issues with insufficient sample sizes in particular strata mentioned above, the confidence intervals reported confirmed the overall approach is adequate. It should however be stressed that the minimal sample sizes only reflect the minimal requirement. If specific strata show particularly low levels of compliance (such as night time drink driving) it should be considered to increase minimal sample size for highly problematic strata in order to obtain a more detailed view of international differences and predictor variables.

Several countries experienced problems in randomly sampling categories of road users that represent a (very) low share of the total traffic volume in a country. This was the case for cycling in some Southern

European countries and for motorcycle use in some Northern countries. Similarly, low numbers of passengers on two wheelers (motorcyclists and cyclists) were included in the database (cf. Yannis & Folla, 2020). In general, and as expected, small numbers of truck and bus drivers were observed in all observations taking this category into account. This shows that although a particular target group may be part of the KPI definition, it is not always possible to impose minimal sample sizes that assure sufficiently narrow confidence levels at sub aggregate level. Methodological requirements should be revised in order to take expected sampling of rare events into account and ad hoc procedures and requirements should be defined for those particular cases.

3.6 Statistical weighting

Statistically weighting of the data collected is necessary in order to estimate a representative KPI value for the total of all kilometres driven (traffic volume) in a country. More specifically, for the total distance driven in a country “by the vehicle/road user type(s) on the road types and within the time periods that are required in the KPI measurement”. In terms of sampling this means that the population is the total traffic volume (in kilometres) of moving vehicles over a certain area (i.e. country or region) over a certain period of time (e.g. one year). Estimates are made by sampling individual vehicles (or road users) at particular locations and moments in time. Hence the question arises as to how each of these individual observations have to be weighed in such a way that the overall average or percentage reflects the overall percentage of vehicles confirming to the rules in the total population. This is a classical problem in sampling from moving populations (e.g. Hakkert & Gitelman, 2007).

Several countries delivered unweighted KPI estimates for different KPIs. This can only be judged appropriate when a proportional sampling scheme respecting all other minimal requirements was used. In practice, practical considerations always need to be taken into account, due to which sampling is always to a certain degree disproportional. Unweighted results, however, cannot be compared with KPI estimates weighted according to traffic volume. This also underlines the importance of defining minimal requirements for weighting of the data in future KPI work.

One possible method to assure representativity is to sample observations strictly proportional to traffic volume in each of the relevant strata. In practice, this is rarely possibly since one can a priori define locations and observations sessions, but it is not realistic to define the number of observations made a priori. For feasibility reasons, numbers of observation sessions are also often based on the minimal required number, rather than on the traffic volume share of particular strata. Consequently, in such disproportional sampling schemes, the number of observations per stratum is not proportional to the traffic volume in each stratum. This disproportionality can be corrected on the basis of external traffic volume information per stratum. When this information is not available, a possible approach is to estimate traffic volume shares based on the traffic intensity counted during the observations in combination with information on the length of the road network in each stratum. In the considerations for sampling weights a general approach to weighting is described (cf. Silverans & Boets, 2020). In practice, implementing the proposed formula appeared to be difficult due to unclarity on the exact calculation methods. Combining estimates of network length in combination with observed traffic intensity cannot be simply combined with a priori traffic volume since the ratio of observed vehicles to traffic volume has to take other weighting factors into account. Due to these technical difficulties Baseline only described general considerations for weighting, not formal recommendations. Since different approaches to weighting can impact the final KPI estimates, it is recommended to define minimal requirements for sampling weights for future KPI estimates. These should include minimal requirements for either network length or a priori traffic volume approaches. Either way, weighted proportions of observations per stratum need to be checked on plausibility and included in the meta-data provided with the final estimates.

A specific element in the statistical analysis regards the exclusion of low volume locations (and roads) from the sampling frame. From a practical point of view this is entirely warranted. Specifically in cases where traffic volumes used for weighting are derived from traffic counts made during the observations in combination with road network length data. In these cases, it should be verified to which degree traffic volume estimates might be biased due to oversampling high volume roads.

3.7 Confidence intervals

Apart from point estimates for the percentage of non-compliant road users, confidence intervals were also required in order to evaluate the accuracy of the KPI estimates at aggregate and disaggregate levels. In some cases, confidence intervals were not calculated, making international comparisons and evaluating the accuracy of the estimates impossible. A recurrent issue with the reported confidence intervals was a huge variability in confidence interval width that cannot be fully explained by variations in numbers of observations. This is mainly due to variations in calculation methods ranging from assuming simple random sampling through complex statistical models taking different aspects of the sampling design into account. In order to improve statistical comparisons between countries, it is recommended to define minimal requirements for calculating confidence intervals (beyond assuming simple random sampling).

