This is the author accepted manuscript of a published article. Please refer to this work as: Snaphaan, T., Hardyns, W., & Pauwels, L. (2022). Expanding the methodological toolkit of criminology and criminal justice with the Total Error Framework. *Journal of Crime & Justice*. https://doi.org/10.1080/0735648X.2022.2114099

Special Issue 'Measurement and Methodology: Addressing Challenges and Exploring Opportunities'
Journal of Crime & Justice

Expanding the methodological toolkit of criminology and criminal justice with the Total Error Framework

Thom Snaphaan[1*], Wim Hardyns[1,2] & Lieven Pauwels[1]

[1] Department of Criminology, Criminal Law and Social Law, Faculty of Law and Criminology, Ghent University, Ghent, Belgium
[2] Safety Sciences, Faculty of Social Sciences, University of Antwerp, Antwerp, Belgium
* Corresponding author (Thom.Snaphaan@UGent.be)

Abstract

The availability and use of new and emerging data sources has increased exponentially. The variety of these data sources offers opportunities to complement, replace, improve or add to conventional data sources. Survey data is one kind of conventional data sources. In survey research, a framework to assess the accuracy of survey data already has been around for quite some time, and goes by the name of the Total Survey Error (TSE) framework (Groves et al., 2004). The philosophy behind this framework has only recently been universalized to (big) data in general in the form of the Total Error Framework (TEF). The current study introduces the TEF to the methodological toolkit of scholars and practitioners in criminology and criminal justice by outlining this generic framework and applying it to an empirical case study (on calculating spatially-referenced crime rates) utilizing two types of administrative data and mobile phone data. The present study discusses the added value and limitations of adapting the TEF, providing guidance to apply the TEF in research and practice. Finally, we propose promising avenues for future inquiries.

Keywords

Accuracy, Big data, Criminal justice, Criminology, Data quality, Methodology, Total Error Framework

1 Introduction

In a world where primary data collection is becoming more challenging, alternatives are increasingly being sought (Japec and Lyberg 2021). For example, traditional survey research faces challenges such as declining response rates, increasing data collection costs, and substantial costs for administering and maintaining census data (de Leeuw and de Heer 2002; Johnson and Smith 2017; Thakuriah, Tilahun, and Zellner 2017). Besides its own challenges, big data can address these challenges because most forms of contemporary big data are actually data of passive solicitation or 'found data' (Japec et al. 2015). These new and emerging data sources have the potential to complement, replace or enhance existing (conventional) data sources (Kitchin 2015; Snaphaan and Hardyns 2021). Big data offer the possibility to enrich existing datasets with finer-grained data that also can be obtained more quickly. New technologies also offer the possibility of collecting data on matters that were previously impossible to measure (at such a scale) (e.g., human mobility through mobile phone data) or that could only be measured at such a coarse-grained level that they were meaningless (e.g., remote sensing data) (Thakuriah, Tilahun, and Zellner 2017). Regarding this evolution from coarse to fine resolutions, Hilbert (2016) makes an analogy to the invention of the telescope in astronomy and the microscope in biology: the impact on the social sciences is immense, given the unprecedented level of fine-grained insights that can be obtained.

In addition to these opportunities, the use of big data also holds several challenges. In this context, reference is often made to the use of (opaque) algorithms and their influences on the way decisions are made. This phenomenon is also referred to as algorithmic governance (Peeters and Schuilenburg 2020) or algocracy (Aneesh 2009), which gives the impression that it are mainly the systems or algorithms used that are biased. However, the challenge lies not so much in the systems or algorithms, but rather in the input: the data itself (Kirkpatrick 2017). Sukel (in Renzenbrink 2019; own translation) notes:

> If an algorithm turns out to be sexist or racist, the blame is often placed on the technology. But if a model does that based on existing data, it is likely that those problems were already there. These problems are just exposed by creating an algorithm for them. It is a waste to then decide not to use the technique. You are throwing away a potential useful tool. It is a signal to say: we have to do this differently.

Indeed, many biases and errors in the outcomes of big data applications can be traced to problems in or with the data (Favaretto, De Clercq, and Elger 2019; Richardson, Schultz, and Crawford 2019). This principle is also known by the adage *garbage in equals garbage out*. In other words, it is highly relevant to look at the (methodological) quality of those data. The focus of this study is on the accuracy of estimates based on various types of measurements, defined in terms of different sources of error. In this context, *error* is defined as the deviation of a measure from its underlying true value (based on Biemer, 2010).

For the methodological quality assessment of survey data, an established standard already exist for quite some time: the Total Survey Error (TSE) framework (Groves et al. 2004). This framework includes all possible errors that can occur in the design and administration of a survey. These errors are considered both individually and in sum to determine the net error in the outcome (Groves et al. 2004; Groves and Lyberg 2010). After all, the net error may be greater than the sum of its parts. Several scholars have advocated assessing the methodological quality of big data or data in general in a similar way as is done with the TSE framework for survey data (e.g., Groves and Lyberg 2010; Japec et al. 2015; Schober et al. 2016; Snaphaan and Hardyns 2021). When insight is obtained into the various sources of error in the data, deliberate choices can be made in the design, processing and analysis methods to minimize or compensate for these errors. In addition, when data sources are combined (e.g., survey data and big data), the strengths of both sources can be leveraged and the weaknesses eliminated, yielding a stronger combined result. Finally, by understanding the net error per data source, deliberate choices can be made between different data sources. Understanding the strengths and weaknesses of specific data sources allows one to assess under what circumstances and for what research questions a data source is best suited. In fact, the rationale behind the TSE framework is extended to (big) data in general, under the name of the Total Error Framework (TEF; Amaya, Biemer, and Kinyon 2020). The added value of the TEF lies in the explicit focus on the decomposition of the various methodological errors, whereby a data source can be evaluated based on its strengths and weaknesses, and, hence, go beyond the idea that a data source is either 'right' or 'wrong'.

The purpose of this contribution is to introduce the TEF to the methodological toolkit of criminology and criminal justice scholars by raising awareness and make scholars in future research think critically about the sources of error that can occur in the different stages of data collection, processing and analysis. Therefore, we will first describe the origins of the TEF by means of a brief outline of the background of the TSE framework. Then, we describe the conceptual framework of the TEF and thereafter we will apply this TEF to a case study relevant to criminology and criminal justice. In the discussion, we will elaborate on the implications for research, policy and practice, and discuss how to put this into practice. In this contribution, we will not measure the individual error components quantitatively and we will not estimate the net error per data source (see Biemer and Amaya 2021).

2 The Total Survey Error framework

The method of random sampling – one of the cornerstones of survey research (Hox, de Leeuw, and Dillman 2008) – focuses largely on the variance in results introduced by the random sampling process itself (Neyman 1934) and consequently makes strong assumptions regarding the non-sampling errors. Early on, many of these assumptions were challenged in practicing survey research (Groves and Lyberg 2010). Deming (1944) was the first to outline multiple error components in probability samples. His notion of survey error ("factors that affect the usefulness of surveys"; Deming 1944, 359) was significantly broader than what is understood today as the Total Survey Error framework. He also included non-statistical and even unmeasurable error components, such as bias introduced by survey sponsorship ("bias of the auspices"; Deming 1944, 363). Several scholars have built on Deming's framework or developed similar initiatives (e.g., Anderson, Kasper, and Frankel 1979; Cochran 1953; Hansen, Hurwitz, and Madow 1953; Kish 1965).

Groves (1989) integrated the ideas of TSE with insights from psychometrics and econometrics. In doing so, he focused specifically on measuring the various error components. In 2004, Groves and colleagues first linked the TSE components to the measurement and representation processes of surveys. By doing so, they illustrated how the worlds of survey researchers (primarily focused on representational processes) and statisticians (primarily focused on measurement processes) converge (Saris 2014). Biemer and Lyberg (2003) paid specific attention to the distinction between sampling and non-sampling errors. They were also the first to attempt to unify notions of continuous improvement (from quality management) within the TSE framework. Relevant to note in this regard is that this approach also allows for non-statistical notions of survey quality (see above). However, it is not the intention to put on the user's hat in order to assess credibility and relevance (see Groves et al. 2009). We explicitly start from this idea, because the quality of the data is independent of its use. A statistical measure can be useful independent of its degree of accuracy and vice versa (Groves and Lyberg 2010). Therefore, these aspects should also be evaluated separately. The Total (Survey) Error framework focuses on the inherent accuracy of data sources.

