

*Stefan Harrendorf*

# Prospects, Problems, and Pitfalls in Comparative Analyses of Criminal Justice Data

## ABSTRACT

Official crime and criminal justice data are influenced by different substantive (e.g., victims' reporting rates), legal (e.g., offense definitions), and statistical (e.g., counting rules) factors. This complicates international comparison. The UN Crime Trends Survey, Eurostat's crime statistics, and the European Sourcebook of Crime and Criminal Justice Statistics try to enhance comparability and document remaining differences. The UN survey and Eurostat rely on the International Classification of Crimes for Statistical Purposes, which has potential but is not yet satisfactorily applied. The European Sourcebook provides the most detailed and best-verified data among the three. Even standardized data need to be compared with extreme caution. Crime levels are not a valid measure of crime in different countries, with the possible exception of completed intentional homicide. Total crime rates depend mainly on the internationally differing quality of police work. Comparisons of crime trends are less problematic but depend on the offenses under comparison being not defined too differently. Indicators expressed as ratios of different system-based values have increased comparability. Owing to immense differences in crime rates and criminal justice variables, mean crime rates for the world or Europe cannot be calculated. Country clusters need to be built very carefully.

International comparison of crime and criminal justice data is complex, almost impossible. Direct comparisons of national statistics on crime and

Electronically published March 9, 2018

Stefan Harrendorf is professor of criminology, criminal law, criminal procedure, and comparative criminal law and justice, University of Greifswald, Germany. He is grateful to Michael Tonry for comments on an earlier draft.

© 2018 by The University of Chicago. All rights reserved.  
0192-3234/2018/0047-0005\$10.00

criminal justice are unreliable not only because of language barriers and translation problems but more importantly because concepts and categories used in national statistics are not created to facilitate international comparisons but to meet information needs of criminal justice agencies, other administrative bodies, and politicians.

National statistics necessarily mirror the criminal law and criminal procedure in a given country. They cannot record behavior that is not considered criminal in a particular country. What is recorded as “theft” or “robbery” depends on how theft or robbery is defined in the criminal law and varies significantly between countries. Selection and definition processes in case processing also vary significantly, depending on the provisions of criminal procedure and criminal law and on practices that have developed around them (Wade 2006; Jehle, Smit, and Zila 2008). International or cross-national analyses cannot be credibly based on simple comparisons of national data. They need to use data from international surveys such as the European Sourcebook of Crime and Criminal Justice Statistics, the United Nations Crime Trends Survey, Eurostat data, and the Council of Europe’s SPACE data on imprisonment and community penalties.

The surveys try to enhance comparability of data in various ways including by standardizing offense definitions. They are based on secondary analyses of data originally collected for administrative purposes by national statistical systems. The surveys can never achieve perfect data comparability (with perplexing consequences for European criminal justice policy; de Bondt 2014). They can, however, carefully document differences and use various adjustments to enhance comparability. Huge amounts of metadata are collected in addition to statistical data (e.g., Aebi et al. 2014).

Researchers often make use of official national data for comparative analyses based on very general *ceteris paribus* assumptions, which are neither critically tested nor very convincing. For example, with regard to Cavadino and Dignan’s (2006) study on relations between political economy and imprisonment rates, David Nelken observed,

Cavadino and Dignan, like most of those comparing a large range of incarceration rates, spend little time on persuading us that crime rates are really the same in all the countries they are comparing. But it is this, the assumption that crime levels are “constant” in the places being compared, that sets the puzzle they are trying to solve. How can

some societies live with high crime rates without concomitant expansion of the prison realm? If countries with higher prison rates were actually dealing with higher threats from crime, this would not be news, and we could hardly say that we were fairly comparing levels of punitiveness. (Rather, we would be showing how neo-liberalism increases both crime and punishment.) On the other hand, it is strange that the good things about more inclusive welfare-oriented or egalitarian social-democratic societies do not also reduce the level, or severity, of crimes being committed, rather than only shaping the response to them. And since our ideal is presumably to live in places that have both low levels of punishment and low crime, it is a pity that this inconvenient point is passed over so quickly. (2010, p. 61)

See also Pakes (2015, p. 6). Indeed, studies show that there is at least some relationship between crime rates and incarceration rates, although it is restricted to the more serious crimes, especially, but not necessarily only, to completed homicide (for homicide: Lappi-Seppälä 2011; Lappi-Seppälä and Lehti 2014; Harrendorf 2017*b*; for other severe offenses in western Europe: Aebi, Linde, and Delgrande 2015).

There are innumerable other examples of careless use of official criminal justice data for testing comparative hypotheses (e.g., Churchill and Laryea [2017] on relations between ethnic diversity and crime; further examples below). As a general rule, one needs to keep in mind that data from different countries depend on different substantive, legal, and statistical factors, making meaningful comparisons very difficult (von Hofer 2000). For example, legal offense definitions differ significantly (Harrendorf 2012), as do statistical recording rules (e.g., whether offenses are recorded when first coming to police attention [input], only after verified [output], or at some intermediate point: Aebi 2008, 2010) and victim reporting rates (van Dijk, van Kesteren, and Smit 2007). The problems multiply concerning the total number of recorded criminal offenses; this is a black box with unknowable contents. The borderline between criminal and noncriminal behavior is drawn differently in different countries, leading to large overall total crime rate differences (Harrendorf 2011). Another confounding difficulty is the connection between overall crime rates and the quality of police work (Harrendorf 2017*a*), a subject I discuss below.

It is thus unwise simply to use national official data for offense categories such as burglary, robbery, homicide, or total crime to test hypoth-

eses about the relationship between crimes and other variables usually correlated with criminal behavior (as, e.g., in Rosenfeld and Messner [2009] or Buonanno, Drago, and Galbiati [2014]). As long as it is not clear that what is recorded as “burglary” is everywhere the same (which it is not; Tonry and Farrington 2005; Aebi et al. 2014, pp. 370–73, 390–92), we cannot use the data as a comparable measure. National crime data are fundamentally incomparable in many respects. Of course, it is possible to compare oranges with apples, as both are fruit. But we should know that the level of comparison is fruit only and not mistake apples for oranges.

Additional problems arise in relation to country clustering. In Rosenfeld and Messner (2009) and Buonanno, Drago, and Galbiati (2014), data from selected European countries are combined to form a country cluster called “EU” or “Europe.” The cluster in Rosenfeld and Messner (2009) consists of Denmark, France, Germany, Greece, Hungary, Ireland, Italy, the Netherlands, and Portugal. That in Buonanno, Drago, and Galbiati (2014) consists of Austria, France, Germany, Italy, the Netherlands, Spain, and the United Kingdom. In Europe, however, official crime rates per 100,000 population differ enormously between countries, without obvious relationship to the true incidence of crime, and depending on nationally distinctive legal, statistical, and case processing factors. For example, total crime levels in Europe in 2010 ranged between 476 per 100,000 population in Armenia and 14,671 in Sweden, a 30-fold difference (Aebi et al. 2014). Variation coefficients (standard deviations divided by the mean) are extremely high, and the mean of country results cannot be validly interpreted to represent all the different countries (Harrendorf 2012). It cannot measure “EU” or “European” crime levels.

Country clustering is a complex but feasible task, as some studies such as Smit, Marshall, and van Gammeren (2008) show. It needs, however, to be done in more sophisticated ways than just by taking a convenience sample of EU countries and using their combined values as a proxy for the EU or Europe.

Von Hofer (2000) sought to raise awareness of the vast methodological problems involved in international comparisons, but to my knowledge no publication has as yet provided a comprehensive overview of prospects, problems, and pitfalls associated with this type of work. That is my aim. In Section I, I briefly discuss reasons why one would want to compare crime and criminal justice data internationally. In Section II, I address the main data collection initiatives in Europe and worldwide,

including the European Sourcebook of Criminal Justice Statistics, the United Nations Crime Trends Survey, and Eurostat's data collection. I discuss the advantages and weaknesses of each and show possible ways to improve data quality and comparability. Since both the United Nations and Eurostat data collection rely on the recently developed International Classification of Crimes for Statistical Purposes for their offense definitions, I also discuss this classification system and show how users can assess data quality of these surveys (UNODC 2015). Section III explains what can and cannot be done with comparative crime and criminal justice data concerning analyses of crime levels and trends. Credible comparisons of crime levels are difficult to achieve. It is preferable to compare ratios that are completely system-based, that is, that consist of a numerator and a denominator both taken from official statistics (e.g., relative growth rates, ratios of convictions to suspects). Official data are a good source for learning about differences in criminal justice systems including, for example, case attrition, punitiveness, and police, prosecution, and court practices; however, official data should be used with extreme caution when the focus is on comparisons of crime problems in different countries. Section IV sums up and suggests how comparative projects using crime and criminal justice data should be planned and conducted.

### I. Why Compare?

"Why Compare?" is the title of the first chapter of David Nelken's important *Comparative Criminal Justice* (2010). The simplest reason is pure scientific interest and curiosity. Comparative work may add to the knowledge base of fundamental research. Criminal justice system differences and commonalities may as well be analyzed with policy questions in mind, for example, to assess use of alternative sanctions available in other systems to decide whether they should be adopted at home. One might seek ways to reduce the prison population by comparative analyses of punitiveness and its determinants, or one might look for functional equivalents in different countries for dealing with dangerous offenders. The aim may also be to identify shared principles and structures to gain deeper understanding of what is and should be going on in different criminal justice systems (Nelken 2010; Ebbe 2013; Pakes 2015; Tonry 2015).

Comparison of crime and criminal justice data, as a special aspect or variant of comparative research, is not necessarily an end in itself, but will often be motivated by policy concerns. De Bondt (2014) has shown

how heavily EU criminal justice policy depends on comparative data on offenses and legal instruments that are subject to harmonization. She has also shown that existing data fall utterly short of this aim.

Comparative studies may also provide additional insights into what national statistics actually mean. If, for example, comparative research shows that the recorded rate of total criminal offenses per 100,000 population in a given country is strongly dependent on the quality of police work, and not on the “reality” of crime, this is important information for a proper understanding of national crime rates (Harrendorf 2017a).

The goal of comparative studies might be to learn more about crime in international perspective (Heiskanen 2010; Aebi and Linde 2012). It might also be to learn more about criminal justice system reactions to crime or more generally about the work and functions of the different actors in the system (Blumstein, Tonry, and van Ness 2005; Smit, van Dijk, and Decae 2012). Both aims are important, yet I try to show why national crime and criminal justice system statistics are not the best place to look for comparative data on crime problems, but are an excellent starting point for efforts to delve into the differences between systems.

## II. How to Compare?

Comparison of crime rates and criminal justice processes in different countries is inherently difficult. Statistical systems exist to meet operational and information needs of criminal justice practitioners and agencies, not the data needs of national or comparative researchers. Statistical systems differ between countries as do criminal codes and criminal justice system processes. Definitions of specific offenses vary from place to place as do lines of demarcation between wrongful behaviors treated as crimes and others handled with administrative penalties or not at all. In this section, I discuss existing efforts to create international data systems that can be used for comparative purposes.

### *A. Problems of International Comparison*

Crime and criminal justice data are strongly dependent on national legal, statistical, and substantive characteristics that negatively affect comparability and cannot be fully controlled for, even with the aid of international surveys (Hofer 2000; Aebi 2008, 2010; Harrendorf 2012).

*Statistical factors* encompass the influence of different statistical practices, such as the use of input, intermediate, or output police crime recording practices. In input systems, cases are recorded when they are reported to the police. In output systems, cases are recorded only after investigations have been completed. Intermediate systems fall in between. Inevitably, information about the case is more detailed and more reliable in output systems. The operational definition of the crime the police use can change between when a case is reported and when it is solved. In output systems, cases in which the offender remained unknown or for which there was insufficient evidence may not be counted, leading to lower rates of recorded crime (Aebi 2008, 2010). Yet that is not necessarily so. In many countries, the police do not have legal authority to drop cases for legal or factual reasons (Elsner, Smit, and Zila 2008). Even cases involving unknown offenders may be referred to the prosecutor (Aebi et al. 2014, pp. 139–40).

Other rules influencing comparability involve counting of multiple or serial offenses or offenders. Imagine a series of 20 burglaries committed between 2013 and 2016 by an individual that were reported to the police in 2016 and for which investigations were completed in 2017. In some countries, this series of events would be recorded as one burglary, in others as 20 (Aebi et al. 2014, pp. 102–3). To complicate matters further, the recording might refer to the year in which the investigation was completed, the year in which the crimes were reported, or the different years in which they were committed.

*Legal factors* encompass national differences in criminal law and procedure. From a criminal law perspective, theft is not *Diebstahl* is not *κράζα* is not *varkaus* is not *kradzież*, although each of these words is simply a translation of the others. Offense definitions vary significantly between countries (Harrendorf 2012; Aebi et al. 2014, pp. 369–403), thereby further reducing comparability. For example, the Anglo-American concept of burglary does not exist in most continental European systems (Tonry and Farrington 2005, p. 3; Harrendorf 2012, p. 42). In some legal systems, subtypes of aggravated theft may be more or less—though never exactly—functionally equivalent.

Problems multiply with efforts to compare the total number of criminal offenses. The borderline between criminal behavior and deviant but noncriminal behavior is different in each system. Some countries exclude minor traffic offenses from criminal law coverage and crime statis-

tics. Others exclude certain minor property offenses or minor bodily injuries (Harrendorf 2012; Aebi et al. 2014).

The rules of criminal procedure also vary. One fundamental contrast is between countries with criminal justice systems governed by the “expediency principle,” which authorizes police, prosecutors, and judges to make decisions in individual cases for prudential reasons, and countries governed by the “legality principle” in which they do not have that authority. For example, in expediency principle countries such as the Netherlands and the United States, the police, the prosecution service, and the courts have authority to drop cases because of established policies, lack of a public interest in proceeding further, or the suspect’s fulfillment of specified conditions. Yet, as a reaction to increasing caseloads while criminal justice funding and staff levels remained the same or were even reduced, even countries that traditionally adhere to the legality principle (like Germany) have usually added important exceptions from this principle to their Codes of Criminal Procedure. This introduces a kind of expediency principle for some crimes, especially petty offenses. This results in final dispositions without formal convictions, thereby increasing attrition between police and conviction statistics. The extent to which these powers are available differs substantially between countries (Wade 2006; Elsner, Smit, and Zila 2008; Jehle, Smit, and Zila 2008).

