ethmig
survey
data

Surveying Ethnic Minorities and Migrants

A Legal Framework for Collecting Data and Other Methodological Issues

REPORT 1 – WG3

PATRICK SIMON, National Institute for Demographic Studies (INED)
ESTHER LOPATIN, Interdisciplinary Center (IDC) Herzliya
ANTONINO MARIO OLIVERI, University of Palermo
AMPARO GONZÁLEZ-FERRER, Spanish National Research Council (CSIC)
CECILIA FORTUNATO, Sapienza University of Rome

TABLE OF CONTENTS

INTRODUCTION

Compared to the general population, researchers face a number of specific challenges when conducting surveys of immigrant and ethnic minority populations. Among the reasons for this added complexity is the fact that migrants and ethnic minorities (EMMs) are typically small in number, while their living conditions and legal statuses are often precarious. Although Font and Méndez (2013) have addressed many of the methodological challenges involved in surveying EMMs, their book does not review the legal framework surrounding survey research with population categories that are considered "sensitive." Legal regulations designed to protect individual liberties and personal data has significant implications for survey design when it restricts access to such populations and to information regarding ethnic background, religion, or race. This is a particularly significant concern for researchers using quantitative surveys. Working Group 3 (WG3) was asked to reflect on these challenges before analyzing the results of the survey mapping conducted by WG1 and WG2.

This report describes the background for collecting data concerning EMMs in Europe by:

1. Emphasizing the relevant EU legal frameworks pertaining to data collection in the EU and some of their implications for conducting surveys of immigrants and ethnic minorities.
2. Reviewing and evaluating the empirical knowledge underlying these frameworks and specific data collection, analysis, and reporting practices.
3. Providing an overview of the methodological challenges faced by researchers who survey EMMs.
4. Alerting EMM survey data producers and data users to the range of integration indicators that could inform their future data production and data analysis efforts.

CHAPTER 1: The Legal Framework for Data Collection in Europe

Patrick Simon

Collecting data on migrants and ethnic minorities raises methodological issues that are particularly acute given the sensitivity of both the target population and the data collected. Indeed, every country in the world has acknowledged, to varying degrees, that data concerning race, ethnicity, nationality, color, and religion require high levels of protection and have developed specific legal guidelines for how EMMs data are collected. This chapter discusses these legal frameworks and their potential impact on data collection among migrant and ethnic minority communities.

The Council of Europe has recently adopted new regulations regarding personal data ([Regulation (EU) 2016/679](#)). These regulations are in the process of being transposed into EU member-state laws, and it would be premature to express full confidence in the results of this process. For this reason, this chapter covers the legal framework prior to the new regulation. Because the new regulations include few changes in terms of "special categories of data" or so-called "sensitive data," however, it is likely that the observations contained in this chapter will remain valid in the future.

Privacy and Data Protection: Principal Regulations (EU Directive 95/46/CE)

At the European level, the legal framework for data protection was established by the Council of Europe [Convention ETS 108 *for the Protection of Individuals with Regard to Au-*](#) [*tomatic Processing of Personal Data* (circa 1981)](#) and by [EU Directive 95/46/EC *on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of such Data* (circa 1995)](#). Both the Convention and the Directive have been more or less consistently transposed into domestic laws, although with noticeable variations. Because a detailed, country-by-country analysis is beyond the scope of this report, it assumes EU-wide consistency in data protection legislation.

In addition to data protection laws, most countries have adopted stringent legal frameworks governing statistics that prescribe codes of conduct and regulate statisticians. Most laws and regulations concerning the use of data and statistics can be consulted on individual supervisory agency websites (for data protection laws) and national statistical institutes (for laws regarding statistics). A survey of these laws is included in the report on [*Ethnic Statistics* and *Data Protection* in the Council of Europe Countries (Simon, 2007)](#).

The primary objective of these laws is to prevent the distribution of the personal data of named persons and to ensure that individuals who have supplied personal information are treated fairly and their information is protected. Only previously specified parties for whom this information is intended are authorized to see an individual file. In principle, this means that unspecified parties are not subsequently allowed to access the file if they were not identified when the data were originally compiled, although exceptions can be made for reasons

of public interest or for official supervisory or punitive purposes. Because the concept of public interest is relatively broad and open to interpretation, the secondary use of data for research or official purposes is theoretically permissible but rarely practiced.