The Total Survey Error (TSE) framework provides guidance on how to sum all errors that may occur in the design, collection, processing, and analysis of survey data (Biemer and Lyberg 2020). Data accuracy refers to the extent to which the actual value differs from the observed value. Conceptually, a distinction is made between observation and non-observation errors on the one hand, and systematic errors and random errors on the other. Observation errors refer to the errors in measurement processes. Non-observation errors refer to deviations between the actual value and the measured value due to representation processes. In addition, observed values can deviate from actual values both systematically and randomly. In other words, these are systematic errors (bias - weighing on validity) and random errors (*variance*[1]- weighing on reliability) that affect the accuracy of survey data. Both have an impact on the accuracy: bias due to systematic deviation and variability due to the instability of measurements.

The goal of TSE, however, is not to carry out every stage of the survey process as flawlessly as possible. After all, that would mean that, given scarce resources, the available budget and/or timing should be exceeded disproportionately. Moreover, even under the best of circumstances, with an unlimited budget and unlimited time, errors (both bias and variability) will always occur. Instead, the goal is to avoid the most egregious errors and reduce other errors to an insignificant and acceptable level (Biemer 2010). For a more detailed account of the creation of the TSE framework, see Biemer and Lyberg (2020), Groves and Lyberg (2010), and Groves et al. (2009).

3 Towards a Total Error Framework

The efforts to extend the TSE framework to a broader palette of data sources can be broadly classified into three stages. These three stages followed each other over the past decade at a rapid pace.

In the **first stage**, the basic principles of the TSE framework were applied to administrative data (Reid, Zabala, and Holmberg 2017; Statistics New Zealand 2016; Zhang 2012). Zhang (2012) builds on Bakker's (2010) idea that the errors that occur in the creation process of survey data will take similar forms in the creation of administrative data. Zhang (2012) created a two-step model for (the integration of) statistical microdata. The first step covers the creation process and related error components of one specific data source. The second step involves the integration of data from different data sources.

Statistics New Zealand has adopted Zhang's (2012) model and operationalized it further by creating concrete quality indicators for the two steps of the model, which should be reported on when integrating data sources. Reid, Zabala, and Holmberg (2017) built on Zhang's model by adding a third step. This third step aims to include consideration of the processes and potential

[1] This concept is also known as noise (e.g., Kahneman, Sibony, and Sunstein 2021).

errors that occur when formatting final statistical products based on integrated datasets (i.e., the result of step 2 in Zhang's model). These second and third steps are very common in practice.

In the **second stage**, big data sources, other than administrative data, were considered in light of some of the tenets of the TSE framework, but these were not comprehensive in terms of error types nor in terms of big data sources (Amaya, Biemer, and Kinyon 2020). For example, Baker (2017) provided an overview of some big data challenges in general and refers to the TSE framework, but did not refer to the specific error types. Other studies considered specific data sources, such as Twitter data (Hsieh and Murphy 2017) or social media in general (Schober et al. 2016), however, these were also not comprehensive in terms of the error types considered.

The **third stage** starts from a comprehensive framework that is applicable (or at least attempts to be applicable) to a broad palette of data sources, both survey data, administrative data sources, and big data sources.[2] The first efforts at this stage were made by the *AAPOR[3] Task Force on Big Data* (Japec et al. 2015). The *Big Data Total Error* (BDTE) framework of the AAPOR Task Force explicitly starts from the TSE framework and adds additional error types that are specific to big data and can create substantial biases and uncertainties in big data products. Biemer (2016) builds on the ideas of the BDTE framework and argues that the error types are easily reducible to column errors, row errors, and cell errors, referring to a table format as often found in relational databases. He argues that this simple representation is applicable to many (structured) data sources. However, such a simplification is difficult to apply to unstructured data (e.g., text or imagery). From this observation, taking into account the huge amount of unstructured data and the lack of direct added value of this simplified representation, Amaya and colleagues (2020) propose a *Total Error Framework* (TEF).[4] This TEF contains a decomposition of error types, analogous to the TSE framework, that can be positioned in the data collection, processing and analysis process (see Amaya, Biemer, and Kinyon 2020). This TEF will be outlined in the next section.

---

[2] It should be mentioned that more general frameworks exist for big data quality assessments (e.g., United Nations Economic Commission for Europe 2014). These more general frameworks also take into account factors that are broader than the TEF. See also the earlier reference to the non-statistical notions of survey quality. In this contribution, the focus is on the accuracy of, or in other words the absence of errors in, the data itself.

[3] American Association for Public Opinion Research.

[4] Lavrakas (2013) was the first who coined the term and laid the conceptual foundation, Amaya, Biemer, and Kinyon (2020) then linked these ideas to big data sources.

4 The Total Error Framework

Amaya, Biemer, and Kinyon (2020) identify important similarities in the creation processes of survey data and big data, and argue that about eight error components recur in both processes. These error components can also be framed within the measurement, representation, and analysis processes, creating a generic structure that can be applied to a broad range of data sources. These error components are outlined in the next three sections. Figure 1 shows how these error components (ovals) relate to the generic process steps (squares).

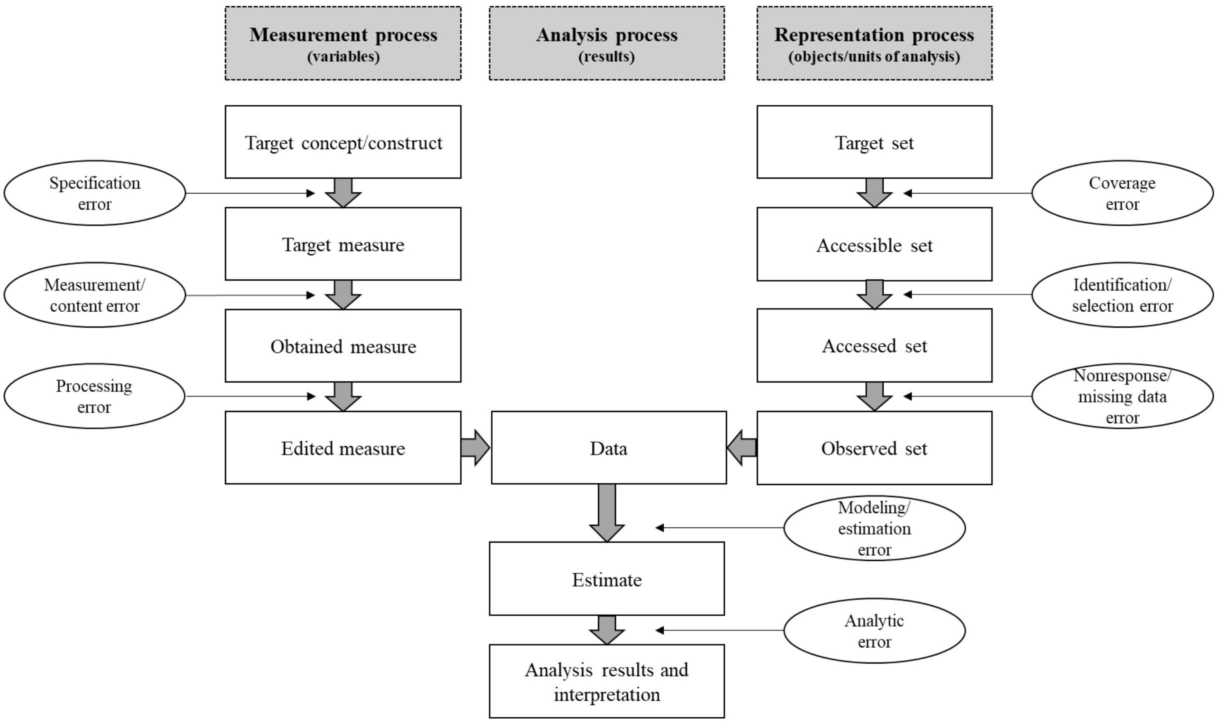

Figure 1. Total Error components (ovals) in relation to a generic process (squares) from theoretical concept and target population to analysis and interpretation of results.

4.1 Errors in measurement processes

The **specification error** occurs when the concept (or construct) needed to answer the research question does not fully match the concept that can be reflected by the data (Amaya, Biemer, and Kinyon 2020). The fact that big data is mostly *found data* (Chen, Li, and Li 2016; Connelly et al. 2016) implies that the researcher has no influence on the data collection procedure and the operationalization of the variables. In fact, this is where the validity issue starts: are we measuring what we want to measure?