Finally, *substantive* factors refer to operational, organizational, and behavioral differences unrelated to legal or statistical rules. The propensity of victims to report an offense to the police is one important factor. Another is the willingness of the police to record and investigate reported offenses. This depends in part on the amount of corruption in a given system. From a global perspective, police corruption is an important problem. According to Transparency International (2013, p. 11), 31 percent of respondents worldwide admit that they or someone from their household bribed a police officer in the preceding year. This is the highest rate among all institutions covered. Police were considered the second-most-corrupt institution, just after political parties (p. 16).

Table 1 summarizes factors that affect official crime statistics. A methodologically sound comparative study would need to control for them, eliminate their influence to the extent possible, and thoroughly disclose the remaining comparability problems. Cross-national comparisons based on unmodified data on recorded offenses with the same translated name (e.g., “theft”) are not credible.



TABLE 1  
Substantive, Legal, and Statistical Factors Influencing Comparability

| Substantive  | Legal                                      | Statistical   |
|--|--|---|
| “True” crime levels (including the “dark figure”)    | Offense definitions                        | Time of recording (input, output, or intermediate)        |
| “True” crime structure (including the “dark figure”) | Content and scope of the criminal law      | Counting rules for multiple current offenses              |
| Reporting by victims                                 | Legality or expediency principle           | Counting rules for serial offenses                        |
| Control activities by the police                     | Diversion, procedural decriminalization    | Counting rules for multiperson offenses                   |
| Police willingness to record offenses                | Plea bargaining; other agreed dispositions | Counting rules for persons suspected of multiple offenses |
| Clearance efforts                                    | Structure of criminal procedure            | Counting rules for multiple sanctions                     |
|  | Age of criminal responsibility             | Minimum age for inclusion in statistics                   |
|  | Sentencing laws and implementation         | Prison population counting rules                          |

The best way to get high-quality comparative data would be to collect them in different countries using identical methods, such as by relying on case files of the courts or prosecution services. This is what Lovett and Kelly (2009) did in their study of attrition in reported rape cases across Europe. This at least eliminates the statistical factors, although legal and substantive factors remain relevant. However, for financial reasons and owing to time restrictions, it is seldom possible to collect comparative data that way. In that case, international surveys provide the next-best sources of information.

The international surveys try to take account of legal and statistical influences. Data are collected by means of a questionnaire, trying to standardize respondents’ replies, thereby increasing comparability, and thoroughly documenting remaining differences (Aebi et al. 2014, pp. 17–21). Controlling for substantive factors would be desirable, too, but that is much more difficult. In theory, some of those factors also can be addressed, for example, by taking account of victim reporting rates from victimization surveys. This is not, however, yet part of the regular methodology of any international survey.

Full data comparability cannot be achieved and inherent limitations of secondary data analysis cannot be overcome. Data are obtained from

very different legal and juridical contexts, produced in criminal justice systems that vary substantially in quality and efficiency, and recorded according to differing statistical rules.

### *B. A Brief Inventory*

Before I discuss methodological details, a brief inventory of international surveys now available may be useful. Because my focus is on crime and criminal justice, I do not discuss surveys of victims, such as the International Crime Victim Survey (van Dijk, van Kesteren, and Smit 2007), or of offenders, such as the International Self-Report Delinquency Study (Junger-Tas et al. 2010). Table 2 provides information on the major international surveys and their coverage.

The oldest international crime survey, which began in 1950, was conducted by Interpol and based on police statistics. It was discontinued in 2006 because of serious quality issues (Interpol 2006; Rubin et al. 2008; Barberet 2009). The oldest ongoing survey is the United Nations Crime Trends Survey (hereafter, the UN Survey; <https://data.unodc.org/>), covering data since 1970 (Lewis 2012) and carried out by the UN Office on Drugs and Crime (UNODC). It is the only data collection with worldwide scope covering all stages of the criminal justice process (police, prosecution, courts, and prisons). Much of the data, especially concerning prosecution and courts, are not very detailed. The UN Survey long had problems with high nonresponse rates, especially from developing countries, leading to predominant coverage of countries from North America and Europe (Rubin et al. 2008; Alvazzi del Frate 2010). Response rates have recently increased, especially from Latin America, but coverage remains poor for Africa and some Asian regions (UN Economic and Social Council 2016, p. 33).

The UNODC also collects data on drug use, prices, seizures, and related subjects by means of its Annual Reports Questionnaire and the Individual Drug Seizure Reports (see the UNODC website). The annual reports also ask about persons brought into formal contact with the justice system in connection with drug-related offenses; comparable questions were removed from the UN Survey, probably to avoid double collection. Results on drug offenses are included in the World Drug Report (e.g., UNODC 2016*b*, pp. 101–2). Another data source for worldwide data, concerning only prisons, is the World Prison Brief, a database compiled by the Institute for Criminal Policy Research at Birkbeck Uni-

versity in London (Lewis 2012).<sup>1</sup> It is based primarily on data provided by national prison departments.

The “official” data collection in Europe is carried out by Eurostat for the European Union member states, European Free Trade Association countries, candidate countries, and potential candidate countries. Data are available from 2005 onward (Eurostat 2017*b*, p. 3). In 2014 Eurostat joined forces with the UNODC in collecting the data for those countries, relying on the UN Survey questionnaire, with supplements of relevance for European Union policy (pp. 4, 9–10). The scope of data collection is thus quite similar to that of the UN Survey with the same limitations, including meager coverage of prosecution and courts.

The most complete data collection initiative for Europe is the European Sourcebook of Crime and Criminal Justice Statistics (hereafter the European Sourcebook), which was patterned on the American Sourcebook of Criminal Justice Statistics (Killias 1995). It was started in 1993 under the auspices of the Council of Europe. After production of a draft model (Council of Europe 1995) and a first regular edition, covering 1990–95 (Council of Europe 1999), the project was continued by an experts group without further Council of Europe funding. Data collection for the fourth edition covering 2003–7 (Aebi et al. 2010) and the fifth covering 2007–11 (Aebi et al. 2014) paralleled European Union projects on specific topics (Jehle and Harrendorf 2010; Heiskanen et al. 2014).

Since 2011, the European Sourcebook has been organized as an association under German law (*eingetragener Verein*), that is, as a legal entity somewhat separate from the current composition of the experts group in charge of questionnaire development and data collection, collation, and validation. The European Sourcebook, like the UN Survey and Eurostat data collections, covers all criminal justice stages from police investigation through execution of sentences. It is the most comprehensive international survey in the number of variables on which data are collected and in the level of detail. It is wider in geographical coverage than the Eurostat survey but narrower than the UN Survey. Data are collected for all member states of the Council of Europe except for microstates.<sup>2</sup>

<sup>1</sup> The database is available at <http://www.prisonstudies.org/world-prison-brief-data>. The institute’s site is at <http://icpr.org.uk/>.

<sup>2</sup> For the fifth edition, data collection was also extended to Kosovo (UN/R 1244/99). Microstates are defined here as states with a population below 100,000. These are Andorra, Liechtenstein, Monaco, and San Marino.

TABLE 2  
International Crime and Criminal Justice Surveys

| Name   | Regional Scope                        | Thematical Focus  | Agency, Funding  | Latest Date Published Data Available                            |
|--|---------------------------------------|---|--|---|
| International Crime Victim Survey/European Survey on Crime and Safety                | World (selected countries);<br>Europe | Victimization; fear of crime; punitiveness; satisfaction with police            | Research group, differing sponsors (e.g., EU, Dutch, and UK governments) | 2004/5 (plus methodological pilot study 2010 in some countries) |
| International Self-Report Delinquency Study  | World (selected countries);<br>Europe | Juvenile delinquency and victimization  | Research group, differing sponsors (e.g., EU)                            | 2012–17 (differing by country)                                  |
| Interpol International Crime Statistics  | World                                 | Official crime data (police)  | Interpol   | 2004 (discontinued)   |
| UN Survey of Crime Trends and Operations of Criminal Justice Systems                 | World                                 | Official crime and criminal justice data (police, prosecution, courts, prisons) | UN Office on Drugs and Crime   | 2015  |
| World Drug Report (Annual Reports Questionnaire and Individual Drug Seizure Reports) | World                                 | Drug-related data (use, supply, crime, seizures)                                | UN Office on Drugs and Crime   | 2015  |
| World Prison Brief   | World                                 | Prison data   | Institute for Criminal Policy Research, Birkbeck, University of London   | 2017 (differing by country)                                     |

|   |   |   |  |      |
|---|---|---|--|------|
| Eurostat Crime and Criminal Justice Statistics                                      | Europe (EU member states, EFTA countries, candidate countries, and potential candidate countries) | Official crime and criminal justice data (police, prosecution, courts, prisons)                     | Eurostat (EU statistical office)   | 2015 |
| European Sourcebook of Crime and Criminal Justice Statistics                        | Europe (Council of Europe member states except microstates)                                       | Official crime and criminal justice data (police, prosecution, courts, prisons, probation agencies) | Research group, differing sponsors (e.g., EU, Dutch, Swiss, and UK governments, Council of Europe) | 2011 |
| Annual Penal Statistics of the Council of Europe (SPACE I + II)                     | Europe (Council of Europe member states)  | Prison data (SPACE I), probation data (SPACE II)  | University of Lausanne (funded by the Council of Europe)   | 2015 |
| European Judicial Systems: Efficiency and Quality of Justice                        | Europe (Council of Europe member states)  | Prosecution and court data  | European Commission for the Efficiency of Justice, Council of Europe                               | 2014 |
| Statistical Bulletin of the European Monitoring Centre for Drugs and Drug Addiction | Europe (EU member states, candidate countries, Norway)  | Drug-related data (use, supply, crime, seizures)  | European Monitoring Centre for Drugs and Drug Addiction, EU  | 2015 |

---

Full coverage has never been achieved but has come quite close. Thirty-nine countries were covered in the fifth edition. Because the United Kingdom includes distinct criminal justice systems for England and Wales, Scotland, and Northern Ireland, each is covered separately (Aebi et al. 2014, p. 9). The UN Survey and Eurostat do the same.

There are a few other multinational data collections in Europe. These include the annual SPACE (Statistiques Pénales Annuelles du Conseil de l'Europe; <http://www.coe.int/en/web/prison/space>) reports on correctional populations. SPACE is overseen and funded by the Council of Europe. Data on incarceration and prisons have been collected since 1983 (SPACE I; e.g., Aebi, Tiago, and Burkhardt 2017) and on noncustodial sanctions and measures since 1992 (SPACE II; e.g., Aebi and Chopin 2016). SPACE II collects data on the execution of community sanctions and measures under the supervision or care of probation agencies (Jehle and Harrendorf 2014).

In addition, a Council of Europe program develops reports on “European Judicial Systems: Efficiency and Quality of Justice.” These are regularly prepared by the European Commission for the Efficiency of Justice (e.g., 2016). It evaluates the quality and efficiency of European court systems including the work of prosecution agencies.

A final regular data report is the Statistical Bulletin of the European Monitoring Centre for Drugs and Drug Addiction (e.g., 2017). The focus is comparable to that of the UN World Drug Report (UNODC 2016*b*). Hence, the only crime and criminal justice data collected pertain to drug offenses, legal responses to them, and seizures; the overall focus is on drug use, supply, health consequences, and treatment. Finally, regarding reconviction data, there is no regular comparative survey, but pilot work has been done (Wartna and Nijssen 2006; Wartna et al. 2014).

In the methodological discussions that follow, I focus only on the three general surveys: the UN Survey, Eurostat, and the European Sourcebook. Since Eurostat and the UN basically use the same methodology and questionnaire, comparisons are between two different approaches.

### *C. General Methodology*

In order to standardize national replies as much as possible, surveys not only collect absolute numbers of different crimes or sanctions imposed but rely on metadata concerning definitions of offenses, prosecutorial case disposition methods, sanctions and measures, and data recording rules. Ideally, the questionnaires used for data collection are well designed and take

account of knowledge from earlier comparative studies and the knowledge of each expert involved concerning his or her national system.

1. *Standard Definitions.* Standard offense definitions are used, which means that national data must be adjusted to fit the definitions. The European Sourcebook augments the standard definition with lists of foreseeable variations and provides clear rules to include or exclude them. Table 3 sets out intentional homicide from the fifth edition questionnaire as an example. Respondents are informed that they should follow the rules as closely as possible and adapt the data they report accordingly. In Germany, for example, assault leading to death (i.e., intentional assault that unintentionally, but negligently, causes the death of the victim) is a distinct offense that is separately recorded in national statistics; according to German law, it would not be considered intentional homicide. Yet to be consistent with Anglo-American concepts of homicide, which often consider such cases as intentional killings, the standard definition requires respondents to add cases of assault leading to death to the officially recorded cases of intentional homicide. This is what is done when Germany replies to the questionnaire.

TABLE 3  
Standard Definition of Intentional Homicide, European Sourcebook

| Intentional Homicide: Standard Definition: Intentional Killing of a Person |       |                       |       |
|--|-------|-----------------------|-------|
| Indicate if Included in or Excluded From:                                  |       |                       |       |
| Police Statistics  |       | Conviction Statistics |       |
| Incl.  | Excl. | Incl.                 | Excl. |
| Include the following:   |       |                       |       |
| • assault leading to death   |       |                       |       |
| • euthanasia   |       |                       |       |
| • infanticide  |       |                       |       |
| • attempts   |       |                       |       |
| Exclude the following:   |       |                       |       |
| • assistance with suicide  |       |                       |       |
| • abortion   |       |                       |       |
| • negligent killing  |       |                       |       |

SOURCE.—Fifth edition questionnaire of European Sourcebook (Aebi et al. 2012, p. 10).