Restrictions on data processing primarily depend on how official regulations are applied. Data protection authorities may supervise projects in either pre- or post-processing phases in response to a complaint or on its own initiative. In the former case, data are screened before the processing phase begins. Article 20 of Directive 95 prescribes "prior checks" when processing operations are "likely to present specific risks to the rights and freedoms of data subjects." This provision echoes precautions pertaining to the collection of sensitive data, signifying that some data types can be collected only after the supervising authority has conducted prior checks. The decision of whether or not a case merits simple notification or requires formal authorization establishes the limitations imposed by the provision. It should be noted that in transposing the EU Directive, all EU countries have not adopted the most restrictive version of these regulations.

Using files for purposes other than those for which they are compiled also has a relatively strategic bearing on how discrimination is monitored. Access to files maintained by public agencies or private companies is strictly regulated. When notifying the protection authority that a file is being created, the individual in charge of processing must indicate its purpose. This notification is a key criterion for receiving authorization. The purpose of every file is therefore strictly defined and must be explicitly stated at the time of data collection. The secondary use of files for historical, statistical, or scientific purposes is nonetheless permitted. In such cases, the text of the law may specifically identify the institutions that are authorized to use data.

"Special Categories of Data": Ethnicity, Race, and Religion as Sensitive Issues

All data protection laws list what are called "special categories of data" or, more specifically, "sensitive data." Article 6 of ETS 108 and Article 8 of Directive 95 define the categories of data concerned and the conditions under which they may be collected as follows:

CONVENTION ETS 108 ARTICLE 6 - SPECIAL CATEGORIES OF DATA

> *"Personal data revealing racial origin, political opinions or religious or other beliefs, as well as personal data concerning health or sexual life, may not be processed automatically unless domestic law provides appropriate safeguards. The same shall apply to personal data relating to criminal convictions."*

DIRECTIVE 95/46/EC ARTICLE 8 – THE PROCESSING OF SPECIAL CATEGORIES OF DATA

> *"Member States shall prohibit the processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, trade-union membership, and the processing of data concerning health or sex life."*

THE 2016 REGULATION repeats the same list:

> *"Processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation shall be prohibited."*

For our purposes, the country of birth and citizenship are not included on the list. For

this reason, collecting migrant data does not fall under the verification protocols stipulated for "sensitive data." The prohibition statement does cover ethnic or racial origins and religion, however. National origin is not always explicitly mentioned, but, insofar as "national origin" is identified with ethnic affiliation in Eastern and Central European countries – it is theoretically subject to the same regulations as ethnic or racial origin. Equating "ethnic" and "national" origins highlights one major difficulty in enforcing the regulations pertaining to sensitive data. Because such categories are not clearly defined in data protection protocols, decisions primarily depend on interpretation by data protection agencies. In some countries, these interpretations are strict and only data that directly refer to "ethnic or racial origin" are prohibited. Other countries interpret such cases more broadly and proxies for ethnicity such as nationality, country of birth, or name may also be categorized as sensitive. In these cases, data protection agencies may take the position that data concerning citizenship and/or place of birth may not be collected or only under limited conditions.

Exemptions

Because the first recital is followed by a list of exemptions, the implications of this apparent prohibition are not straightforward. The combination of a prohibition followed by a relatively long list of exceptions under which certain data may nevertheless be collected reflects the spirit of these laws. Their objective is not to prevent the processing of sensitive data *per se*, but to call attention to their specificity while strengthening safeguards. The often-cited view that data protection laws prohibit the collection of sensitive data is therefore an exaggeration. It is true, however, that safeguards such as banning sampling strategies have the potential to severely limit data collection and impede research concerning ethnic minorities.

LIST OF EXEMPTIONS IN DIRECTIVE 95

Article 8 (1) Member States shall prohibit the processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, trade-union membership, and the processing of data concerning health or sex life.

Paragraph 1 shall not apply where:

Article 8 (2a) the data subject has given his explicit consent to the processing of those data, except where the laws of the Member State provide that the prohibition referred to in paragraph 1 may not be lifted by the data subject's giving his consent.

Article 8 (2b) processing is necessary for the purposes of carrying out the obligations and specific rights of the controller in the field of employment law insofar as it is authorized by national law providing for adequate safeguards.

Article 8 (2c) processing is necessary to protect the vital interests of the data subject or of another person where the data subject is physically or legally incapable of giving his consent.

Article 8 (2d) processing is carried out in the course of its legitimate activities with appropriate guarantees by a foundation, association or any other non-profit-seeking body with a political, philosophical, religious or trade-union aim and on condition that the processing relates solely to the members of the body or to persons who have regular contact with it in connection with its purposes and that the data are not disclosed to a third party without the consent of the data subjects.