Measurement/content error refers to all errors that occur during the collection of the data. In survey research, this includes, for example, question wording effects, memory effects, and interviewer effects. Intentional falsification of data also falls under this category. Measurement errors are the most damaging type of errors in measurement procedures, as they relate to the reliability of the measurements and consequently affect their precision. In the context of big data, the causes of measurement/content errors are diverse. These include errors in the measurement process, transcription errors, conversion errors, incorrect reading of mechanical devices, outdated data, et cetera (Amaya, Biemer, and Kinyon 2020).

Processing error simply means the degree to which data is processed without error. Processing errors arise from processes such as data entry, coding, conversion of variables, et cetera. In the case of big data, these errors are often related to processing and/or transforming the data. In addition, big data is frequently linked with other data sources, which causes processing errors to creep into the data due to incorrect data linking (Amaya, Biemer, and Kinyon 2020).

4.2 Errors in representational processes

Coverage error refers to the mismatch between the units returned in the data source and the target population (Amaya, Biemer, and Kinyon 2020). Under-coverage implies that certain units from the target population are not reflected in the data source. Over-coverage means that certain units that are outside the target population are reflected in the data source. There may also be over-coverage due to duplicates in the data source. In addition to these coverage errors that pertain between units, it is also possible for a coverage error to occur within units. This within-unit coverage error occurs when errors have crept into the selected units, making them different from the unselected units and therefore not a representative reflection of the desired subset (in this case, the unselected).

Identification/selection error refers to those errors that result from the identification and selection of a (usually random) subset of the target population rather than the entire population (Amaya, Biemer, and Kinyon 2020; Zhang 2012). In the context of survey research, this type of errors is referred to as 'sampling errors', which is due to several factors (e.g., sampling size, sampling design and weighting procedures), but sampling is often not the case in big data research. One of the advantages of big data is that due to exhaustivity – which is next to velocity one of the two key boundary characteristics that differentiate it from 'small' data (Kitchin and McArdle 2016) – there is generally speaking no need to rely on sampling. If this aspiration is approached, the resulting estimates are little to unaffected by sampling error. However, one should be aware that large numbers (analogous to large sample sizes) in statistical analyses will quickly produce significant results, even in the absence of a meaningful difference in the entire population. The risk of non-selection is always lurking, think for example of hidden populations, which are particularly relevant in criminological research (e.g., Curtis 2010). When using multiple data sources, merging them into a dataset can create identification problems, as this requires a unique identifier, which is not always available in the case of big data sources.

The **nonresponse/missing data error** refers to all errors that result in missing data (Amaya, Biemer, and Kinyon 2020). In survey research, this often involves unit or item non-response. Despite the fact that big data sets often do not involve "response" as in a survey, the dichotomy of missing items and missing units also translates to this type of data source. With big data, however, the potential causes of missing data are more diverse because the data sources are more varied (see below). It should be noted that in the case of big data, these missing data are often associated with or caused by coverage errors (under-coverage) or sampling errors (non-selection).

4.3 Errors in analysis processes

Modeling/estimation error includes those errors that arise from the shortcomings in applied weighting corrections and imputation methods for missing data (Amaya, Biemer, and Kinyon 2020). The methods that can be applied for big data and survey data are similar, but the challenges are different. The mechanisms responsible for missing data are mostly unknown in the case of big data, in contrast to survey data (e.g., Little and Rubin 2002). In addition, in a big data set, little additional information (i.e., additional variables) is usually available per record. These properties complicate possible weighting and imputation in the case of missing data. Analogous to methods used with survey data, however, it is possible to use external calibration data with big data. However, as described above, linking big data sets in this case can create challenges.

The **analytic error** includes errors made by users of the data when analyzing and interpreting it. In this category of errors, there are also many similarities between survey data and big data. One example is the ecological (Robinson 1950) or atomistic fallacy, whereby conclusions are always interpreted at the wrong aggregation level (or relations between variables at a higher aggregation level are assumed to be valid at a lower aggregation level and vice versa). Generalization errors and reification fallacies are also examples of this. These errors involve, respectively, drawing conclusions about complex properties based on incomplete or generalized measurements and fallacies of misplaced concreteness. A strength of big data compared to survey data is the high granularity of the data (e.g., spatiotemporal), which makes inferences at different levels possible (and potentially valid). A weakness is that big data tend to be univariate and do not allow for (quasi)experimental research, as this requires high-dimensional data (Amaya, Biemer, and Kinyon 2020).

5 Applying the Total Error Framework

In this section, we will elaborate on an empirical case study where we apply the Total Error Framework to assess the suitability of data sources for an illustrative, but relevant example. We start with a brief outline of the research problem in this case study, subsequently describe the datasets used and lastly compare the data sources based on the Total Error Framework. The results of this empirical study are published in Rummens et al. (2021). In this article, we assess the data sources used from a methodological point of view.[5]

5.1 Research problem

In short, this case study relates to the *denominator dilemma*, which encompasses the issue that the population-at-risk (denominator) is particularly problematic for crime (numerator) in calculating crime rates (Boggs 1965). A crime rate is "a statistic often used to represent the risk of criminal events [and that] help[s] to reveal clusters of crime in space and/or time based on an underlying population at risk" (Malleson and Andresen 2015: 112). Although the residential population is frequently used in representing the population-at-risk, Harries (1981: 148) already mentioned more than 40 years ago that the "uncritical application of population as a denominator for all crime categories may yield patterns that are at best misleading and at worst bizarre". Nowadays, new and emerging data sources enable scholars to estimate the ambient population as a population-at-risk (e.g., Malleson and Andresen 2016). The ambient population concerns the number of people that are actually present in a given area at a given time (Andresen, 2006).

5.2 Data

In this case study, we compare two measures of the population-at-risk for the purpose of calculating crime rates in Ghent, Belgium. We assess the population-at-risk based on both administrative data (from the population register) and mobile phone data (from call detail records). Besides, we use police-registered crime data to compose the numerator in calculating crime rates. The units of analysis in this study are grid cells of 200 by 200 meters (*n*=4,254 grid cells for Ghent, Belgium). This grid level is created by spreading a uniform grid over a geographical area. All data were mapped to this grid.

[5] In view of the scope of this contribution, we will only briefly discuss how the TEF can be applied to these data sources. An extensive, detailed and limitative application of this framework would cover a complete or even several contributions per data source. As such we do not claim we present an exhaustive description of all error components in the considered data sources. The reader should bear this in mind when interpreting the results.

The first measure of the population-at-risk is composed from the residential population measured by **population register (Dutch: 'Rijksregister') data**. These data were obtained from the City of Ghent, who provided the most recent data available at that time (2018). The City of Ghent mapped these data to the grid level and masked grid cells with ≤ 4 but ≠ 0 inhabitants by equalizing all these cells to have 4 inhabitants (6.28 percent of the grid cells). Cells with 0 inhabitants were labeled as such (48.19 percent of the grid cells).

The second measure of the population-at-risk concerns the ambient population measured by **mobile phone data**. These data were composed of call detail records (CDRs) from Proximus, the largest mobile phone operator in Belgium (with a market share of 38.7 percent in 2019; Proximus 2020). The counts of mobile phones were extrapolated to the population, taking into account this market share and its distribution. These mobile phone data are tied to cell towers and as a consequence spatially distributed in the form of so-called Thiessen polygons (i.e. a hexagonal grid; with 288 cells for Ghent). The individual size of these cells depends on the population density: the higher the population density, the more cell towers, hence, the smaller the Thiessen polygons (see Rummens et al. 2021). Proximus mapped these data to the grid level and excluded cells with ≤ 3 phones present (1.06 percent of the data points in the raw data). Mention that the amount of raw data (i.e. counts of unique mobile phones in a grid cell for a time period of one hour) is 595,858,852 for a period of three months (October – December 2019). Divided by the number of days (*n*=92) and 24 hours per day for which we have data and unique phones are counted, the average number of individual phones in Ghent counts 269,740, which approximates the number of inhabitants for that year (*n*=261,483), however, taking into account the tourist function of Ghent, a higher number is as would be expected. In Figure 2, the spatial distribution of both measures of the population-at-risk is shown.