It is, of course, not always possible to adapt national data to fit the standard definitions and to follow all of the inclusion and exclusion rules. Respondents are accordingly asked to indicate whether they were able to follow the rules, as table 3 illustrates. The system of standard definitions has a prescriptive purpose—to achieve maximum standardization—but also aims to document remaining differences. Conformity with the inclusion and exclusion rules is reported separately for police and conviction statistics. Usually, it is much easier to adhere to the standard definition on the police level than on the conviction level, as conviction statistics depend more strongly on legal offense categories. For some offenses, especially theft of a motor vehicle, burglary, and domestic burglary, this results in a large proportion of countries being completely unable to provide conviction data, typically because these categories do not exist as separate criminal code offenses and are also not separately identifiable in statistics (Harrendorf 2012, pp. 39, 42).

Eurostat and the UN Survey until recently also used quite similar standard definitions but did not have sophisticated systems for identifying variations. The surveys simply asked “Was this definition applied in your country?” and provided space for comments. Most respondents tick “yes” or “no” but provide no comments.

More recently, the UNODC and Eurostat switched to another classification model, the International Classification of Crimes for Statistical Purposes (ICCS). The idea was to create an event-based classification system that did not rely on criminal law definitions in order to enhance comparability (UNODC 2015, p. 8). The classification system is comprehensive and aims at classifying all possible criminal acts. The earlier systems of standard definitions were selective, focusing only on specific crimes and the total number of offenses.

Apart from this criterion of exhaustiveness, the ICCS includes additional criteria of mutual exclusivity of classifications and statistical feasibility, that is, the capacity in principle of identifying these acts in national statistics (UNODC 2015, pp. 12–13). The classification is hierarchical, involving one top level and up to three further sublevels of increasing detail. The 11 level 1 categories are set out below:

- acts leading to death or intending to cause death,
- acts leading to harm or intending to cause harm to the person,
- injurious acts of a sexual nature,
- acts against property involving violence or threat against a person,



- acts against property only,
- acts involving controlled psychoactive substances or other drugs,
- acts involving fraud, deception, or corruption,
- acts against public order, authority, and provisions of the state,
- acts against public safety and state security,
- acts against the natural environment, and, finally,
- other criminal acts not elsewhere classified.

For the level 2 category of intentional homicide, for example, the ICCS gives this standard definition: “Unlawful death inflicted upon a person with the intent to cause death or serious injury” (UNODC 2015, p. 33).

Like the European Sourcebook approach, the ICCS provides inclusions and exclusions for each offense, but they are often more detailed. For intentional homicide:

**Inclusions:** Murder; honour killing; serious assault leading to death; death as a result of terrorist activities; dowry-related killings; femicide; infanticide; voluntary manslaughter; extrajudicial killings; killings caused by excessive use of force by law enforcement/state officials.

**Exclusions:** Death due to legal interventions; justifiable homicide in self-defence; attempted intentional homicide (0102); homicide without the element of intent is non-intentional homicide (0103); non-negligent or involuntary manslaughter (01031); assisting suicide or instigating suicide (0104); illegal feticide (0106); euthanasia (0105).

The codes in the exclusion lists refer to other offense categories of the ICCS. They result from implementation of the two principles of mutual exclusivity and exhaustiveness. The inclusion and exclusion lists are elaborated in a large number of footnotes, providing standard definitions for many of the items mentioned in the lists (like “murder” or “honour killing”). The classification system, finally, requires recording of several disaggregation variables, concerning the event, the victim, the perpetrator, and some further descriptive data (UNODC 2015, p. 21).<sup>3</sup> The document explaining and presenting the classification system is 130 pages long.

<sup>3</sup> The variables for events are as follows: attempted/completed; type of weapon used; situational context; geographic location; date and time; type of location; motive; cybercrime related; reported by. For the victim: sex; age; age status (minor/adult); victim-perpetrator relationship; citizenship; legal status (natural/legal person); intoxication status; economic sector of business victim. For the perpetrator: sex; age; age status (minor/adult); victim-perpetrator relationship; citizenship; legal status (natural/legal person); intoxication status; economic activity status; recidivist status. For further descriptive data: threats included;

The ICCS is very ambitious. However, everything depends on how it is used in practice. It is not meant to be a legal classification system; it is not a model penal code for all countries in the world. Hence, the ICCS cannot alter national legal factors influencing data comparability. Even an “event-based” classification system does not change dependency of national statistics on national criminal law. Of course, the influence of such factors can be reduced if countries comply with the classification system. But that is also true of the standard definition system used by the European Sourcebook, which uses classifications as much event-based as the ICCS.

Perhaps the UN and the European Union will have the influence and power in coming years to persuade more and more countries to use the ICCS categories in national statistical systems. This would be an important improvement. It is, however, not realistic to expect that any country with an existing statistical system will ever completely switch to the ICCS. That would render all existing national data incomparable with new data. Apart from that, statistics would largely lose their connection to national laws. This would be very helpful for international comparisons but would create serious problems in the national contexts in which the data are primarily used.

National statistical offices could parallel code their national data according to the ICCS, that is, use the national and the international classifications in parallel. This would mean substantial additional work for practitioners and statistical systems. It is more realistic to expect that national statistical offices will often cross-code data for the ICCS from national categories (Jehle 2012*b*, p. 138). This is not different from the current approach for the European Sourcebook.

In practice, both UNODC and Eurostat (which uses the UNODC questionnaires plus some additional modules; Eurostat 2017*b*, pp. 4, 9) continue to employ a simplistic system for collection of metadata. As before, they inform respondents about the required definitions. Regarding compliance, they ask “Do data comply with this definition (yes/no)?” again with the possibility to provide comments (UNODC 2016*a*). Most respondents are not interested in providing lengthy textual explanations; the added value of the ICCS, as now used, is almost certainly minimal. What is the use of a sophisticated classification system if there is no prac-

---

aiding/abetting included; accessory/accomplice included; conspiracy/planning/preparation included; incitement to commit crime included.

tical way to monitor compliance? The European Sourcebook system is—as of now—superior in practice, but the ICCS has great potential.

Standard definitions are needed both for offenses and for other variables that are subject to variation because of legal and statistical factors. A comparative study, for example, recently showed that the concept of a “cleared case” in police statistics differs significantly between countries (Brå 2015). This could be the first, initial suspicion, as in Germany, or it could require a suspicion so strong that it justifies an indictment or an equivalent decision by the prosecuting authority as in Sweden. The number of cleared cases is not recorded in any existing international survey, but the same problems arise concerning the definition of the “suspect”; clearance in general means that a suspect has been identified. No survey yet provides such a definition.

For other concepts, such as “conviction,” both the European Sourcebook and the UN Survey feature standard definitions. In the European Sourcebook, definitions are structured similarly to offense definitions and are accompanied by lists of ambiguous cases with clear instructions about inclusion and exclusion. Respondents are asked for each rule whether they followed it. The fifth edition questionnaire provides that “conviction means that the person was found guilty, according to the law, of having committed an offence and therefore has a criminal record.” Inclusions are “court convictions; sanctions imposed by the prosecutor (or by the court, but on application of the prosecutor and without a formal court hearing) that lead to a formal verdict and count as a conviction (e.g. penal order, *Strafbefehl*); convictions of minors in regular criminal proceedings . . .; convictions of minors in juvenile criminal proceedings.” Exclusions are “sanctions imposed by the prosecutor that do not lead to a formal verdict and do not count as a conviction (e.g. conditional disposals); sanctions/asures imposed by the police; sanctions/asures imposed by other state bodies; reactions on criminal or deviant behaviour of minors imposed in family court or youth welfare proceedings” (Aebi et al. 2012, p. 69).

The UN Survey definition is as follows: “‘Persons Convicted’ means persons found guilty by any legal body authorized to pronounce a conviction under national criminal law, whether or not the conviction was later upheld. The total number of persons convicted should also include persons convicted of serious special law offences but exclude persons convicted of minor road traffic offences, misdemeanours and other petty offences” (UNODC 2016a). The UN Survey definition thus also involves

inclusions and exclusions, but compliance is verified only by asking about it and providing opportunity to give comments. This lack of information about compliance with definitions is a fundamental problem for both the UN Survey and Eurostat because they mainly use the same questionnaire. As with offense definitions, this is a crucial weakness.

2. *Rules of Statistical Recording.* Another system is used for rules relating to the statistical counting of offenses, offenders, convictions, and other factors. In some cases, a preferential counting unit is given. For example, concerning prosecution statistics, the European Sourcebook questionnaire states, “The counting unit required here is the case (in the sense of proceedings relating to one person only) dealt with by prosecuting authorities” (Aebi et al. 2012, p. 50). Afterward, respondents can choose which counting unit they actually used (case, proceedings, person, other). In many other cases, no preferential method for counting cases is specified. For example, the European Sourcebook questionnaire simply documents how multiple offenses by the same offender, or single offenses committed by multiple offenders, are counted (as one case or as multiple cases). Regarding statistical differences, this documentary method (without additional standardization) is used not only in the European Sourcebook but also in UN and Eurostat data collections because it is assumed that such statistical rules cannot be changed afterward.

While this is true for aggregate statistical databases, even for these traditional national statistics, different counting rules may already be applied. For example, national statistics might include data on the input of cases, on pending cases, and on the output of cases at the police level. If international surveys now simply ask about the stage of investigation the data refer to, without providing information about the preferred stage, they miss an important opportunity for further standardization. Since increasing numbers of national statistical systems are based on an electronic database that contains all individual recorded cases, with the option to aggregate statistics as needed, possibilities for later adaptations of data to comply with counting rules have now even increased.

Another reason not to standardize statistical counting rules might be that it would seem arbitrary to choose one preferred recording method among several in use. Yet this is not true: The preferred method can be identified easily by using two simple rules. First, if one of several possible rules of statistical recording provides better data quality, that rule is to be preferred. Second, if there is no superior rule, the statistical rule that is most commonly applied should be used.

For police data, output statistics in principle deliver better-quality data, since police then have more information on which to base classifications. Some cases, however, that appear in input statistics may not appear in output statistics, especially in countries where cases for which an offender was not found or which were evidentially insufficient are not recorded in the output (Aebi 2008, p. 208). Intermediate systems are the most problematic, because it will be unclear how far the police have advanced in investigating a crime. In such statistics this will be subject to variation (Aebi 2008, p. 208).

The ICCS takes no account of statistical recording practices, for example, concerning counting of multiple offenses or offenders. The reason is that the ICCS is a crime classification system, not a full model for a statistical recording system. However, this seriously reduces the potential of the ICCS to produce truly comparable data in the future.

There are no other international approaches aiming at standardizing counting rules. The UN *Manual for the Development of a System of Criminal Justice Statistics* acknowledges that different counting units make it difficult to compare national crime and criminal justice statistics but does not prescribe standardization (UN Statistics Division 2003, p. 19).

3. *Respondents.* A crucial question is who completes the questionnaires. In all international data collections, usually one questionnaire is sent to each country.<sup>4</sup> For the UN Survey and Eurostat, the questionnaires go through official channels and the respondents are state employees working in national statistical offices, the police, the prison administration, and other government agencies.

The European Sourcebook, because it is an independent research enterprise, can be more flexible. Many respondents are researchers at universities or other research institutions. The main selection criterion is expertise concerning the national criminal justice system. The contact persons are referred to as national correspondents. This capacity to choose the best-qualified national correspondent leads to improved data quality. Unfortunately, it is not always possible to find a national correspondent in every country. This is especially a problem for countries in which criminology is not well developed in universities. In these coun-

<sup>4</sup> In some cases, there might also be more questionnaires per country, as for the United Kingdom, which—because of the different criminal justice systems—receives three questionnaires, one for England and Wales, one for Scotland, and one for Northern Ireland.

tries, all depends on the willingness of national statistical offices to cooperate. Problems finding a qualified correspondent for the European Sourcebook are intensified at least for some countries because it has not been possible so far to remunerate correspondents, even though the work involved is substantial.

4. *Presentation of Data.* Data from international surveys are presented as rates per 100,000 population, or as percentages of a larger total (e.g., the percentage of women among offenders registered on the police level). Yet it is of almost equal importance that the raw data (the absolute values and all metadata) are made publicly available, in order to allow users to better assess the quality of the data and for more precise computation of indices, change rates, and the like. This is especially important for the European Sourcebook, in which rates per 100,000 population are truncated and presented without decimals or with only one decimal. All three data collections make the absolute values for the collected data available. For the UN Survey and the European Sourcebook, all metadata are readily accessible. For Eurostat, the individual metadata are also available on the internet, but finding them is somewhat difficult.<sup>5</sup>

#### D. *Quality Assessment*

Complex data collections like these need sophisticated and thorough data validation routines. In principle, three different types of checks can, and should, be carried out: consistency, trends, and other sources checks (Harrendorf and Smit 2010, pp. 146–47).

*Consistency* (or internal validity) checks aim at internal consistency. Some of the reported data are interrelated; therefore, some general rules can be fixed. For example, all subcategories within a breakdown need to be smaller than the total (e.g., the number of female offenders needs to be smaller than the total number). Where a total is broken down completely into subcategories, these subcategories should usually sum to 100 percent, yet this is not necessarily the case. Imagine a breakdown by sanction type: in many countries, a principal sanction rule is applied for the total (i.e., only the most severe sanction is counted), but it is not always also applied for the subcategories. In such a country, the subcat-

<sup>5</sup> See <https://data.unodc.org/> for the UN Crime Trends Survey, <http://ec.europa.eu/eurostat/web/crime/database/> for Eurostat, and <https://wp.unil.ch/europeansourcebook/data-base/> for the European Sourcebook. The direct link to the Eurostat national metadata is [http://ec.europa.eu/eurostat/cache/metadata/Annexes/crim\\_esms\\_an2.pdf](http://ec.europa.eu/eurostat/cache/metadata/Annexes/crim_esms_an2.pdf).

egories would add up to more than 100 percent if some offenders receive multiple sanctions in a single court decision.

Criminal procedure can be envisioned as a process of case selection and definition during which cases are successively filtered out. This is often visualized as a funnel (e.g., Cole and Smith 2011, p. 14; Jehle 2015, p. 9). A conceptualization more consistent with the process is a picture of a series of gateways and gatekeepers (Harrendorf 2017*a*). Attrition is a natural occurrence and happens at and between different stages of the process, for example, within the police level between input and output, between the police and prosecution, within the prosecution level between input and output, and between the prosecution and the court (Harrendorf, Jehle, and Smit 2014).