4 Future perspectives

4.1.1 Scope of current KPIs

For several KPIs widening the scope of the observations might be interesting. Distraction, for instance, is mainly focussed on the use of electronic communication devices. Obviously, other types of distraction are also a risk factor (e.g. tuning radio, in vehicle information systems, manipulating objects, reading, ...). Given the efforts invested in sampling and observing huge samples of drivers, it should be considered if the range of risks observed can be extended (cf. Boets, 2022).

The KPI on vehicle safety is currently limited to safety ratings of newly registered vehicles (cf. Wardenier, 2022). In order to obtain a complete picture of the safety of a country's entire vehicle fleet, this should be extended from newly registered vehicles to the total vehicle park per country. This requires scientific modelling to rescale Euro NCAP star ratings of the past. For a start, this extension could be limited to personal cars, but extensions into professional and goods transport can also be considered. Analogous possibilities to extend the observations made during the field work and data collection for KPI measurements are described in the respective KPI reports.

4.1.2 Trendline

At the time of publication, 25 EU Member States formed a consortium under coordination of SWOV to reply to the follow-up call for tenders "MOVE/C2/2022-54 — Technical Assistance for the development and collection of Road safety Key Performance Indicators (KPI)". Four other countries will also join as observers. The objective of the project is to continue the Baseline work on the KPIs defined in the EU Road Safety Policy Framework 2021-2030, by reviewing and adapting the current methodological guidelines and to collect new data for all KPIs in the period 2023-2025. In parallel the current set of 8 KPIs will be extended with experimental and complementary indicators in order to cover road safety indicators so far left out of scope. One of these indicators will concern enforcement of traffic offences.

In this sense Baseline will be literally become the first baseline measurement relative to which future developments will be monitored and evaluated. Equally literally, Trendline sets out to depict historical trends in national KPI performance over time.

References

Boets, S. (2022). Baseline report on the KPI Distraction. Baseline project, Brussels: Vias institute.

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.

Boets, S. Schumacher, M., Stelling, A., Jankowska-Karpa, D., Pavlou, D. (2021). Methodological guidelines – KPI Distraction. Baseline project, Brussels: Vias institute

European Commission (2019). COMMISSION STAFF WORKING DOCUMENT EU Road Safety Policy Framework 2021-2030 - Next steps towards "Vision Zero". Brussels, 19.6.2019 SWD(2019) 283 final.

European Commission (2022). Communication on Workshop on the Methodology Network Wide Road Safety Assessment, 21 december 2022.

Martensen, H.; Daniels S. (2020) Hoeveel slachtoffers kunnen we vermijden door veiliger te rijden? – Omvang van belangrijke risicofactoren in het verkeer in België, Brussel, België: Vias institute – Kenniscentrum Verkeersveiligheid

Nuyttens Nina. (2022). Baseline report on the KPI Post-crash care. Baseline project, Brussels: Vias institute

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

Silverans, P., & Boets, S. (2021). Considerations for sampling weights. Baseline project, Brussels: Vias institute". www.baseline.vias.be

Teuchies, M. et al. (2021). Methodological guidelines – KPI speeding. Baseline project, Brussels: Vias institute

Van den Berghe, W. (2022). Baseline report on the KPI Infrastructure. Baseline project, Brussels: Vias institute

Van den Broek B. & Aarts, L. (2022). Baseline report on the KPI Speeding. Baseline project, Brussels: Vias institute"

Van den Broek B. & Aarts, L. (2022). Baseline report on the KPI Safety belt and Child restraint systems. Baseline project, Brussels: Vias institute"

Van den Broek, Bijleveld, Aarts, & Bos, (2021). Data Quality Control Version 2.0, November 11, 2021. https://baseline.vias.be/storage/minisites/baseline-data-quality-control.pdf

Vanhove, S. (2023). Baseline policy report.

Wardenier, N. (2022). Baseline report on the KPI Vehicle Safety. Baseline project, Brussels: Vias institute

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

Yannis, G., Folla K. (2022). Baseline report on the KPI Driving under the Influence of Alcohol. Baseline project, Brussels: Vias institute