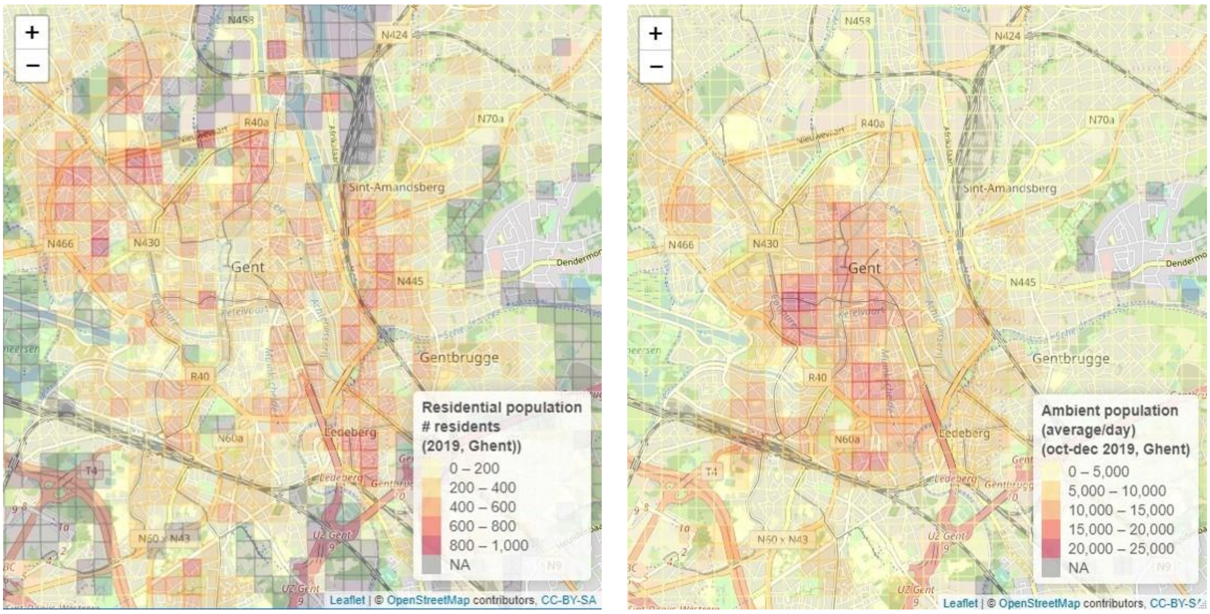

Figure 2. Spatial distribution of two measures of the population-at-risk: residential population measured by population register data (left) and ambient population measured by mobile phone data (right).

The third data source used in this study is another form of administrative data: **crime data** registered and provided by the Local Police force of Ghent. In accordance with the available mobile phone data, these crime data span a time period of three months (October, November, December) in 2019. In this study, crime data for three crime types are used: aggressive theft, battery, and bicycle theft; covering both violent and property crime types. Aggressive robbery includes (attempts of) purse snatching or robbery using weapons or threats. Battery includes international use of force or violence resulting in injuries, where intra-familial violence is excluded. Bicycle theft includes (attempts of) theft of locked or unlocked bicycles in the public space. The dataset contains details on the location at the address level and the exact time or time range during which the crime (presumably) took place. In cases where only a time range was available, we assumed the crime event took place at the midpoint of this range. The addresses were geocoded to add x,y coordinates to the crime data. In cases where the house number was missing and only the street name was mentioned, we randomly assigned a grid cell from the grid cells that overlap the street (respectively in 67.80%, 35.06% and 34.15% of the cases). In the case of unidentifiable street names, we deleted these crime events from the dataset (respectively in 13.56%, 1.56% and 0.98% of the cases). Finally, the dataset contained 49 cases of aggressive theft, 293 cases of battery, and 571 cases of bicycle theft.

5.3 A Total Error perspective on population register data

Census and population register data are frequently used in academic research (Bakker, 2012). In criminological research, these data are either to be used as control variables or as comparators to other (big) data sources (e.g., Snaphaan and Hardyns 2021). This data source has particular relevance in the context of the research problem at hand, because the residential population is the favored denominator in calculating crime rates (Andresen and Jenion 2010). These data can be obtained from the population register and the units of analysis are individuals. A population register is defined as “an individualized data system, that is, a mechanism of continuous recording, and/or of coordinated linkage, of selected information pertaining to each member of the resident population of a country in such a way to provide the possibility of determining up-to-date information concerning the size and characteristics of that population at selected time intervals” (United Nations 1969: Chap. I.A.).

Perhaps, the most frequently invoked error in using the residential population as a measure of the population-at-risk is that this measure does not accurately reflect the population-at-risk for particular crime types (Harries 1981), which relates to the **specification error** as a part of the measurement process. The presence of **measurement errors** is ostensibly lower than for example with survey (e.g., census) data, however, the errors in these administrative data should not be underestimated, with for example persons or other entities that may have an interest in being registered in a particular way or severe administrative delays in the data (Bakker, 2012). **Processing errors** may for example arise due to administrative practices of the register keeper that may lead to biased entries (Bakker, 2012).

In the representation process, register data are assumed to be comprehensive since these span the total population. Nevertheless, **coverage errors** may occur due to inconsistencies in the registration practices: deaths and emigrations that have not been reported before updating the register cause over-coverage, while births and immigrations that have not been reported timely cause under-coverage. **Identification or selection errors** could be introduced when linking datasets, which is very common with register data to broaden the dataset and include more variables from other registers or surveys (Bakker, 2010). The problem of missed links corresponds with non-response error in surveys. If missed links are non-random, they will lead to biased outcomes. Population estimates based on register data contain generally lesser **missing data**, with the exception of those introduced by under-coverage and identification/selection errors.

Regarding the analysis processes, **modeling and estimation errors** occur while making adjustments in the dataset for missing data, for example by weighting or imputing values (Wallgren and Wallgren 2007). As described above, there is relatively little missing data in population estimates based on register data. **Analytical errors** are related to the analytical strategy rather than the specific data source. Calculating (spatially-referenced) crime rates can hold the specific risk of introducing ecological or atomistic fallacies, where results are interpreted at an inappropriate level of aggregation. In the case of the ecological fallacy, relationships at a higher level of geographical aggregation (e.g., neighborhood level) are interpreted at a lower level (e.g., individual level; Robinson 1950).

5.4 A Total Error perspective on mobile phone data

Mobile phone data have been frequently used in (environmental) criminological research in recent years (Snaphaan and Hardyns 2021) and also have relevance for police services. These data can, for example, be usefully employed to express crime rates more validly and improve the predictive accuracy of predictive policing models (Rummens et al. 2021).

In terms of the measurement processes, a number of errors can be identified. First, the **specification error**: while these data estimate the number of individuals (i.e. the unit of analysis) that are actually at a certain place and time, there is no continuous measurement of an individual's location, because a mobile phone (based on call detail records) only makes contact with a cell tower when trafficking communications or at regular intervals. Thus, human mobility is not fully reflected using these mobile phone data. **Measurement errors** are caused by the non-accuracy of measurements due to the technical process of how mobile phones communicate with cell towers (Malleson and Andresen 2016). **Processing errors** can occur when mobile phone data are converted in their raw form to a format to facilitate further analysis, e.g., when the geographic level is adjusted. When mobile phone data are used to calculate crime rates, one is interested in how many individuals (i.e., mobile phones) were present in a certain place at a certain time. For this purpose, the data are obtained as so-called Thiessen polygons: a representation centered around point measurements of the cell towers. The size of these polygons fluctuates: the closer the cell towers are to each other, the smaller the polygons (therefore measurements in urban areas tend to be more accurate than measurements in rural areas). Converting these Thiessen polygons to smaller geographic units creates processing errors to a greater or lesser extent, depending on the method used.

The errors that occur in the representation processes are related to the degree to which the target population is approached by the research units. A major challenge in this context starts with the **coverage error**: first, there is under-coverage, because not everyone has a cell phone and, moreover, this use is not distributed proportionally across the population, which is problematic when aiming at the general population as the target population (Keusch et al. 2020). Second, duplicates may occur when one person carries two or more cell phones. **Selection error** occurs when the telecom provider's system does not select all relevant units within the geographical area and time frame of interest. For example, **missing data** occurs when a person who does have a cell phone is not carrying the device, but is within the geographic area and time frame of interest.

The **modeling and estimation error** occurs when the mobile phone data are extrapolated to the general population. This is often done because the mobile phone operator only has data on part of the sales market (i.e. market share). In this study, Proximus (2020) holds a market share of 38.7% in 2019 and conducted an extrapolation based on the distribution of this market share compared to the total population. This extrapolation relies on a number of assumptions, which cause errors. The **analytical error** applies in a similar way as to the register data (see above).