Numbers can be expected to decrease during the process, with the number of offenses recorded at the police level being larger than the number of suspects found, the number of suspects larger than the number of persons indicted, that number larger than the number of persons convicted, and that number larger than the number of persons sent to prison. This can also be used for internal consistency checks, yet once again some deviations are possible. For example, the number of suspects can be larger than the number of recorded offenses if the following three criteria are cumulatively fulfilled: for a given offense clearance rates are high, the offense is at least sometimes committed by multiple offenders, and counting rules treat offenses by multiple offenders as one offense but register each suspect separately. The same effect can occur for offenses with high clearance rates if offenses are registered on an input basis and suspects are registered only later when they are identified. In such a system, an offense might be recorded in 2014 but the suspect not be identified until 2015. Because of this time lag, offender ratios (suspects per 100 registered offenses) above 100 percent are possible. Homicide data often show offender ratios above 100 percent (Harrendorf 2017*a*).<sup>6</sup>

Internal validity checks can also identify implausible outliers outside the acceptable range. Because of the large variation in international data, this is, however, restricted to extreme cases.

*Trend* checks look at time series for a certain variable and look for odd “jumps” in the magnitude of a variable. When this occurs between survey

<sup>6</sup> For further examples of consistency checks, see Harrendorf and Smit (2010, pp. 146–47).

waves, this may indicate a change in the way a category was cross-coded from the national system or a different handling of the inclusion and exclusion rules of offense definitions. Or there might be changes in criminal or procedural law or in the rules of statistical recording. Of course, even large changes could result from substantive factors such as changes in priorities or efficiency of police work or in the true incidence of crime.

Finally, *other sources* checks compare results with values a country reported to another international survey for a similar variable. Data on the prison population reported to the European Sourcebook could, for example, be compared to UN Survey data, SPACE I data, and World Prison Brief data. Trend and other sources checks can show huge differences between data reported for one survey wave or to one source and another survey wave or source because of, among other things, a different understanding or handling of standard offense definitions. Malby (2010a, p. 57) gives an excellent example concerning huge differences in the trend for drug trafficking and the total number of drug offenses in Germany according to several international surveys.

For the fifth edition of the European Sourcebook, internal consistency, trends, and other sources checks were programmed in SPSS, with an Excel output file per country listing all inconsistencies found. These then need explanation or correction. A validation system is used in which different members of the experts group act as regional coordinators for groups of countries. Each regional coordinator is in charge of data validation for his or her countries. The validation file helps here, as the regional coordinator can decide which inconsistencies can already be explained by the existing metadata (e.g., a major change in criminal law) or by other factors, and which need to be reported back to the national correspondent. The correspondent will then receive a list of inconsistencies that need correction or explanation. A second validation system, based on responsibilities for specific chapters of the sourcebook (police, prosecution, courts, corrections, offense definitions), leads to further improvement of quality control.

Eurostat also reports that it applies consistency, trend, and other sources checks and resolves inconsistencies together with the national contacts, either by correction of the value or by explanation (Eurostat 2017a, p. 27). Since the UN Survey and Eurostat share the data collected, this automatically leads to a validation of UN data for these countries. To what extent UN data for other countries are validated is not clear. Yet it can be expected that—differently from earlier years (cf. Harrendorf



2013*b*)—at least some validation is carried out. Some internal, unpublished reports show that there have been some efforts to validate UN data at least since 2009/10 (Harrendorf 2010). On the basis of the available information, the Europe Sourcebook group applies the strictest validation procedures.

Harrendorf (2012, 2013*b*) showed that there is an additional possibility for data validation, relying on variation coefficients of reported data. There is huge variation in reported international data even for Europe in total levels of registered crimes per 100,000 population and in other variables. Therefore, it is not easy to say when a value is totally outside the acceptable range. Harrendorf found out that there is an almost perfect linear relationship between the means of offense, suspect, and conviction rates for different offenses in the European Sourcebook and UN Survey data and the respective standard deviations. On the basis of this assumption, exceptionally high variation coefficients for particular offenses or years indicate a problem with data quality or comparability. Harrendorf (2013*b*) was able to confirm that variation coefficients for unvalidated data were, in principle, higher than for validated data. This allows use of variation coefficients as an additional validation tool on a summary (not country-specific) level. Exceptionally high variation coefficients hint at a data problem and indicate that a closer look at the data is necessary.

As long as international surveys publish all their raw data (absolute values and all metadata), these checks can also be carried out after publication (e.g., Harrendorf and Smit 2010, pp. 146–47). Users of international surveys should take a critical look at data before using them comparatively, keeping in mind the quality indicators discussed. If a time series has odd and sudden increases or decreases between adjacent years, it is necessary to find out why. If the data look strange, they should not be trusted! Identified errors should be corrected, for example, by replacing the data with data from another international survey for an (almost) identical variable. For time series, interpolation might also be an option. There may, however, be a plausible explanation for odd values, so the metadata should be examined before discarding a value.

#### *E. Shortcomings and Possible Improvements*

Even the best methodology cannot eliminate all negative effects on data comparability of legal and statistical differences between countries.

The influences of substantive factors are not even taken into account by current surveys. There is also no useful information available on the relative intensity of the effects of different legal, statistical, and substantive factors on data collected in different countries. It is doubtful that this information will ever be obtained. Thus, it is impossible to use statistical weighting to adjust for substantive, legal, and statistical differences between countries (Aebi 2008, p. 217).

The UN Survey and Eurostat do not have a system to record deviations from standard definitions and their inclusions and exclusions. This could easily be changed and would substantially improve data quality and comparability. The European Sourcebook system, in which correspondents are asked to provide explicit information about inclusions and exclusions, could serve as a model. Apart from that, legal factors, especially offense definitions, are addressed adequately in all three surveys. More troubling is how statistical recording rules are handled. Here, it would be useful to switch to prescribed or at least preferred counting rules combined with precise questions regarding compliance with them.

Finally, international surveys should begin systematically to collect and collate data on substantive factors that influence data comparability. This is of immense importance, as the 30-fold difference between official total crime rates in Sweden and Armenia demonstrates; it cannot be explained by legal and statistical differences. Substantive factors such as differing rates of victim reporting or police quality and efficiency are even more influential; likewise for other variables. Some essential data, such as the prevalence and incidence of victimization, reporting rates, and citizen satisfaction with police work, can be taken from victimization surveys. The European Sourcebook has incorporated some data from international and, in the fifth edition (Aebi et al. 2014), national victimization surveys. This approach needs to be made more systematic and focused on key substantive factors.

While, in principle, the proposed changes should improve comparability of national data in international surveys, they pose formidable management challenges (Harrendorf 2012). Data quality is now strongly dependent on how thoroughly national correspondents complete questionnaires and how much effort they invest in enhancing data comparability by adhering to standard definitions and rules. Doing so increases workloads as respondents may need to combine data from several national statistical categories to conform to standard definitions. However, because of time restrictions or methodological misunderstandings, corre-

spondents sometimes make mistakes or fail to follow rules. Some use more or less unmodified data from national statistics, although modifications were necessary and possible. Others misunderstand definitions or inclusion and exclusion rules. Such omissions, misunderstandings, and errors cannot always be identified.

It is therefore important for project managers to stay in close contact with national correspondents while questionnaires are being completed and validated. This is why the European Sourcebook uses regional coordinators with responsibility for only a few countries. It is also why conferences of all national correspondents held during the data collection phase have proven important.

### III. Making Sense of Comparative Data

Crime and criminal justice data are produced by criminal justice practitioners and measure the quality and quantity of their work. The data do not measure the reality of crime, the true number of acts, or omissions that violate criminal laws in a given country. There is no constant or knowable relationship between crimes that are committed and crimes that come to the attention of criminal justice system agencies. This is a problem for any national research study but presents greater difficulties for comparative studies because of national differences in legal, statistical, and substantive factors. And just as the ratio between committed crimes and those that come to the attention of the police varies from year to year and between offenses, it also varies between countries.

#### *A. Rate Comparisons*

That Sweden in 2010 had the highest European crime rate per 100,000 population does not mean that Sweden is the most dangerous country in Europe. There may be many other explanations. First, the criminal law may be used extensively, defining petty wrongs or administrative violations as criminal that are handled administratively or not at all in other countries. Second, statistical counting rules may inflate crime rates. Police use of input rather than output statistics tends to produce higher crime levels (Aebi 2008, 2010). Crime rates also increase if a principal offense rule is not used—recording each of several simultaneous offenses separately, for example, robbery and murder committed in the same event—rather than only the most serious. Likewise, separately recording each of a series of similar offenses inflates crime rates. Sweden

does all of these things (Brå 2015, pp. 9–11). High crime rates may also be based on substantive factors such as a higher victim reporting rate or greater likelihood that the police record crimes (von Hofer 2000).

For these reasons, it is often advised not to make direct comparisons of rates per 100,000 but to use trend comparisons (Aebi et al. 2014, p. 21; Eurostat 2017*b*, p. 36). This is, however, not the full picture. It is important to find out what internationally varying crime levels mean in order to understand when direct comparison is feasible and when it is not. There is, for example, evidence that total crime rates are mainly a function of the quality of police work (Harrendorf 2017*a*). The better the police performance, the higher the crime rate. This is shown in figure 1.

There is an almost perfect linear relationship in figure 1 between police performance in a country and the total number of recorded offenses; the correlation coefficient is 0.80 ( $R^2 = 0.65$ ).<sup>7</sup> Diverse factors explain this relationship. When the police are known or seen to perform well, more incidents are reported and more cases are recorded because of proactive police activities. Low levels of corruption mean that suspects cannot often avoid recording and subsequent prosecution in exchange for money or other favors. The strong correlation suggests that legal or statistical factors are secondary (for further details, see Harrendorf [2017*a*]). Similar results occur for minor offenses. The correlation coefficient for the relation between the Police Performance Index (PPI)<sup>8</sup> and the theft rate per 100,000 population in 2010 was also 0.80.

For severe offenses, there is no clear relation between crime rates and the PPI. The correlation coefficient for robbery is 0.16, for the total of attempted and completed homicides –0.09, and for attempted homicide

<sup>7</sup> Albania and Sweden were excluded as outliers. Albanian reports of total offenses to the European Sourcebook seem to include only cleared cases. For Sweden the number of total recorded offenses is artificially high, partly because of statistical counting rules (Brå 2015).

<sup>8</sup> The PPI (Pare 2014) is based on five variables (reporting of crimes by victims, satisfaction with the police reaction, general satisfaction with police work, businesses' view of police trustworthiness with respect to law enforcement, and victimization by corruption). Four of these variables were taken from the International Crime Victims Survey for 2004/5 (van Dijk, van Kesteren, and Smit 2007) and one (businesses' trust) from the World Economic Forum survey (Porter et al. 2004). The index is largely consistent with one proposed by van Dijk (2008) but replaces the homicide clearance rate with the corruption measure, which is an improvement since there is significant variation in the definition of what "clearance" means (Pare 2014; Brå 2015). The PPI was used unmodified even though the index uses data from 2003 and 2004. This assumes that there were no significant changes in police performance between these years; this is confirmed by the correlation coefficient for the PPI with crime data for 2004 being similar to that for 2010.

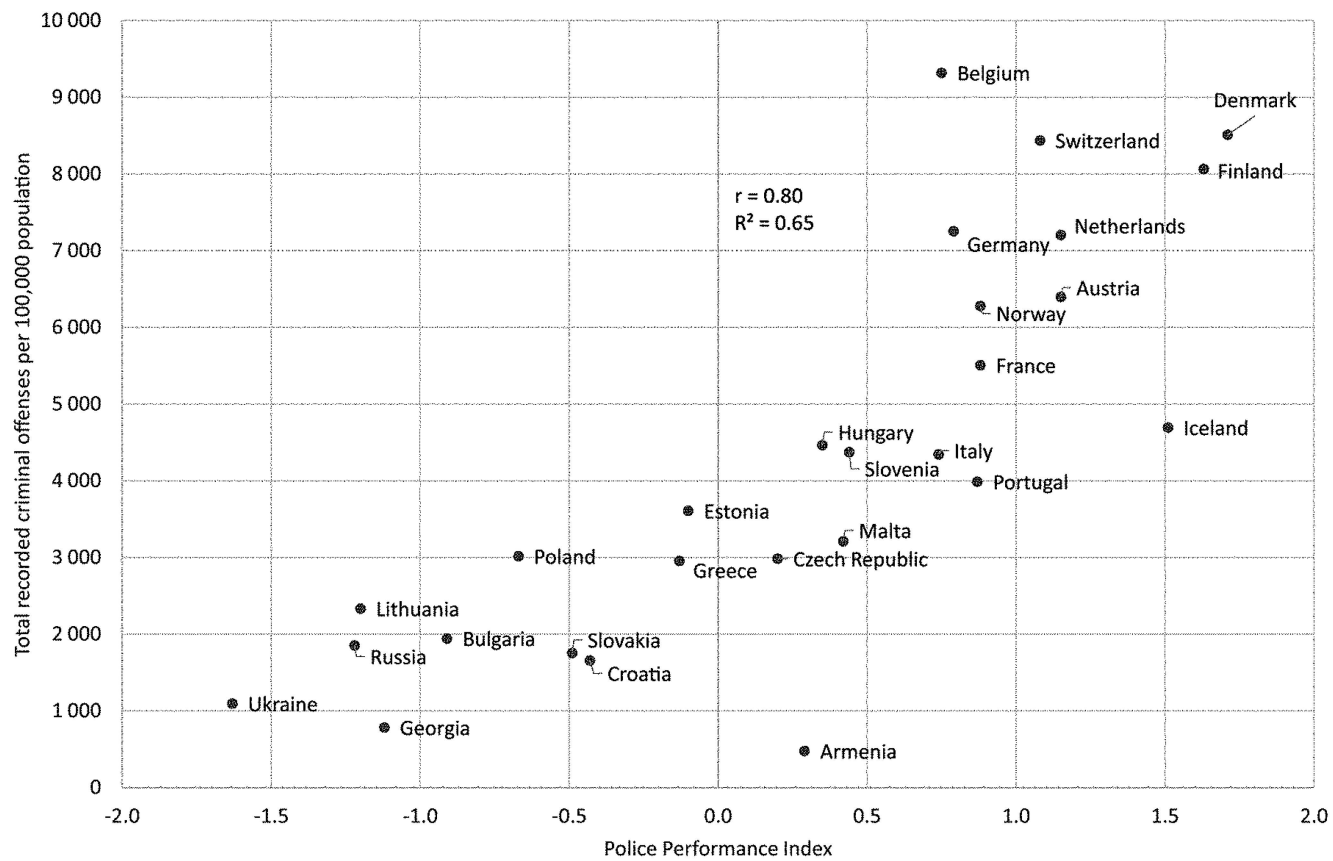


FIG. 1.—Police performance and total crime per 100,000 population, 2010. Source: raw data for the fifth edition European Sourcebook (Aebi et al. 2014); Police Performance Index (PPI; Pare 2014). Albania and Sweden are excluded as outliers.