5.5 A Total Error perspective on police-registered crime data

The most obvious form of administrative data available to law enforcement agencies are the data they generate themselves. These data exist in different forms and systems. A large amount of crime and disorder does not end up in the registrations (i.e., the dark number) and this causes several problems (e.g., Skogan 1977). A large number of prior studies have focused on measurement errors in crime registrations (e.g., Biderman and Lynch 1991; Buil-Gil et al. 2021; Lynch and Addington 2007; Mosher, Miethe, and Phillips 2002), however, none of them from a Total Error perspective.

Looking at the measurement processes in the case of police-registered crime data, a number of errors can be identified. First, it is possible that the concepts one is interested in do not correspond to the available data (i.e. **specification error**). For example, when estimating the number of cybercrime incidents from police data, we overlook the fact that a large part of the cybercrime incidents are not reported to the police but to other institutions (Van de Weijer, Leukfeldt, and Bernasco 2019; De Kimpe et al. 2021). Second, **measurement/content errors** can occur when the concepts are actually measured. These include incorrect and incomplete registrations (Klinger and Bridges 1997), intended or unintended and both on the side of the reporter and on the side of the police. Third, there may be **processing errors**. An example in the context of calculating spatially-referenced crime rates are geocoding errors. (Historical) records of crime and disorder are usually not provided with GPS coordinates, so other location data (e.g., street, house number, postal code, place of residence) must be used in order to conduct geographical analyses. The extent to which these data are missing varies by crime type, but a 100% geocoding ratio is never achieved in practice (Andresen et al. 2020; e.g., Hardyns et al. 2019), so there is always a loss of precision or even units of analysis.

Considering representational processes, the first thing to mention explicitly is what the units of analysis are. In this case, the units of analysis are the crime incidents. The **coverage error** occurs, since the records are not a representative reflection of crime that is actually committed. In other words, both the dark and grey number (e.g., Bottomley and Pease 1986; Skogan 1977) apply, and duplicate registrations are threats in the context of this error type. These phenomena in turn have specific causes, such as individuals' varying willingness to report (depending on individual and contextual factors; Warner and Pierce 1993). For example, in our case, bicycle theft is among the most recorded crime types in Belgium and is actually the most recorded crime type in the region of Flanders (Federal Police 2021a). However, the results from the most recent Belgian Security Monitor indicate that bicycle theft is among the crime types with the lowest citizens' willingness to report: in the 12 months prior to the data collection, 10% of Belgian households reported being a victim of bicycle theft, but only 48.1% reported this to the police (Federal Police 2021b). **Selection error** occurs when (random) selections are made in the data, for example in time and place, that do not or do not sufficiently reflect the target population. This type of error also occurs when, for example, an incorrect or incomplete selection is made in the crime types. Similar research with emergency request data shows that the nature of the calls is highly diverse (e.g., Ratcliffe 2021) and aggregating across crime types is mostly inappropriate (Andresen and Linning 2012). The **non-response error** or **missing data** can have many causes in the case of administrative data (Liao et al. 2021). For example, because certain items (i.e. *item missing*) in the data have not been transmitted by the reporter or registered by the government agency. For example, the exact location of the event (e.g., missing street and/or house number). But also as a result of the coverage error and the selection error, missing data arises, such as missing units (i.e. *unit missing*) as a result of not reporting or not registering.

The errors that occur in the analytical processes are logically related to the scope of the study and the analytical strategy which translates the data into results. Looking at the **modeling/estimation error**, the mechanisms of missing data in police data are well-known, however, the imputation methods could be improved (Lynch and Jarvis 2008). The **analytic errors** in this case study are related to the potential inappropriate level of aggregation at which the results are interpreted (see above).

5.6 Comparison and summary

To assess the actual differences in the outcomes of these analyses, Table 1 shows the point-biserial correlation coefficients for the two measures of the population-at-risk in relation to the three crime types studied. The estimates based on mobile phone data yield stronger correlations with the different crime types over the various months than the estimates based on population register data. This was as expected, since all three crime types have a mobile target (see Wikström 1991) and therefore a static measure of the population (from the population register) was hypothesized to not be a valid proxy for the target. Hence, using a more dynamic measure of the (ambient) population is more valid and also stronger associated with the three crime types.

Table 1. Point-biserial correlation coefficients of crime vs. population-at-risk based on respectively population register data and mobile phone data at the grid level (Rummens et al. 2021).

Crime type	Month	Population register data	Mobile phone data	Correlation difference
Aggressive theft	Oct	0.12***	0.16***	0.04*
	Nov	0.08***	0.18***	0.10*
	Dec	0.12***	0.13***	0.01*
Battery	Oct	0.23***	0.26***	0.03*
	Nov	0.23***	0.25***	0.02*
	Dec	0.21***	0.22***	0.01*
Bicycle theft	Oct	0.34***	0.47***	0.13*
	Nov	0.27***	0.41***	0.14*
	Dec	0.23***	0.35***	0.12*

* significant at <0.05; ** significant at <0.01; *** significant at <0.001

In Figure 3, we summarized the given examples for the three data sources discussed in this empirical application, in the format of the model from Figure 1. The respective data sources are indicated with icons (see Legend in Figure 1).

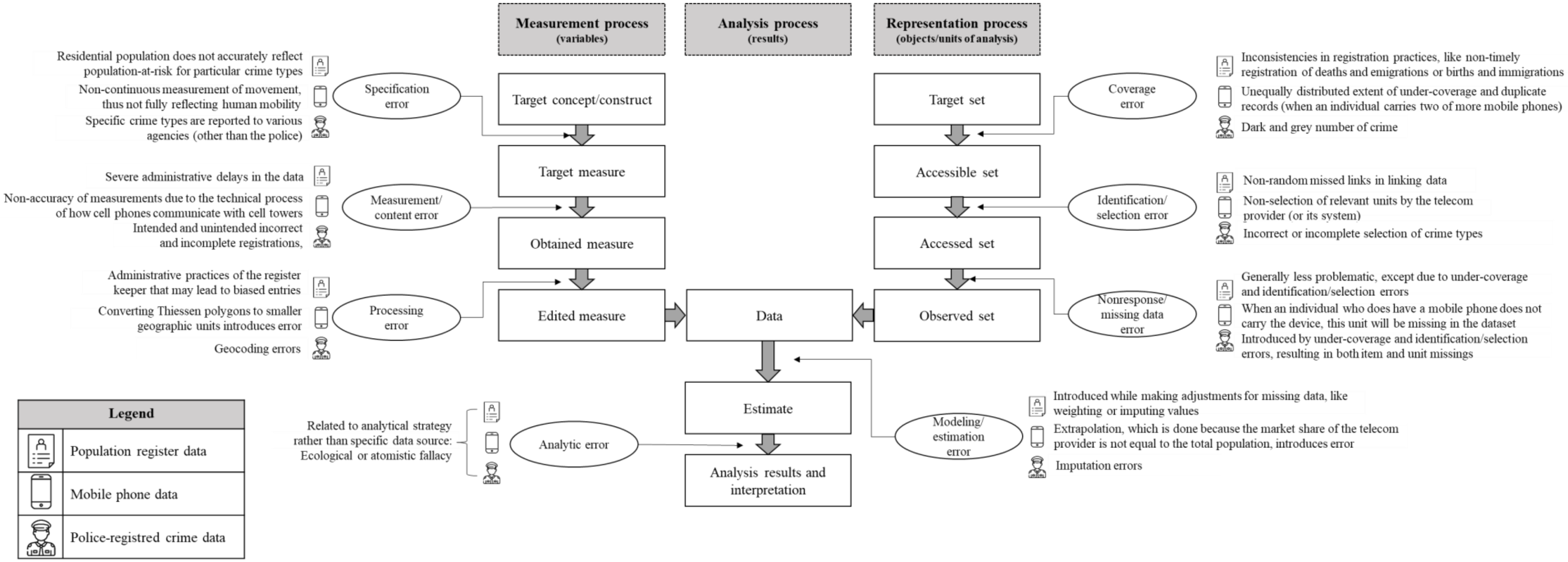

Figure 3. Summary of the given examples of errors in the three data sources from the empirical case study.

6 Discussion

The purpose of this contribution was to introduce the Total Error Framework in the field of criminology and criminal justice, to provide guidance for the methodological quality assessment of (big) data. Because, as Kreuter (2021, 615) puts it: "as with survey data, we will need robust frameworks and metrics to assess the quality of the data provided by governments, academic institutions and the private sector, and to guide us in using such data." We have shown that the TEF is suitable for assessing the accuracy of data derived from various data sources. In addition, we hope that this will contribute to a broadening of the discussion on the use of big data, which often boils down to the legal question of whether the data *may be* used or the ethical question how we *should like* to use them. We would like to let a question precede this: *can* we use the data at all in the context of the research question(s), in other words, do they provide a sufficiently valid and reliable measure of the concepts and population of interest? Even more, we hope that in future research, scholars go beyond the opportunistic selection of data sources and select the data source that measures the concept (or construct) of interest the best. Particularly in the context of research with big data sources, big data hubris (Lazer et al. 2014) or the belief that "bigger is better" is lurking. To prevent this, TEF can provide a framework to make deliberate choices. In line with Biemer (2010), we put forward the TEF as a theoretical framework for optimizing (choices regarding) data by maximizing data quality within budgetary constraints.

In conclusion, this study found that the TEF is applicable to new and emerging data sources relevant to criminology and criminal justice. By looking at (big) data sources from a TEF perspective, we assessed the accuracy of these data in the light of the concepts that were intended to be measured. In the examples given, it can be concluded that specific errors are present for both measures of the population-at-risk, and the choice have to be made by a trade-off of these errors, also in the light of the broader assessment of data quality (see below). For example, in this case, the specification error for mobile phone data is less problematic than for population register data, but for the coverage error the opposite is true. The fact that mobile phone data are considerably more dynamic than population register data brings new opportunities and weighs heavily for the user, however, other errors should not be overlooked. Specifically, for the three crimes types used in the worked example the mobile phone data would be more suitable given the mobile targets and, as a consequence, the varying risk of victimization over space and time. For crime types with immobile targets (e.g., residential burglary) this might be not the case.

The question remains: what are the implications for criminology and criminal justice, and how should we bring this into practice? In general, it should be mentioned that given the predominantly *found* nature of big data sources and in contrast to conventional data sources (such as surveys and observations) which can be considered as *made* data, there is usually little influence that can be exerted on the measurement processes. Nevertheless, the errors can be identified so that they can be taken into account when interpreting the results. Besides, it is indeed possible to influence the selection and analysis processes. Moreover, understanding the errors in the various processes also promotes appropriate trade-offs in the use of a data source and/or its combination with other data sources. Prior research shows that by combining data sources, the properties or strengths of one data source, can override the shortcomings in another data source (Biemer and Amaya 2021). Besides the use of the TEF in the selection phase of data sources, we would also suggest using the framework to be more explicit about assumptions and choices that are made during the research process. Last, this framework can also be subject matter in methodological research (see below).

For criminology, from a scientific perspective, there are legitimate concerns about a lapse into new forms of empiricism, where “data can speak for itself” without a need for theory (Kitchin 2014). A more productive way to look at this departs from a scientific-realist approach, as analytic criminology aims to do (Bunge 2006; Wikström and Kroneberg 2022). This approach starts from the perspective that causes and mechanisms are not “invented,” but discovered; and that reality therefore exists independently of ourselves. The better we understand how reality works, the better we can explain social phenomena (such as crime). Knowledge about reality is built on the one hand by observation (empirical study) and on the other hand by reasoning (theorizing), which are of equal and independent importance (Wikström and Kroneberg 2022). In order to guarantee ‘pure’ observations in a mechanism-based approach, and thus draw valid conclusions with respect to the tested theories, it is necessary that the data used are sufficiently accurate. Therefore, it is essential to be able to assess very diverse types of data sources from an overarching framework, thus enabling scientific justification for using or not using particular data sources.

For the sake of a first introduction to the criminological literature, we have limited ourselves in this contribution to a conceptual description of the TEF. To go a step further, the error types could also be quantified, estimating a measure of the *total error* in a (combined) dataset. In this way, quantitative statements can be made about the appropriateness of using the dataset for the specified purposes and multiple data sources can be compared. Therefore, the total Mean

Squared Error (MSE) could be used (e.g., Biemer and Amaya 2021) or one could explore the use of a Multi-Trait Multi Method (MTMM) approach (Saris and Andrews 1991) in this context. Nevertheless, an important added value of the TEF is that it allows us to identify the strengths and weaknesses of a data source, which is many times more fruitful than stating that a data source is “good” or “bad”.

> No statistical dataset perfectly measures exactly what we want it to. At present we cannot provide a single generic measure to summarize data quality. (...) Instead of judging a dataset as ‘good’ or ‘bad’, the framework identifies the strengths and weaknesses of a dataset in an objective way, with reference to its original purpose. (Statistics New Zealand 2016, 5)

Although it was not the scope of this contribution, in practice it makes sense to look at non-statistical parameters of quality as well, and thus to look more broadly than just at the accuracy of the data itself. According to Karr et al. (2006) there are three hyperdimensions of data quality, related to (1) process, (2) data and (3) user. However the framework introduced in this article relates to the first two dimensions, the latter is ignored. The focus of the TEF is on the accuracy of (big) data, but the notion of total data quality is broader. Additionally, we have chosen a breadth view over a depth view. Therefore, the given examples do not present an exhaustive overview of the errors in the data sources that were discussed. The aim was to demonstrate multidimensionality, not provide a limitative assessment of the errors in the data sources.

Future endeavors in criminology and criminal justice will increasingly use new and emerging data sources. Given the varied characteristics of these diverse data sources, it is fruitful to use a universal framework to assess errors affecting the accuracy of these data. Therefore, we recommend using this framework in making deliberate choices in future situations. Future methodological research should include a comprehensive and (attempted) limitative overview of the errors in specific data sources. Additionally, the TEF provides a general framework to assess the accuracy of data sources, independently of the purpose of the data source (e.g., theory-testing or predictive analysis). However, in predictive analysis, there are specific quantitative measures of (predictive) accuracy available. Future research should assess how this TEF relates to these measures of accuracy in predictive modelling, such as precision, recall and F1-score.

As Braga and Clarke (2014) argue, there is a need to leverage a broader palette of data sources, more microdata, and a greater variety of research designs to advance our knowledge and understanding of criminological phenomena. Big data come with specific opportunities and challenges, but one of the largest opportunities is certainly the availability of a wide variety of data (Kitchin 2015). It is necessary to consider these data critically and to always pre-test the methodological quality, in particular the accuracy, of each data source before making statements based on it. The introduction of the TEF in this contribution should provide the necessary levers to do so. We therefore invite future research to consider the methodological quality of the great diversity of big data sources in criminology and criminal justice based on this Total Error Framework.

References

Amaya, Ashley, Paul P. Biemer, and David Kinyon. 2020. "Total error in a big data world: Adapting the TSE framework to big data." *Journal of Survey Statistics and Methodology* 8: 89–119. https://doi.org/10.1093/jssam/smz056

Anderson, Ronald, Judith Kasper, and Martin Frankel. 1979. *Total survey error: Applications to improve health surveys.* San Francisco, CA: Jossey-Bass.

Andresen, Martin A. 2006. "Crime measures and the spatial analysis of criminal activity." *British Journal of Criminology* 46: 258–285.

Andresen, Martin A., and Greg W. Jenion. 2010. "Ambient populations and the calculation of crime rates and risk." *Security Journal* 23(2): 114–133. https://doi.org/10.1057/sj.2008.1

Andresen, Martin A., and Shannon J. Linning. 2012. "The (in)appropriateness of aggregating across crime types." *Applied Geography* 35 (1–2): 275–282. https://doi.org/10.1016/j.apgeog.2012.07.007

Andresen, Martin A., Nick Malleson, Wouter Steenbeek, Michael Townsley, and Christophe Vandeviver. 2020. "Minimum geocoding match rates: An international study of the impact of data and areal unit sizes." *International Journal of Geographical Information Science* 34 (7): 1306–1322. https://doi.org/10.1080/13658816.2020.1725015

Aneesh, A. 2009. "Global labor: Algocratic modes of organization." *Sociological Theory* 27(4): 347–370.

Baker, Reginald P. 2017. "Big data: A survey research perspective." In *Total Survey Error: Improving Quality in the Era of Big Data*, edited by Paul P. Biemer, Edith De Leeuw, Stephanie Eckman, Brad Edwards, Frauke Kreuter, Lars Lyberg, Clyde Tucker, and Brady West, 47–70. Hoboken: Wiley.

Bakker, Bart F. M. 2010. "Micro-integration: State of the art." In *Report on WP1. State of the art on statistical methodologies for data integration*, 77–107. Luxembourg: Eurostat.

Bakker, Bart F. M. 2012. "Estimating the validity of administrative variables." *Statistica Neerlandica* 66(1): 8–17. https://doi.org/10.1111/j.1467-9574.2011.00504.x

Biderman, Albert D., and James P. Lynch. 1991. *Understanding crime statistics: Why the UCR diverges from the NCS.* New York: Springer-Verlag.