0.11. For completed homicide, there is a negative correlation ( $r = -0.64$ ); see figure 2. A similar result was found in a worldwide analysis on the connection between police performance and completed homicide rates in 77 countries ( $r = -0.72$ ; Pare 2014, p. 264).

Pare identified several possible explanations for the negative connection between the PPI and the homicide rate, concluding that all have some relevance, but that the strength of the effects of each is unclear (Pare 2014; see Harrendorf 2017a). These are the explanations offered:

- higher clearance and conviction ratios that may operate as deterrents,
- different controls of problem behaviors and crimes that may escalate into homicide (e.g., excessive drinking, burglary),
- strict enforcement of firearm and weapon laws,
- successful interventions against violent hot spots and criminal gangs,
- different measures to pacify conflicts, separate conflict parties, and to protect victims,
- use of force, bound by the principle of proportionality; deadly force as *ultima ratio*, and
- providing alternatives to revenge, vigilantism, and vendettas.

In addition, different levels of legitimacy and procedural justice may directly affect citizens' willingness to abide by the law (Tyler 2006).

These correlations between police performance and crime rates have implications for data comparability. For the total of crime and for minor offenses, rates can be compared but are mainly an indirect measure of police performance, more or less unrelated to the reality of crime. That rates for most severe crimes are not correlated with the PPI does not mean that these rates are totally unrelated to police performance. It can, however, be expected that other variables, especially the incidence of a given crime, have greater influence on crime rates for severe offenses. This hypothesis is supported by the strong negative correlation between police performance and completed homicide rates. The above-mentioned plausible explanations for such a correlation all imply an increase of the true amount of homicide in a society, and not only of cases recorded by the police. There is no plausible mechanism by which weak police performance would only increase police recording of completed homicides.

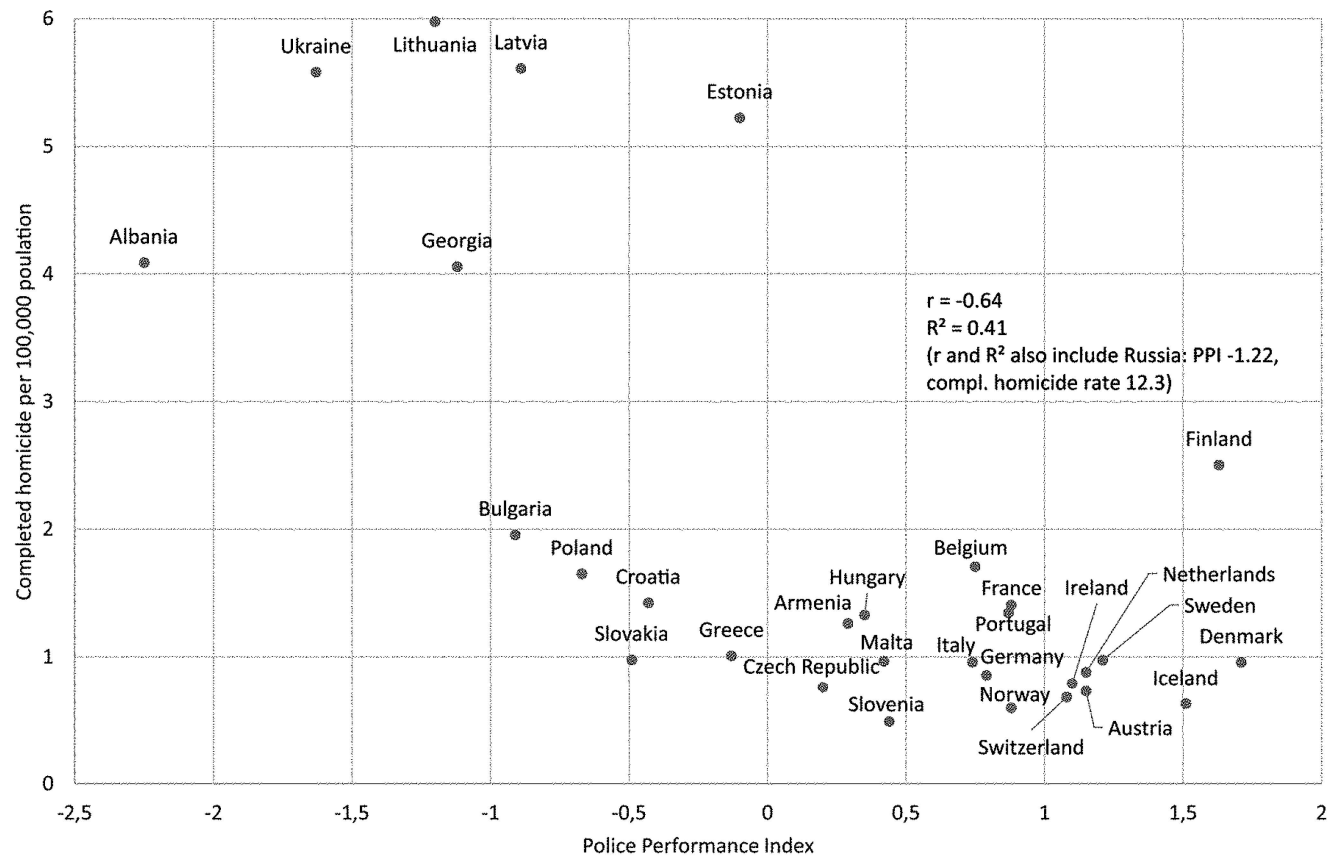


FIG. 2.—Police performance and completed homicide rates per 100,000 population, 2010. Source: raw data for the fifth edition European Sourcebook (Aebi et al. 2014), supplemented by World Health Organization causes of death statistics; PPI (Pare 2014).

Crime levels for serious offenses (except homicide) therefore need to be compared with extreme caution, as they are not a valid measure either of police performance or of the incidence of offenses. This supports the widespread view that completed homicide is the only offense category for which police data may come close to the true picture of crime (Malby 2010*b*; UNODC 2014). Yet there may still be some doubts.

Clearance rates for homicide are usually high, but identification of deaths resulting from homicide is not straightforward. Problems exist especially with respect to missing bodies and persons whose circumstances are unknowable, and with clinical differentiation between natural and unnatural deaths. Concerning the latter, there can be severe problems in cases that do not intuitively appear to be violent deaths, such as poisonings, especially if the victim was ill or old (Mätzler and Wirth 2016). Problems also arise from the systems used to certify deaths. In Germany, for example, death certificates are typically issued by nonspecialized, often family, doctors. They often invest little time in the exercise and, in order not to irritate the bereaved and risk losing them as future patients, often do not adhere to the rules (such as undressing the body and checking it fully for any signs of unnatural death; Rückert 2000; Arbeitsgruppe der AOLG 2011). Partly as a result (and also because of insufficient funding), postmortems are seldom ordered: only in 2 percent of all registered deaths (Stang 2015), compared with, for example, 17 percent examined by coroners in England and Wales (Ministry of Justice 2016). Thus in Germany it is estimated that only one in three homicides is recorded as such, leading to a very large undercount (Brinkmann et al. 1997). More postmortems would reduce the number of unidentified homicides. National differences in postmortem rates will substantially influence comparability of homicide rates. Efforts made to find missing persons are also subject to international variation.

Finally, in dysfunctional criminal justice systems, offenders may get away with murder by bribing a police officer or because no one dares or even wishes to report the murder to the police. Low postmortem rates and lack of efforts to find missing persons can also serve as indices for weak performance of the criminal justice system. In a dysfunctional system, more homicides should go unnoticed by the police. That the correlation between the PPI and the rate of completed homicides is negative could mean, however, that more homicides are recorded in dysfunctional systems. This could be explained by the hypothesis of a higher incidence of homicides. Together, these two inconsistent hypotheses lead to the expectation that the number of recorded cases will increase in a dysfunc-



tional system, but also that many cases go unnoticed. High official homicide rates thus may imply a high level of undetected offenses and low homicide rates a low one.

The UN Survey and Eurostat collect few data on suspects or convictions for specific offenses. The European Sourcebook by contrast provides a detailed breakdown of data by offense type for police-recorded offenses, suspects, and convicted persons. In principle, it is even more questionable to compare these latter rates, since data on suspects in countries with input statistics are recorded later in the process than offense data. Data on convicted persons are necessarily recorded later than police data. Conviction data are more strongly influenced by attrition processes than data on suspects, and data on suspects are subject to additional attrition compared with data on police-recorded offenses. Since attrition processes differ significantly between countries, depending on the architecture of the criminal justice system and criminal procedure rules and practices, comparability is further reduced for variables relating to later stages of the process (Wade 2006; Jehle, Smit, and Zila 2008).

For cases that are not filtered out, however, offense classification will increase in precision, as the initial suspicion that a particular offense was committed increases to certainty beyond reasonable doubt. Depending on the research question being asked, conviction data or data on suspects are thus not necessarily inferior to offense data (Aebi and Linde 2012).

The concerns discussed above concerning rate comparisons also apply to conviction rates and suspects. However, another option for rate comparisons becomes available when combining data on offenses, suspects, and convicted persons. Attrition processes can be compared on the basis of the relation of these rates to each other in different countries (Jehle 2012*a*; Harrendorf 2017*a*).

Incarceration rates are compared at least as often as police-recorded completed homicide rates are. The level of incarceration is often taken as a direct measure of punitiveness (Hinds 2005; Cavadino and Dignan 2006). Differences in crime rates for serious offenses are seldom taken into account, even though they might at least partly explain differences in use of imprisonment (for homicide: Lappi-Seppälä 2008, 2011; Harrendorf 2017*b*; for other serious offenses: Aebi, Linde, and Delgrande 2015). Incarceration rates are, however, negatively correlated to the PPI (Harrendorf 2017*b*). This is shown in figure 3.

The negative correlation coefficient of  $-0.69$  is similar in magnitude to the coefficient for the relationship between the PPI and completed homicides. The (positive) correlation between the completed homicides

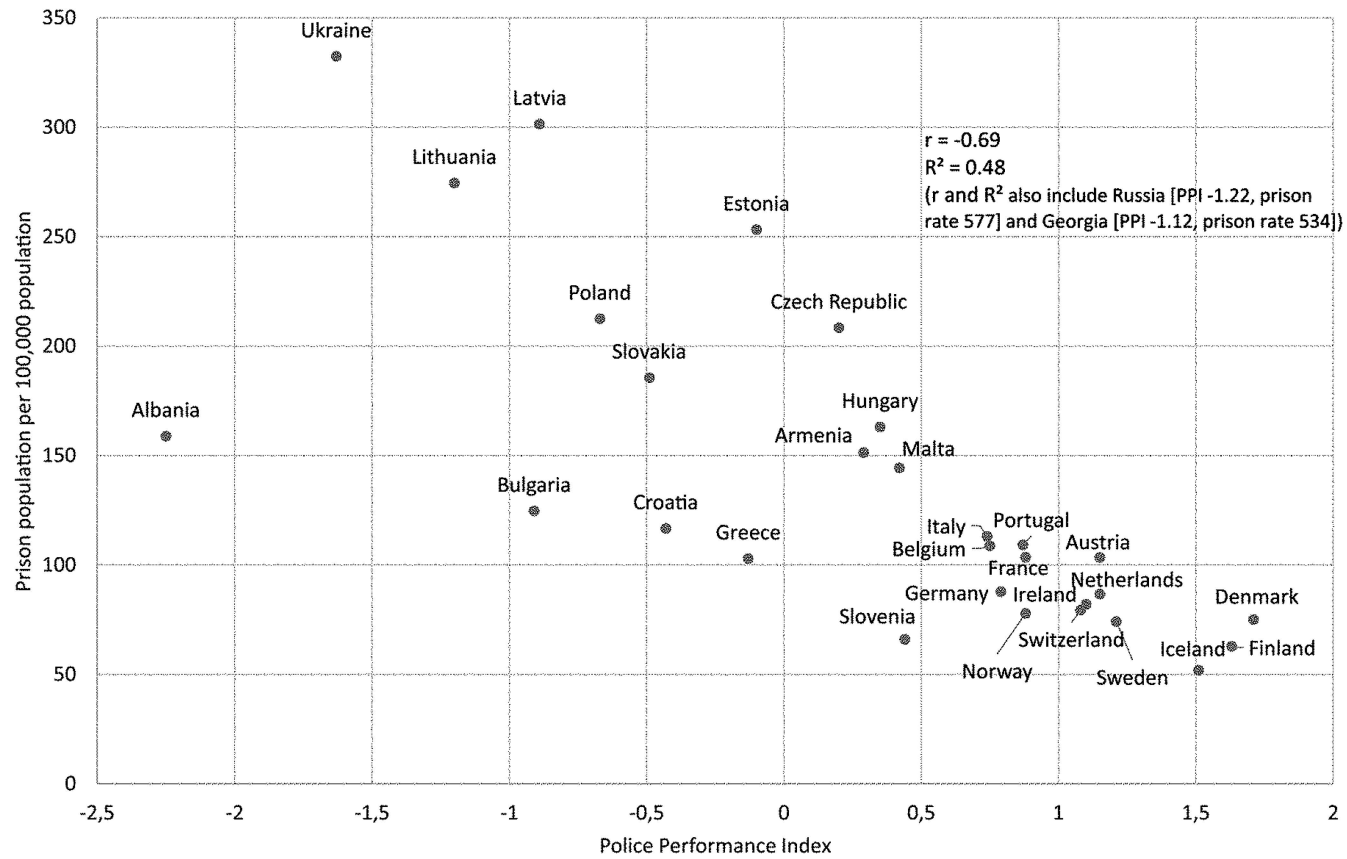


FIG. 3.—Police performance and prisoners per 100,000 population, September 1, 2010. Source: raw data for the fifth edition European Sourcebook (Aebi et al. 2014); PPI (Pare 2014); for some countries, a different reference date within 2010 was used.

rate and incarceration rates is a bit stronger (0.85; Harrendorf 2017*b*). The reason might be that there is a direct effect of a high number of homicides on incarceration: Where many homicides occur, more people are imprisoned for longer periods. The negative correlation of both rates with the PPI suggests an additional plausible explanation: a high imprisonment rate, like low total crime rates and high homicide rates, may indicate a dysfunctional criminal justice system. Countries with dysfunctional systems are probably more prone to punitive responses to crime. Harsh punishments might be necessary to defend the legal order. By contrast, systems that are seen to be just, fair, and trustworthy can be more self-confident and restrict punishment to a minimum (Lappi-Seppälä 2008; Harrendorf 2017*b*; also see Hegel 1821, sec. 218).

### *B. Trend Comparisons*

Rate comparisons are possible but need to be carefully made. Interpretation of differences found is never straightforward, and only police-recorded completed homicide rates can serve as a plausible proxy for real-world crime levels. Warnings about rate comparisons between countries are appropriate (Aebi et al. 2014, p. 21; Eurostat 2017*b*, p. 36). Trend comparisons have the advantage that the values being compared are change rates relative to a reference year. They are not directly influenced by legal or statistical differences between countries (except when there are major changes in criminal procedure or in recording rules during the period under study). Yet for a meaningful comparison of trends it is still necessary to know that there are no fundamental differences in offense definitions and recording practices. It is also necessary to understand that change rates are influenced by different substantive factors. An increase or decrease in a crime rate is not necessarily due to a change in the incidence of the offense in society. It might, for example, be due to changes in the willingness of victims to report. Thus even seemingly identical trends in different countries may have different substantive causes. It is impossible to explain developments in crime rates without taking into account other sources of information such as victimization surveys.

Some elementary data comparability is a prerequisite for trend comparisons, making it important which offenses are compared. The European Sourcebook provides standard offense definitions together with inclusion and exclusion rules. It also asks correspondents to indicate whether these rules have been followed. These answers can be used to

estimate the overall comparability of data for offense categories. Offenses with definitions that are consistent in all respects in many countries should, in principle, have higher comparability than offenses for which many correspondents had to deviate from the definitions and rules. If a country was unable even to report data for a certain offense, this may also indicate a serious problem with the offense definition. For example, the offense may not be separately identifiable (Harrendorf 2012).

Figures 4 and 5 show the overall conformity with offense definitions at the police level for the fifth edition of the European Sourcebook. The percentage of countries completely unable to report data is generally low, but there are many offense definitions that could not be followed for the majority of countries. Sexual assault, robbery, theft, domestic burglary, and money laundering are the only offenses with conformity rates of 50 percent or higher.<sup>9</sup>

Other meaningful comparisons are possible, for example, by evaluating the answers about applicability of the different inclusion and exclusion rules in detail (Harrendorf 2012). Space limitations preclude pursuit of that topic in this essay.

<sup>9</sup> The standard definition of sexual assault is “sexual contact with a person against her/his will or with a person who cannot validly consent to sexual acts. Include the following: any sexual acts committed with violence or threat of violence, any sexual acts committed with abuse of authority or undue pressure, any sexual acts committed against a helpless person, any sexual acts committed against a marital partner against her/his will, acts considered as rape, acts considered as physical sexual abuse of a child, attempts. Exclude the following: any verbal or any other form of non-physical molestation, pornography” (Aebi et al. 2014, pp. 383–84). Compared to the fourth edition European Sourcebook (Aebi et al. 2010), the offense definition for sexual assault was changed, as conformity rates and data availability for the earlier definition were poor (cf. Harrendorf 2012). Obviously, the change of the definition increased comparability. The standard definition of robbery is “stealing from a person with force or threat of force. Include the following: muggings (bag-snatchings), theft immediately followed by force or threat of force used to keep hold of the stolen goods, attempts. Exclude the following: pick-pocketing, extortion, blackmailing” (Aebi et al. 2014, pp. 387–88). The standard definition of theft is “depriving a person or organization of property with the intent to keep it. Include the following: minor (e.g. small value) theft, burglary, theft of motor vehicles, attempts. Exclude the following: embezzlement (including theft by employees), robbery, receiving/handling stolen goods” (p. 388). The standard definition of domestic burglary is “gaining access to closed private premises (e.g. by use of force against an object) with the objective to steal goods. Include the following: theft from an attic or basement in a multi-dwelling building, theft from a secondary residence (even if unoccupied), attempts. Exclude the following: theft from a factory, shop, office, etc., theft from a detached garage, shed, barn or stable, theft from a fenced meadow/compound” (p. 392). The standard definition of money laundering is “specific financial transactions to conceal the identity, source, and/or destination of money or non-monetary property deriving from criminal activities. Include the following: receiving and handling illegally obtained (but not stolen) non-monetary property, attempts. Exclude the following: receiving/handling stolen property, violations of the ‘know-your-customer’ rule (i.e. negligence in identification of customer’s identity or origin of funds)” (p. 395).

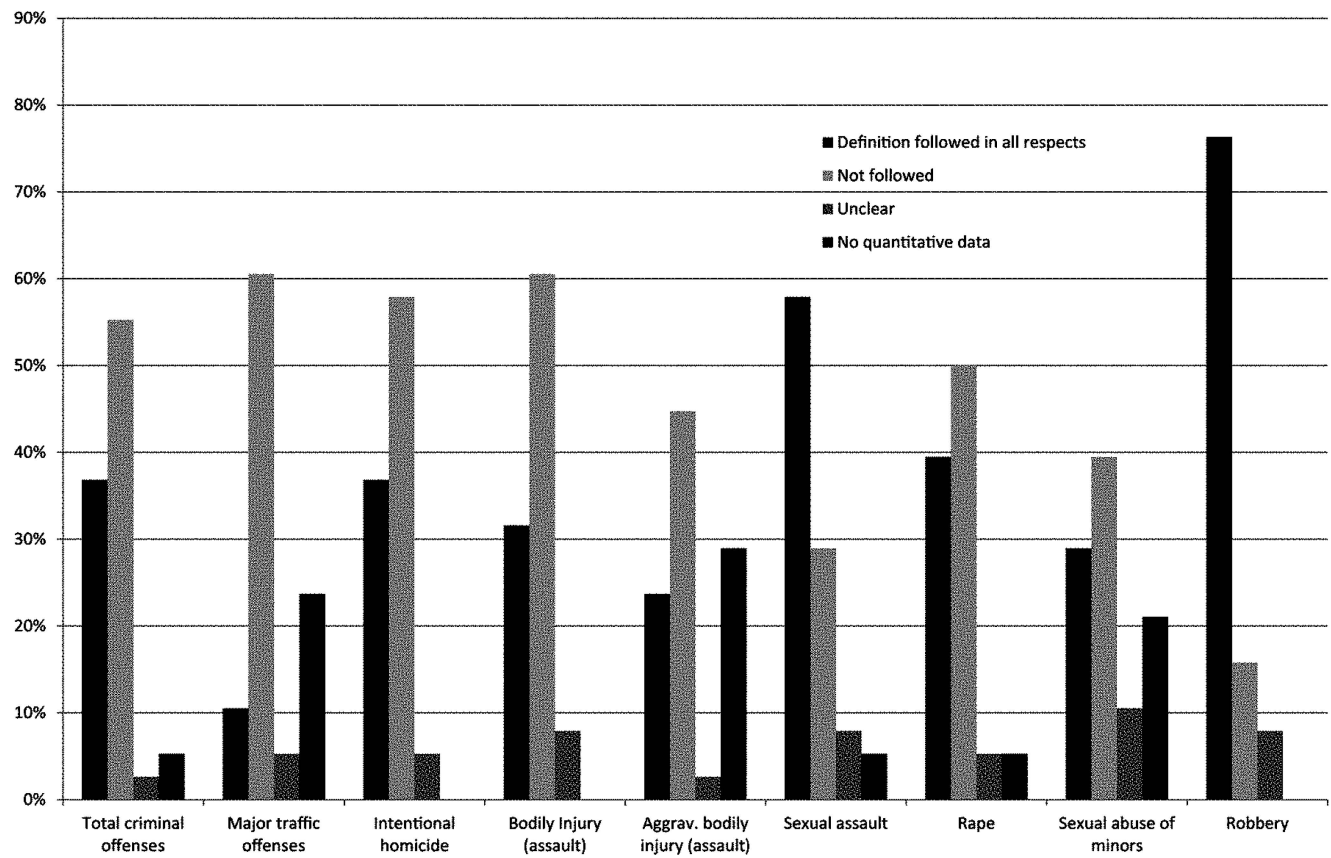


FIG. 4.—Overall conformity with offense definitions, police level, part I. Source: raw data for the fifth edition European Sourcebook (Aebi et al. 2014).

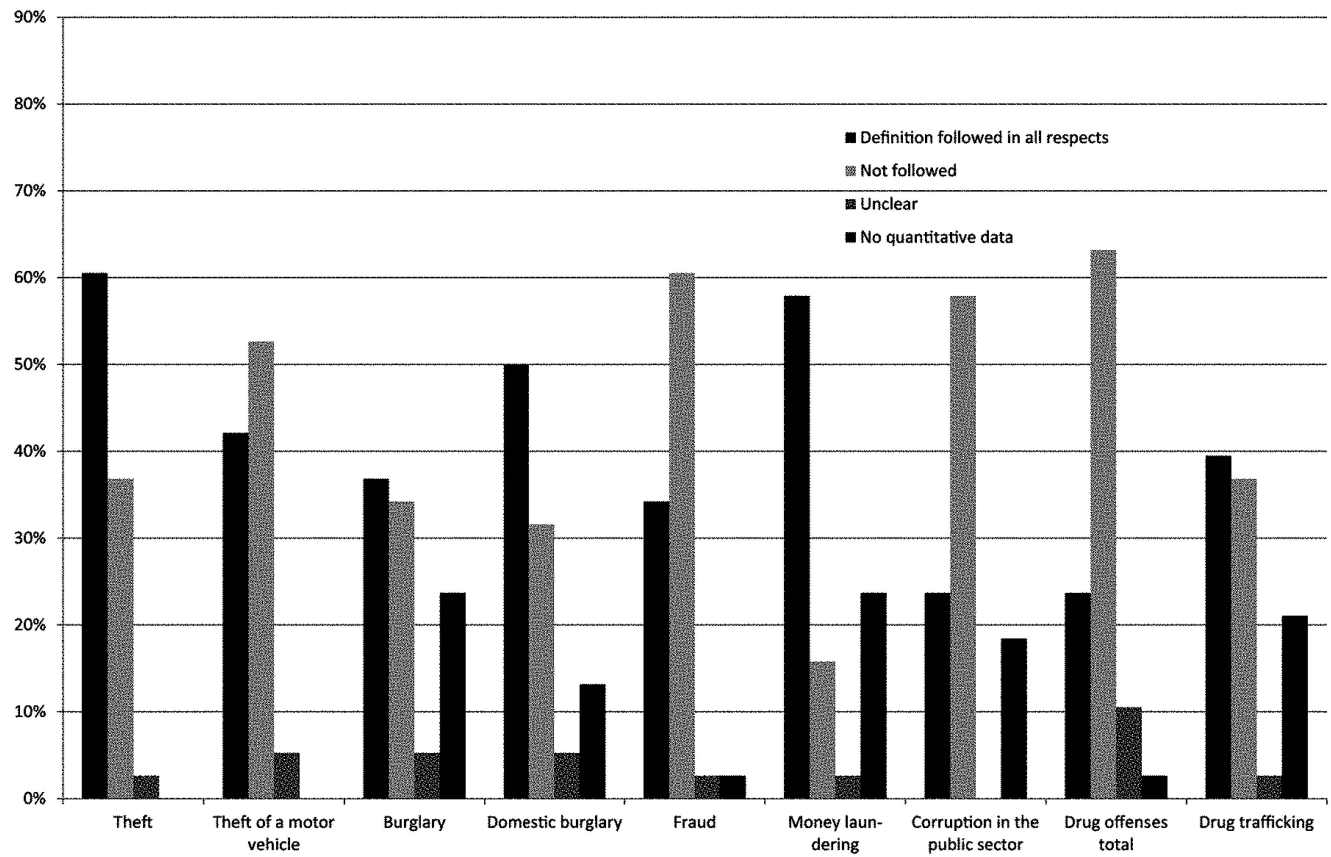


FIG. 5.—Overall conformity with offense definitions, police level, part II. Source: raw data for the fifth edition European Sourcebook (Aebi et al. 2014).

However, some limitations of assessments of comparability based on overall conformity rates need to be mentioned. First, overall conformity rates are sensitive to the number of items on the inclusion and exclusion lists. The probability that a country cannot follow a definition in all respects increases with the number of items. The high conformity rate for robbery is likely also due to its short list of inclusions and exclusions compared with, for example, fraud.<sup>10</sup>

Second, some rules are more important for comparability than others because they affect larger proportions of offenses or offenders. Data comparability for theft, for example, is affected more by exclusion of minor thefts than by inclusion of robbery.