Biemer, Paul P. 2010. “Total survey error: Design, implementation, and evaluation.” *The Public Opinion Quarterly* 74 (5): 817–848.

Biemer, Paul P. 2016. “Errors and inference.” In *Big Data and Social Science: A Practical Guide to Methods and Tools*, edited by Ian Foster, Rayid Ghani, Ron S. Jarmin, Frauke Kreuter, and Julia Lane, 265–297. Boca Raton, FL: CRC Press

Biemer, Paul P., and Ashley Amaya. 2021. Total error frameworks for found data. In *Big data meets survey science. A collection of innovative methods*, edited by Craig A. Hill, Paul P. Biemer, Trent D. Buskirk, Lilli Japec, Antje Kirchner, Stas Kolenikov, and Lars E. Lyberg, 133–161. Hoboken: John Wiley & Sons.

Biemer, Paul P., and Lars E. Lyberg. 2003. *Introduction to survey quality*. Hoboken: Wiley.

Biemer, Paul P., and Lars E. Lyberg. 2020. Total Survey Error. In *SAGE Research Methods Foundations*, edited by Paul Atkinson, Sara Delamont, Alexandru Cernat, Joseph W. Sakshaug, and Richard A. Williams. https://www.doi.org/10.4135/9781526421036888444

Boggs, Sarah L. 1965. “Urban crime patterns.” *American Sociological Review* 30(6): 899–908.

Bottomley, A. Keith, and Ken Pease. 1986. *Crime and punishment: Interpreting the data*. Milton Keynes: Open University Press.

Braga, Anthony A., and Ronald V. Clarke. 2014. “Explaining high-risk concentrations of crime in the city: Social disorganization, crime opportunities, and important next steps.” *Journal of Research in Crime and Delinquency* 51: 480–498. https://doi.org/10.1177/0022427814521217

Buil-Gil, David, Angelo Moretti, Samuel H. Langton. 2021. “The accuracy of crime statistics: Assessing the impact of police data bias on geographic crime analysis.” *Journal of Experimental Criminology*. https://doi.org/10.1007/s11292-021-09457-y

Bunge, Mario. 2006. *Chasing reality. Strife over realism*. Toronto: University of Toronto Press.

Chen, Yifei, Yuenan Li, and Jonathan Li. 2016. “Investigating the influence of tree coverage on property crime: A case study in the city of Vancouver, British Columbia, Canada.” *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* XLI-B2: 695-702. https://doi.org/10.5194/isprs-archives-XLI-B2-695-2016

Cochran, William G. 1953. *Sampling techniques*. New York: Wiley.

Connelly, Roxanne, Christopher J. Playford, Vernon Gayle, and Chris Dibben. 2016. “The role of administrative data in the big data revolution in social science research.” *Social Science Research* 59: 1–12.

Curtis, Ric. 2010. Getting good data from people that do bad things: Effective methods and techniques for conducting research with hard-to-reach and hidden population. In *Offenders and offending. Learning about crime from criminals*, edited by Wim Bernasco, 163–180. London: Willan.

De Kimpe, Lies, Michel Walrave, Thom Snaphaan, Lieven Pauwels, Wim Hardyns, and Koen Ponnet. 2021. “Research Note: An investigation of cybercrime victims’ reporting behavior.” *European Journal of Crime, Criminal Law, and Criminal Justice* 29(1): 66–78. https://doi.org/10.1163/15718174-bja10019

De Leeuw, Edith D., and Wim de Heer. 2002. Trends in household survey nonresponse: A longitudinal and international comparison. In *Survey nonresponse*, edited by Robert M. Groves, Don A. Dillman, John L. Eltinge, and Roderick J. A. Little, 41–54. New York: Wiley.

Deming, William E. 1944. “On errors in surveys.” *American Sociological Review* 9 (4): 359–369.

Favaretto, Maddalena, Eva De Clercq, and Bernice S. Elger. 2019. “Big Data and discrimination: Perils, promises and solutions. A systematic review.” *Journal of Big Data* 6: Article 12. https://doi.org/10.1186/s40537-019-0177-4

Federal Police. 2021a. Criminaliteitsstatistieken. Available online: http://www.stat.policefederale.be/criminaliteitsstatistieken/interactief/ (Accessed 15 April 2022).

Federal Police. 2021b. Veiligheidsmonitor 2018. Grote tendensen. Available online: http://www.moniteurdesecurite.policefederale.be/assets/pdf/2018/reports/Grote_tendensen_Analyses_VMS2018.pdf (Accessed 15 April 2022).

Groves, Robert M. 1989. *Survey errors and survey costs*. New York: Wiley.

Groves, Robert M., and Lars Lyberg. 2010. “Total Survey Error: Past, present, and future.” *Public Opinion Quarterly* 74 (5): 849–879. https://doi.org/10.1093/poq/nfq065

Groves, Robert M., Floyd J. Fowler Jr., Mick P. Couper, Eleanor Singer, and Roger Tourangeau. 2004. *Survey methodology*. Hoboken: Wiley.

Groves, Robert M., Floyd J. Fowler Jr., Mick P. Couper, James M. Lepkowski, Eleanor Singer, and Roger Tourangeau. 2009. *Survey methodology* (2nd ed.). Hoboken: Wiley.

Hansen, Morris H., William N. Hurwitz, and William G. Madow. 1953. *Sample survey methods and theory*. 2 vols. New York: Wiley.

Hardyns, Wim, Thom Snaphaan, and Lieven Pauwels. 2019. "Crime concentrations and micro places: An empirical test of the 'law of crime concentration at places' in Belgium." *Australian and New Zealand Journal of Criminology* 52(3): 390–410. https://doi.org/10.1177/0004865818807243

Harries, Keith D. 1981. "Alternative denominators in conventional crime rates." In *Environmental criminology*, edited by Paul J. Brantingham, and Patricia L. Brantingham, 147–165. Prospect Heights: Waveland Press.

Hilbert, Martin. 2016. "Big data for development: A review of promises and challenges." *Development Policy Review* 34(1): 135–174. https://doi.org/10.1111/dpr.12142

Hox, Joop J., Edith D. de Leeuw, and Don A. Dillman. 2008. The cornerstones of survey research. In *International handbook of survey research*, edited by Edith D. de Leeuw, Joop J. Hox and Don A. Dillman, 1–17. New York: Taylor & Francis Group.

Hsieh, Y. Patrick, and Joe Murphy. 2017. Total Twitter error: Decomposing public opinion measurement on Twitter from a total survey error perspective. In *Total Survey Error in Practice*, edited by Paul P. Biemer, Edith D. de Leeuw, Stephanie Eckman, Brad Edwards, Frauke Kreuter, Lars E. Lyberg, N. Clyde Tucker and Brady T. West, 23–46. Hoboken: Wiley.

Japec, Lilli, and Lars Lyberg. 2021. Big data initiatives in official statistics. In *Big data meets survey science. A collection of innovative methods*, edited by Craig A. Hill, Paul P. Biemer, Trent D. Buskirk, Lilli Japec, Antje Kirchner, Stas Kolenikov, and Lars E. Lyberg, 275–302. Hoboken: John Wiley & Sons.

Japec, Lilli, Frauke Kreuter, Marcus Berg, Paul Biemer, Paul Decker, Cliff Lampe, Julia Lane, Cathy O'Neil, and Abe Usher. 2015. "Big data in survey research: AAPOR task force report." *Public Opinion Quarterly* 79 (4): 839–880. https://doi.org/10.1093/poq/nfv039

Johnson, Timothy P., and Tom W. Smith. 2017. Big data and survey research: Supplement or substitute? In *Seeing cities through big data: Research, methods and applications in urban informatics*, edited by Piyushimita V. Thakuriah, Nebiyou Tilahun, and Moira Zellner, 113–126. Cham: Springer.

Kahneman, Daniel, Olivier Sibony, and Cass R. Sunstein. 2021. *Noise: A flaw in human judgment.* New York: Little, Brown Spark.

Karr, Alan F., Ashish P. Sanil, and David L. Banks. 2006. "Data quality: A statistical perspective." *Statistical Methodology* 3(2): 137–173. https://doi.org/10.1016/j.stamet.2005.08.005

Keusch, Florian, Sebastian Bähr, Georg-Christoph Haas, Frauke Kreuter, and Mark Trappmann. 2020. "Coverage error in data collection combining surveys with passive measurement using apps: Data from a German national survey." *Sociological Methods & Research* 1–38. https://doi.org/10.1177/0049124120914924

Kirkpatrick, Keith. 2017. "It's not the algorithms, it's the data." *Communications of the ACM* 60 (2): 21–23. https://doi.org/10.1145/3022181

Kish, Leslie. 1965. *Survey sampling*. New York: Wiley.

Kitchin, Rob, and Gavin McArdle. 2016. "What makes big data, big data? Exploring the ontological characteristics of 26 datasets." *Big Data & Society*, 3(1): 1–10. https://doi.org/10.1177/2053951716631130

Kitchin, Rob. 2014. "Big data, new epistemologies and paradigm shifts." *Big Data & Society* 1(1): 1–12. https://doi.org/10.1177/2053951714528481

Kitchin, Rob. 2015. "The opportunities, challenges and risks of big data for official statistics." *Statistical Journal of the IAOS* 31 (3): 471–481.

Klinger, David A., and George S. Bridges. 1997. "Measurement error in calls-for-service as an indicator of crime." *Criminology* 35 (4): 705–726.

Kreuter, Frauke. 2021. "What surveys really say." *Nature* 600: 614–615. https://doi.org/10.1038/d41586-021-03604-1

Lavrakas, Paul J. 2013. "Applying a total error perspective for improving research quality in the social, behavioral, and marketing sciences." *Public Opinion Quarterly* 77 (3): 831–850.

Liao, Dan, Marcus E. Berzofsky, G. Lance Couzens, Ian Thomas, and Alexia Cooper. 2021. Improving quality of administrative data: A case study with FBI's national incident-based reporting system data. In *Big data meets survey science. A collection of innovative methods*, edited by Craig A. Hill, Paul P. Biemer, Trent D. Buskirk, Lilli Japec, Antje Kirchner, Stas Kolenikov, and Lars E. Lyberg, 217–244. Hoboken: John Wiley & Sons.

Little, Roderick J. A., and Donald B. Rubin. 2002. *Statistical analysis with missing data.* 2nd ed. Hoboken: Wiley.

Lynch, James P., and John P. Jarvis. 2008. "Missing data and imputation in the Uniform Crime Reports and the effects on national estimates." *Journal of Contemporary Criminal Justice* 24 (1): 69–85. https://doi.org/10.1177/1043986207313028

Lynch, James P., and Lynn A. Addington, eds. 2007. *Understanding crime statistics: Revisiting the divergence of the NCVS and UCS.* Cambridge: Cambridge University Press.

Malleson, Nick, and Martin A. Andresen. 2015. "The impact of using social media data in crime rate calculations: Shifting hot spots and changing spatial patterns." *Cartography and Geographic Information Science* 42(2): 112–121. https://doi.org/10.1080/15230406.2014.905756

Malleson, Nick, and Martin A. Andresen. 2016. "Exploring the impact of ambient population measures on London crime hotspots." *Journal of Criminal Justice* 46: 42–63. https://doi.org/10.1016/j.jcrimjus.2016.03.002

Mosher, Clayton J., Terance D. Miethe, and Dretha M. Phillips. 2002. *The mismeasure of crime*. Thousand Oaks: Sage.

Neyman, Jerzy. 1934. "On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection." *Journal of the Royal Statistical Society* 97(4): 558–625.

Peeters, Rik, and Marc Schuilenburg. 2020. The algorithmic society: An introduction. In *The algorithmic society*, edited by Mark Schuilenburg and Rik Peeters, 1–15. New York: Routledge.

Proximus. 2020. Financiële resultaten van de Proximus Groep—Volledig jaar 2019. Available online: https://www.proximus.com/nl/news/2020/financial-results-q4-2019# (accessed on 15 April 2022).

Ratcliffe, Jerry H. 2021. “Policing and public health calls for service in Philadelphia.” *Crime Science* 10: Article 5. https://doi.org/10.1186/s40163-021-00141-0

Reid, Gill, Felipa Zabala, and Anders Holmberg. 2017. “Extending TSE to administrative data: A quality framework and case studies from Stats NZ.” *Journal of Official Statistics* 33 (2): 477–511. https://doi.org/10.1515/jos-2017-0023

Renzenbrink, Tessel. 2009. “Algoritmen voor een eerlijkere stad.” *Tada.City*, June 3. Accessed 12 March 2021. https://tada.city/nieuws/algoritmen-voor-een-eerlijkere-stad/

Richardson, Rashida, Jason Schultz, and Kate Crawford. 2019. “Dirty data, bad predictions: How civil rights violations impact police data, predictive policing systems, and justice.” *New York University Law Review* 94: 192-233.

Robinson, William S. 1950. “Ecological correlations and the behavior of individuals.” *American Sociological Review* 15 (3): 351–357.

Rummens, Anneleen, Thom Snaphaan, Nico Van de Weghe, Dirk Van den Poel, Lieven Pauwels, and Wim Hardyns. 2021. “Do mobile phone data provide a better denominator in crime rates and improve spatiotemporal predictions of crime?” *ISPRS International Journal of Geo-Information* 10(6): Article 369. https://doi.org/10.3390/ijgi10060369

Saris, Willem E. 2014. Total survey error. In *Encyclopedia of quality of life and well-being research*, edited by Alex C. Michalos. Springer. https://doi.org/10.1007/978-94-007-0753-5_3034

Saris, Willem E., and Frank Andrews. 1991. “Evaluation of measurement instruments using a structural modeling approach.” In *Measurement errors in surveys*, edited by Paul P. Biemer, Robert M. Groves, Lars E. Lyberg, Nancy A. Mathiowetz, and Seymour Sudman, 575–597. New York: Wiley.

Schober, Michael F., Josh Pasek, Lauren Guggenheim,Cliff Lampe, and Frederick G. Conrad. 2016. “Social media analyses for social measurement.” *Public Opinion Quarterly* 80 (1): 180–211. https://doi.org/10.1093/poq/nfv048

Skogan, Wesley G. 1977. “Dimensions of the dark figure of unreported crime.” *Crime & Delinquency* 23 (1): 41–50. https://doi.org/10.1177/001112877702300104

Snaphaan, Thom, and Wim Hardyns. 2021. “Environmental criminology in the big data era.” *European Journal of Criminology* 18(5): 713–734. https://doi.org/10.1177/1477370819877753

Statistics New Zealand. 2016. “Guide on reporting on administrative data quality.” Accessed 26 March 2021. https://www.stats.govt.nz/methods/guide-to-reporting-on-administrative-data-quality

Thakuriah, Piyushimita V., Nebiyou Tilahun, and Moira Zellner. 2017. Big data and urban informatics: Innovations and challenges to urban planning and knowledge discovery. In *Seeing cities through big data: Research, methods and applications in urban informatics*, edited by Piyushimita V. Thakuriah, Nebiyou Tilahun, and Moira Zellner, 11–45. Cham: Springer.

United Nations Economic Commission for Europe. 2014. “A suggested framework for the quality of big data.” Accessed 26 March 2021. https://statswiki.unece.org/download/attachments/108102944/Big%20Data%20Quality%20Framework%20-%20final-%20Jan08-2015.pdf

United Nations. 1969. *Methodology and evaluation of population registers and similar systems* (Series F, no. 15). New York: Statistical Office of the United Nations.

Van de Weijer, Steve G. A., Rutger Leukfeldt, and Wim Bernasco. 2019. “Determinants of reporting cybercrime: A comparison between identity theft, consumer fraud, and hacking. *European Journal of Criminology* 16(4): 486–508. https://doi.org/10.1177/1477370818773610

Wallgren, Anders, and Britt Wallgren. 2007. *Register-based statistics: Administrative data for statistical purposes.* Chichester: Wiley.

Warner, Barbara D., and Glenn L. Pierce. (1993). “Reexamining social disorganization theory using calls to the police as a measure of crime.” *Criminology* 31 (4): 493–517.

Wikström, Per-Olof H. 1991. *Urban crime, criminals, and victims. The Swedish experience in an Anglo-American comparative perspective.* New York: Springer.

Wikström, Per-Olof H., and Clemens Kroneberg. 2022. “Analytic criminology: Mechanisms and methods in the explanation of crime and its causes.” *Annual Review of Criminology* 5: 179–203. https://doi.org/10.1146/annurev-criminol-030920-091320

Zhang, Li-Chun. 2012. “Topics of statistical theory for register-based statistics and data integration.” *Statistica Neerlandica* 66 (1): 41–63. https://doi.org/10.1111/j.1467-9574.2011.00508.x