Homicide is an offense with low conformity rates but probably has good international comparability. Some other offenses have high conformity in definitions but still lack comparability, for example, because of strong influence of substantive factors. Money laundering is such an offense, as the exceptionally high variation coefficients show (Harrendorf 2012). High conformity is, however, at least an indicator of increased comparability. Trend comparisons are therefore especially likely to be valid for theft, robbery, sexual assault, and homicide and at the police level also for domestic burglary.<sup>11</sup>

The total of criminal offenses for most purposes is not reliable for international comparison. First, it is a black box with unknown content with respect to the offenses covered, mainly because borderlines between criminal and other forms of deviant behavior are drawn differently in each country (Harrendorf 2011). The total rate is mainly a measure of the quality of police work (this is also for the theft rate). Yet the trend for total criminal offenses is important for international comparison, as it shows the overall workload of the system. This cannot be shown by any of the specific offenses. That the total rate mainly measures the quality of police work also does not rule out comparison, but offers cautions about the meaning of varying trends and rates.

<sup>10</sup> Standard definition (Aebi et al. 2014, p. 393): “deceiving someone or taking advantage of someone’s error with the intent to unlawfully gain financial benefits, thereby causing the deceived person to enter any operation that will be damaging to his/her or a third person’s financial interests. Include the following: attempts. Exclude the following: receiving/handling stolen property, forgery of documents, passports etc., tax and customs offences, subsidy fraud, fraud involving welfare payments, money laundering, forgery of money/payment instruments, consuming goods or services, breaching of trust/embezzlement.”

<sup>11</sup> On a convictions level, many countries are unable to report data, since domestic burglary is not a criminal law concept in their country (cf. Aebi et al. 2014, p. 373).

### *C. Building and Comparing System-Based Indicators*

Trends can thus be compared more easily than rates because the influence of legal and statistical differences is minimized. Another way to improve comparability is to build indicators as ratios of two different, system-based variables. By doing so, influences of different legal and substantive factors such as offense definitions or reporting rates are controlled for, assuming that the influence of these factors is the same for the numerator and denominator variables. The same is true for statistical factors such as different counting rules. Counting rules can, however, be controlled effectively only if both the numerator and denominator variables stem from the same statistics or from different statistics that apply the same methodology. In practice, different national statistical systems often have different counting rules.

System-based indicators are therefore a good approximation for major system-based differences, for example, attrition processes or punitiveness (Jehle 2012*a*; Smit, van Eijk, and Decae 2012; Harrendorf 2013*b*; Harrendorf, Jehle, and Smit 2014). Attrition, for example, can be measured by comparing ratios such as the offender ratio (suspects per 100 recorded offenses) and the conviction ratio (convicted persons per 100 suspects). For punitiveness, several possible system-based measures are superior to a simple comparison of incarceration rates. One such possibility is a ratio using the total number of prisoners or of convicted prisoners as the numerator and the total number of convictions as the denominator (cf. Harrendorf 2017*b*, pp. 145–51; also see Lappi-Seppälä [2008, pp. 327–28], relating the numbers of prisoners to the number of police-recorded crimes). Since the prison stock mainly depends on the number and length of unsuspended prison sentences issued by the courts, this is a plausible indicator of relative harshness (Smit 2009; Smit, van Eijk, and Decae 2012; Harrendorf 2013*a*). Of course such an indicator is still highly correlated with the imprisonment rate itself (Lappi-Seppälä 2008, pp. 327–28). Results discussed above in Section III.A also showed that incarceration rates can at least be seen as a proxy for punitiveness and—together with low total crime rates and high homicide rates—are an indicator for a dysfunctional criminal justice system.

### *D. Country Clustering*

The immense differences in the rates for crime and criminal justice variables also complicate the development and comparative interpreta-



tion of country clusters. Variation coefficients are a measure of relative variation, as they are a ratio calculated by dividing the standard deviation by the mean. Variation coefficients are often higher than 100 percent (Harrendorf 2012, 2013*b*). Thus, the mean does not represent the individual country results. Even if data for all countries in the world were available, neither the population-weighted nor the unweighted mean would provide a meaningful world representation. This is also true for Eurostat or European Sourcebook data with respect to Europe. Only arbitrary results would be produced if means were generated for a convenience sample of countries located on the same continent. For each sample of countries chosen, the mean would be very different from that for any other sample. I criticized that approach in the introduction.

Country clustering is useful only for groups of countries in which the legal architecture and practices are comparable. Smit, Marshall, and van Gammeren (2008) illustrate an innovative empirical approach to country clustering. By use of Categorical Principal Components Analysis, they identified groups of countries that resemble each other in their rates for different variables, but also taking into account some geographical, political, and cultural characteristics. They concluded that it is feasible and sensible to cluster European countries into four large groups: East, Central, North/West, and South. They summarized rules to determine the exact positioning of countries:

First, all countries that used to be Soviet states are placed in the category “East.”

Secondly, all “countries in transition,” that is, all countries that used to be in the sphere of influence of the Soviet Union before the 1990s, are placed in the category “Central.” The former Yugoslavian countries are not in this category.<sup>12</sup>

The remainder of the countries are divided into two categories “North/West” and “South” on geographical grounds only. With “South” meaning south of the Pyrenees and the Alps. The USA and Canada are placed in the category “North/West.” (2008, p. 186)

Of course, this approach to country clustering is only an example. There are other sensible possibilities. Most important is that country clusters

<sup>12</sup> The countries from former Yugoslavia are placed in the “South” category. On substantive reasons not to put these countries into the “Central” category, with respect to the criminal justice policy of former Yugoslavia, see Flander and Meško (2016).

need to be construed on the basis of a clear theoretical concept. Such a concept can be derived from differences and similarities of criminal justice systems or rely on a typology of political economies (cf. Cavadino and Dignan 2006).

#### IV. Planning and Carrying Out a Meaningful Comparative Study

I have tried to provide an overview of the prospects, problems, and pitfalls of comparative studies. Achieving data comparability is difficult, almost impossible. Data are strongly influenced by different substantive, legal, and statistical factors, which make it impossible simply to use unmodified data from different national statistic systems. Usually, it is necessary to rely on international surveys for more or less comparable data. These surveys try to increase comparability by adapting data to specific standards and thoroughly documenting remaining differences. Data from the UN Survey, the Eurostat data collection, and the European Sourcebook have distinct advantages and disadvantages.

I tried to explain what can, and cannot or should not, be done with international crime and criminal justice system data. In comparing national data from international surveys, it is important to follow certain rules. The main ones are summarized below.

1. Do not use comparative crime and criminal justice data if the research question can be answered by relying on data from international victim or offender surveys. International comparability for these surveys is better.
2. Do not use comparative crime and criminal justice data to investigate the true incidence of crime in different countries; international victim or offender surveys are better for this. However, as an exception, it may be feasible to compare levels and trends for completed homicide in different countries and use them as indicators.
3. Comparisons of rates for crime and criminal justice variables between countries should be made extremely cautiously. This does not rule them out, but the interpretation of differences found may be complex.
4. Crime rates are based on the work of actors involved in the criminal justice process and are necessarily influenced by the quality

and efficiency of their work. Some rates, like the total of criminal offenses, can be seen as a proxy for qualitative police performance.

5. Trend comparisons are more reliable than rate comparisons, as the influence of legal and statistical factors is reduced. They should be preferred, if possible.
6. Comparability can be improved by controlling for the influence of distorting factors by using indicators calculated as ratios of two different variables.
7. Punitiveness in international comparisons should not be measured only by incarceration rates. Other indicators, such as the ratio between the size of the prison population on a given date and the number of convictions that year, are also instructive.
8. Country clustering is a suspect task, since data variations between countries are huge. Mean crime rates for the world or even for Europe cannot be credibly calculated. Country clusters are potentially feasible only for countries for which data are highly similar.
9. The best way to obtain comparable data for different countries is to conduct a multicountry study using an identical methodology, for example, by relying on case files of the courts or prosecution services.
10. If you have to rely on secondary analysis of statistical data, never use national data unmodified in comparative projects. Use data from international surveys.
11. Choose the survey that best provides the variables you need and that fits the regional scope of your study.
12. For European studies, the European Sourcebook is preferable to using Eurostat or UN Survey data because it much more fully documents differences in offense definitions and recording practices and has a better validation process.
13. If data look strange, do not trust them! Look critically at data before using them and check for internal consistency, inexplicable increases or decreases in trends, and differing values for the same or comparable variables from other surveys.
14. Try to correct wrong or suspect data by replacing them with data for an (almost) identical variable from another international survey.
15. Remember in comparing offense-related cross-national data that data for some offenses are much more reliable and comparable than for others.

16. When drawing from the European Sourcebook, data for theft, robbery, sexual assault, and homicide, and at the police level also for domestic burglary, are relatively comparable.

## REFERENCES

- Aebi, Marcelo F. 2008. "Measuring the Influence of Statistical Counting Rules on Cross-National Differences in Recorded Crime." In *Crime and Criminal Justice Systems in Europe and North America, 1995–2004*, edited by Kauko Aromaa and Markku Heiskanen. Helsinki: Helsinki United Nations Institute.
- . 2010. "Methodological Issues in the Comparison of Police-Recorded Crime Rates." In *International Handbook of Criminology*, edited by Shlomo Giora Shoham, Paul Knepper, and Martin Kett. Boca Raton, FL: CRC Press.
- Aebi, Marcelo F., Galma Akdeniz, Gordon Barclay, Claudia Campistol, Stefano Caneppele, Beata Gruszczyńska, et al. 2012. *European Sourcebook of Crime and Criminal Justice Statistics*. Questionnaire Covering the Years 2007–2011. Lausanne: University of Lausanne.
- . 2014. *European Sourcebook of Crime and Criminal Justice Statistics—2014*. 5th ed. Helsinki: Helsinki United Nations Institute.
- Aebi, Marcelo F., Bruno Aubusson de Cavarlay, Gordon Barclay, Beata Gruszczyńska, Stefan Harrendorf, Markku Heiskanen, et al. 2010. *European Sourcebook of Crime and Criminal Justice Statistics—2010*. 4th ed. Den Haag: Boom.
- Aebi, Marcelo F., and Julien Chopin. 2016. *Council of Europe Annual Penal Statistics. SPACE II Survey 2015: Persons Serving Non-custodial Sanctions and Measures in 2015*. Strasbourg: Council of Europe.
- Aebi, Marcelo F., and Antonia Linde. 2012. "Conviction Statistics as an Indicator of Crime Trends in Europe from 1990 to 2006." *European Journal on Criminal Policy and Research* 18(1):103–44.
- Aebi, Marcelo F., Antonia Linde, and Natalia Delgrande. 2015. "Is There a Relationship between Imprisonment and Crime in Western Europe?" *European Journal on Criminal Policy and Research* 21:425–46.
- Aebi, Marcelo F., Mélanie M. Tiago, and Christine Burkhardt. 2017. *Council of Europe Annual Penal Statistics. SPACE I—Prison Populations: Survey 2015*. Updated April 25. Strasbourg: Council of Europe.
- Alvazzi del Frate, Anna. 2010. "Crime and Criminal Justice Statistics Challenges." In *International Statistics on Crime and Justice*, edited by Stefan Harrendorf, Markku Heiskanen, and Steven Malby. Helsinki: Helsinki United Nations Institute.
- Arbeitsgruppe der AOLG. 2011. *Verbesserung der Qualität der äußeren Leichenschau*. Berlin: Arbeitsgruppe der AOLG.
- Barberet, Rosemary. 2009. "The Legacy of INTERPOL Crime Data to Cross-National Criminology." *International Journal of Comparative and Applied Criminal Justice* 33(2):193–210.

- Blumstein, Alfred, Michael Tonry, and Ashley Van Ness. 2005. "Cross-National Measures of Punitiveness." In *Crime and Punishment in Western Countries, 1980–1999*, edited by Michael Tonry and David P. Farrington. Vol. 33 of *Crime and Justice: A Review of Research*, edited by Michael Tonry. Chicago: University of Chicago Press.
- Brå (Brottsförebyggande rådet). 2015. *The Clearance Rate in Sweden and in Other Countries*. Stockholm: Brå.
- Brinkmann, Bernd, S. Banaschak, H. Bratzke, U. Cremer, G. Drese, C. Erfurt, et al. 1997. "Fehlleistungen bei der Leichenschau in der Bundesrepublik Deutschland: Ergebnisse einer multizentrischen Studie." *Archiv für Kriminologie* 199:1–12, 65–74.
- Buonanno, Paolo, Francesco Drago, and Roberto Galbiati. 2014. "How Much Should We Trust Crime Statistics? A Comparison between UE and US." LIEPP Working Paper no. 19. Paris: Sciences Po.
- Cavadino, Michael, and James Dignan. 2006. "Penal Policy and Political Economy." *Criminology and Criminal Justice* 6(4):435–56.
- Churchill, Sefa Awaworyi, and Emmanuel Laryea. 2017. "Crime and Ethnic Diversity: Cross-Country Evidence." *Crime and Delinquency*. Online First, September 23.
- Cole, George F., and Christopher E. Smith. 2011. *Criminal Justice in America*. 6th ed. Belmont, MA: Wadsworth.
- Council of Europe. 1995. *European Sourcebook of Crime and Criminal Justice Statistics*. Draft model. Strasbourg: Council of Europe.
- . 1999. *European Sourcebook of Crime and Criminal Justice Statistics*. Strasbourg: Council of Europe.
- de Bondt, Wendy. 2014. "Evidence Based EU Criminal Policy Making: In Search of Matching Data." *European Journal on Criminal Policy and Research* 20(1):23–49.
- Ebbe, Obi N. I. 2013. "The Purpose of Comparative and International Criminal Justice Systems." In *Comparative and International Criminal Justice Systems: Policing, Judiciary, and Corrections*. 3rd ed. Boca Raton, FL: CRC Press.
- Elsner, Beatrix, Paul Smit, and Josef Zila. 2008. "Police Case-Ending Possibilities within Criminal Investigations." *European Journal on Criminal Policy and Research* 14(2–3):191–201.
- European Commission for the Efficiency of Justice. 2016. *European Judicial Systems: Efficiency and Quality of Justice*. CEPEJ Studies no. 23. Strasbourg: Council of Europe.
- European Monitoring Centre for Drugs and Drug Addiction. 2017. *Statistical Bulletin 2017*. Lisbon: EMCDDA. [http://www.emcdda.europa.eu/data/stats2017\\_en](http://www.emcdda.europa.eu/data/stats2017_en).
- Eurostat. 2017a. *Crime and Criminal Justice Statistics*. Summary Quality Report on the 2016 Data Collection. Luxembourg: Eurostat.
- . 2017b. *Crime and Criminal Justice Statistics: Methodological Guide for Users*. 2017 Version. Luxembourg: Eurostat.
- Flander, Benjamin, and Gorazd Meško. 2016. "Penal and Prison Policy on the Sunny Side of the Alps: The Swan Song of Slovenian Exceptionalism?" *European Journal on Criminal Policy and Research* 22:565–91.

- Harrendorf, Stefan. 2010. *A Validation of Selected 10th and 11th CTS Variables as a Pretest for a General Automated Validation of the Crime Trends Survey Results*. Final report to the UN Office on Drugs and Crime. Göttingen: University of Göttingen.
- . 2011. "How to Measure Punitiveness in Global Perspective: What Can Be Learned from International Survey Data?" In *Punitivity: International Developments*, vol. 1, *Punitiveness: A Global Phenomenon?* edited by Helmut Kury and Evelyn Shea. Bochum, Germany: Brockmeyer.
- . 2012. "Offence Definitions in the European Sourcebook of Crime and Criminal Justice Statistics and Their Influence on Data Quality and Comparability." *European Journal on Criminal Policy and Research* 18(1):23–53.
- . 2013a. "Methodische Überlegungen zu Möglichkeiten und Grenzen vergleichender Punitivitätsmessung auf der Grundlage internationaler Kriminalitätssurveys." In *Täter-Taten-Opfer: Grundlagenfragen und aktuelle Probleme der Kriminalität*, edited by Dieter Dölling and Jörg-Martin Jehle. München-gladbach, Germany: Forum Verlag Godesberg.
- . 2013b. "Towards Comparable International Crime and Criminal Justice Statistics: Where Do We Stand? What Can We Expect?" In *Criminology, Criminal Policy and Criminal Law in an International Perspective: Essays in Honour of Martin Killias on the Occasion of His 65th Birthday*, edited by André Kuhn, Christian Schwarzenegger, Pierre Margot, Andreas Donatsch, Marcelo F. Aebi, and Daniel Jositsch. Zürich: Stämpfli.
- . 2017a. "Attrition in and Performance of Criminal Justice Systems in Europe: A Comparative Approach." *European Journal on Criminal Policy and Research*. Online First, September 7.
- . 2017b. "Justizieller Umgang mit kriminellem Verhalten im internationalen Vergleich: Was kann 'Comparative Criminal Justice' leisten?" *Rechtswissenschaft* 8(2):113–52.
- Harrendorf, Stefan, Jörg-Martin Jehle, and Paul Smit. 2014. "Attrition." In *Recording Community Sanctions and Measures and Assessing Attrition: A Methodological Study on Comparative Data in Europe*, edited by Markku Heiskanen, Marcelo F. Aebi, Willem van der Brugge, and Jörg-Martin Jehle. Helsinki: Helsinki United Nations Institute.
- Harrendorf, Stefan, and Paul Smit. 2010. "Attributes of Criminal Justice Systems: Resources, Performance, and Punitivity." In *International Statistics on Crime and Justice*, edited by Stefan Harrendorf, Markku Heiskanen, and Steven Malby. Helsinki: Helsinki United Nations Institute.
- Hegel, Georg W. F. 1821. *Grundlinien der Philosophie des Rechts*. Berlin: Nicolai.
- Heiskanen, Markku. 2010. "Trends in Police-Recorded Crime." In *International Statistics on Crime and Justice*, edited by Stefan Harrendorf, Markku Heiskanen, and Steven Malby. Helsinki: Helsinki United Nations Institute.
- Heiskanen, Markku, Marcelo F. Aebi, Willem van der Brugge, and Jörg-Martin Jehl, eds. 2014. *Recording Community Sanctions and Measures and Assessing Attrition: A Methodological Study on Comparative Data in Europe*. Helsinki: Helsinki United Nations Institute.

- Hinds, Lynette R. 2005. "Crime Control in Western Countries, 1970 to 2000." In *The New Punitiveness: Trends, Theories, Perspectives*, edited by John Pratt, David Brown, Mark Brown, Simon Hallsworth, and Wayne Morrison. Cullompton, Devon, UK: Willan.
- Interpol General Assembly. 2006. *Resolution AG-2006-RES-19: Proposal to Discontinue the Production of Crime Statistics*. Rio de Janeiro: Interpol.
- Jehle, Jörg-Martin. 2012a. "Attrition and Conviction Rates of Sexual Offences in Europe." *European Journal on Criminal Policy and Research* 18(1):145–61.
- . 2012b. "How to Improve the International Comparability of Crime Statistics." In *New Types of Crime: Proceedings of the International Seminar Held in Connection with HEUNI's Thirtieth Anniversary 20 October 2011*. Helsinki: Helsinki United Nations Institute.
- . 2015. *Criminal Justice in Germany: Facts and Figures*. 6th ed. München-gladbach, Germany: Forum Verlag Godesberg.
- Jehle, Jörg-Martin, and Stefan Harrendorf, eds. 2010. *Defining and Registering Criminal Offences and Measures: Standards for a European Comparison*. Göttingen, Germany: Universitätsverlag Göttingen.
- . 2014. "How to Record Data on Community Sanctions and Measures and the Work of Probation Agencies across Europe: The Approach of the European Sourcebook." In *Organized Crime, Corruption, and Crime Prevention: Essays in Honor of Ernesto U. Savona*, edited by Stefano Caneppele and Francesco Calderoni. New York: Springer.
- Jehle, Jörg-Martin, Paul Smit, and Josef Zila. 2008. "The Public Prosecutor as Key-Player: Prosecutorial Case-Ending Decisions." *European Journal on Criminal Policy and Research* 14(2–3):161–79.
- Junger-Tas, Josine, Ineke Haen Marshall, Dirk Enzmann, Martin Killias, Majone Steketeer, and Beata Gruszczyńska, eds. 2010. *Juvenile Delinquency in Europe and Beyond: Results of the Second International Self-Report Delinquency Study*. New York: Springer.
- Killias, Martin. 1995. "The European Sourcebook of Crime and Criminal Justice Statistics." *European Journal on Criminal Policy and Research* 3(4):108–17.
- Lappi-Seppälä, Tapio. 2008. "Trust, Welfare, and Political Culture: Explaining Differences in National Penal Policies." In *Crime and Justice: A Review of Research*, vol. 37, edited by Michael Tonry. Chicago: University of Chicago Press.
- . 2011. "Explaining Imprisonment in Europe." *European Journal of Criminology* 8:303–28.
- Lappi-Seppälä, Tapio, and Martti Lehti. 2014. "Cross-Comparative Perspectives on Global Homicide Trends." In *Why Crime Rates Fall and Why They Don't*, edited by Michael Tonry. Vol. 43 of *Crime and Justice: A Review of Research*, edited by Michael Tonry. Chicago: University of Chicago Press.
- Lewis, Chris G. 2012. "Crime and Justice Statistics Collected by International Agencies." *European Journal on Criminal Policy and Research* 18(1):5–21.
- Lovett, Jo, and Liz Kelly. 2009. *Different Systems, Similar Outcomes? Tracking Attrition in Reported Rape Cases across Europe*. London: Child and Woman Abuse Studies Unit, London Metropolitan University.

- Malby, Steven. 2010a. "Drug Crime." In *International Statistics on Crime and Justice*, edited by Stefan Harrendorf, Markku Heiskanen, and Steven Malby. Helsinki: Helsinki United Nations Institute.
- . 2010b. "Homicide." In *International Statistics on Crime and Justice*, edited by Stefan Harrendorf, Markku Heiskanen, and Steven Malby. Helsinki: Helsinki United Nations Institute.
- Mätzler, Armin, and Ingo Wirth. 2016. *Todesermittlung: Grundlagen und Fälle*. 5th ed. Heidelberg: C. F. Müller.
- Ministry of Justice, England and Wales. 2016. *Coroners Statistics Annual 2015*. London: Ministry of Justice.
- Nelken, David. 2010. *Comparative Criminal Justice: Making Sense of Difference*. London: Sage.
- Pakes, Francis. 2015. *Comparative Criminal Justice*. 3rd ed. Abingdon, UK: Routledge.
- Pare, Paul-Philippe. 2014. "Indicators of Police Performance and Their Relationships with Homicide Rates across 77 Nations." *International Criminal Justice Review* 24(3):254–70.
- Porter, Michael E., Klaus Schwab, Xavier Sala-i-Martin, and Augusto Lopez-Claros. 2004. *The Global Competitiveness Report, 2003–2004*. New York: Oxford University Press.
- Rosenfeld, Richard, and Steven F. Messner. 2009. "The Crime Drop in Comparative Perspective: The Impact of the Economy and Imprisonment on American and European Burglary Rates." *British Journal of Sociology* 60(3): 445–71.
- Rubin, Marilyn M., Richard Culp, Peter Mameli, and Michael Walker. 2008. "Using Cross-National Studies to Illuminate the Crime Problem: One Less Data Source Left Standing." *Journal of Contemporary Criminal Justice* 24(1): 50–68.
- Rückert, Sabine. 2000. *Tote haben keine Lobby: Die Dunkelziffer der vertuschten Morde*. Hamburg: Hoffmann & Campe.
- Smit, Paul. 2009. "Nederland in internationaal perspectief." In *Criminaliteit en Rechtsbandhaving 2008*, edited by Sandra N. Kalidien and A. Th. J. Eggen. Den Haag: Boom.
- Smit, Paul, Ineke Haen Marshall, and Mirjam van Gammeren. 2008. "An Empirical Approach to Country Clustering." In *Crime and Criminal Justice Systems in Europe and North America, 1995–2004*, edited by Kauko Aromaa and Markku Heiskanen. Helsinki: Helsinki United Nations Institute.
- Smit, Paul, Anneke van Eijk, and Rob Decae. 2012. "Trends in the Reaction on Crime in Criminal Justice Systems in Europe in 1990–2007: A Comparison of Four European Regions." *European Journal on Criminal Policy and Research* 18 (1):55–82.
- Stang, Michael. 2015. *Rechtsmedizin: Defizite bei Leichenschau und Obduktion*. [http://www.deutschlandfunk.de/rechtsmedizin-defizite-bei-leichenschau-und-obduktion.676.de.html?dram:article\\_id=338753](http://www.deutschlandfunk.de/rechtsmedizin-defizite-bei-leichenschau-und-obduktion.676.de.html?dram:article_id=338753).
- Tonry, Michael. 2015. "Is Cross-National and Comparative Research on the Criminal Justice System Useful?" *European Journal of Criminology* 12(4):505–16.



- Tonry, Michael, and David P. Farrington. 2005. "Punishment and Crime across Space and Time." In *Crime and Punishment in Western Countries, 1980–1999*, edited by Michael Tonry and David P. Farrington. Vol. 33 of *Crime and Justice: A Review of Research*, edited by Michael Tonry. Chicago: University of Chicago Press.
- Transparency International. 2013. *Global Corruption Barometer 2013*. Berlin: Transparency International.
- Tyler, Tom R. 2006. *Why People Obey the Law*. Princeton, NJ: Princeton University Press.
- UN Economic and Social Council. 2015. *Annual Report Questionnaire*. Part 4: *Extent and Patterns of and Trends in Drug Crop Cultivation and Drug Manufacture and Trafficking*. Vienna: UN Economic and Social Council.
- . 2016. *World Crime Trends and Emerging Issues and Responses in the Field of Crime Prevention and Criminal Justice: Note by the Secretariat*. E/CN.15/2016/10. Commission on Crime Prevention and Criminal Justice, 25th session, May 23–27. Vienna: UN Economic and Social Council.
- UNODC (UN Office on Drugs and Crime). 2014. *Global Study on Homicide 2013: Trends, Contexts, Data*. Vienna: UNODC.
- . 2015. *International Classification of Crimes for Statistical Purposes*. Vienna: UNODC.
- . 2016a. *United Nations Survey of Crime Trends and Operations of Criminal Justice Systems (UN-CTS)—2016*. Vienna: UNODC.
- . 2016b. *World Drug Report 2016*. Vienna: UNODC.
- UN Statistics Division. 2003. *Manual for the Development of a System of Criminal Justice Statistics*. ST/ESA/STAT/SER.F/89. New York: UN Statistics Division.
- van Dijk, Jan. 2008. *The World of Crime*. Thousand Oaks, CA: Sage.
- van Dijk, Jan, John van Kesteren, and Paul Smit. 2007. *Criminal Victimization in International Perspective: Key Findings from the 2004–2005 ICVS and EU ICS*. Den Haag: Boom.
- von Hofer, Hanns. 2000. "Crime Statistics as Constructs: The Case of Swedish Rape Statistics." *European Journal on Criminal Policy and Research* 8(1):77–89.
- Wade, Marianne. 2006. "The Power to Decide: Prosecutorial Control, Diversion and Punishment in European Criminal Justice Systems Today." In *Coping with Overloaded Criminal Justice Systems: The Rise of Prosecutorial Power across Europe*, edited by Jörg-Martin Jehle and Marianne Wade. Berlin: Springer.
- Wartna, Bouke S. J., Ian Knowles, Ian Morton, Susan M. Alma, and Nikolaj Tollenaar. 2014. "Comparison of Reoffending Rates across Countries: An International Pilot Study." In *National Reconviction Statistics and Studies in Europe*, edited by Hans-Jörg Albrecht and Jörg-Martin Jehle. Göttingen, Germany: Universitätsverlag Göttingen.
- Wartna, Bouke S. J., and L. T. J. Nijssen. 2006. *National Studies on Recidivism: An Inventory of Large-Scale Recidivism Research in 33 European Countries*. WODC Studies on Recidivism. Fact Sheet 2006-11. Den Haag: Research and Documentation Center, Netherlands Ministry of Justice.