Article 8 (2e) the processing relates to data which are manifestly made public by the data subject or is necessary for the establishment, exercise or defense of legal claims.

Article 8 (3) Paragraph 1 shall not apply where processing of the data is required for the purposes of preventive medicine, medical diagnosis, the provision of care or treatment or the management of health-

care services, and where those data are processed by a health professional subject under national law or rules established by national competent bodies subject to the obligation of professional secrecy or by another person also subject to an equivalent obligation of secrecy.

Article 8 (4) Subject to the provision of suitable safeguards, Member States may, for reasons of substantial public interest, lay down exemptions in addition to those laid down in paragraph 2 either by national law or by decision of the supervisory authority.

Article 8 (5) Processing of data relating to offences, criminal convictions or security measures may be carried out only under the control of official authority, or if suitable specific safeguards are provided under national law, subject to derogations which may be granted by the Member State under national provisions providing suitable specific safeguards. However, a complete register of criminal convictions may be kept only under the control of official authority.

Not all of the grounds for exemption listed in **Article 8** are equal, and most apply only relatively narrowly. As regards to data collection in social science surveys, explicit consent (2a), health data (8 (3)), and the public interest (8(4)) generally apply. **Recital 34 of Directive 95** clarifies the scope of the allowable exceptions for reasons of public interest by linking them to public health and social protection. This is based on a specific argument that the public interest prevails in the domains of scientific research and public statistics in which the public interest prevails.

RECITAL (34) OF 95:

"Whereas Member States must also be authorized, when justified by grounds of important public interest, to derogate from the prohibition on processing sensitive categories of data where important reasons of public interest so justify in areas such as public health and social protection - especially in order to ensure the quality and

cost-effectiveness of the procedures used for settling claims for benefits and services in the health insurance system - scientific research and government statistics; whereas it is incumbent on them, however, to provide specific and suitable safeguards so as to protect the fundamental rights and the privacy of individuals."

Consent

The notion of consent is central to regulations and oversight of personal data collection in general and specifically of the collection of data that can potentially harm individuals. Consent is defined in **Article 2 (h) of Directive 95** as follows: "The data subject's consent shall mean any freely given specific and informed indication of his wishes by which the data subject signifies his agreement to personal data relating to him being processed." Many national laws are even more specific and require written consent, which can be difficult to obtain during standardized statistical data collection. The practical problems associated with obtaining written consent when statistics are compiled under normal conditions have been sharply criticized by statisticians and researchers. Revisions of the French data-processing law that transposed Directive 95 sparked considerable debate, for example. Statisticians regarded the requirement of written consent as "inappropriate" because it infringes on the underlying trust needed for an effective research subject-researcher relationship. Written consent was also perceived as implying doubt about the guarantee of data anonymity. Finally, obtaining consent was considered difficult to reconcile with actual survey conditions in the field. In all but the most favorable cases, it seems either nearly impossible to obtain written consent or so difficult that it undermines effective data collection.

There is an inherent contradiction in the consent requirement for collecting "sensitive data." Social scientists complain about simultaneously being forced to reassure respondents and justify collecting their data while also asking them to sign a document that resembles a police form (although less intrusive consent forms are also common). On the other hand, from a minimalist perspective voluntarily completing a questionnaire can be interpreted as implying consent. With few exceptions, implicit consent is assumed with anonymous statistical and scientific surveys. Because they are intended to be exhaustive, however, census forms are nearly always mandatory, although they typically preserve an element of consent, by including optional items.

According to the [newer regulations in Article 4](#), "'consent' of the data subject means any freely given, specific, informed and unambiguous indication of the data subject's wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her." Only actual case studies can reveal whether or not such specifications change the impact of consent on the collection of sensitive data, however.

To summarize, data that mention an individual's country of birth or citizenship/nationality are not subject to specific control. Data that reveal or refer to ethnic and racial origins or religion, however, are formally prohibited, although data protection laws authorize their collection whenever:

1. The law requires that they be collected. For example, anti-discrimination laws may include such provisions, although only the United Kingdom's Race Relations Act and the Minorities Act (Minderhedennota, 1979, amended as Allochtonenbeleid in 1989) in the Netherlands require the collection of these data. Most national laws regarding minority populations to include clauses that provide exceptions to bans on collecting of "ethnic" data.

2. Tax liabilities and faith-based organizations render the collecting data regarding religious affiliations necessary.

3. The notion of the "public interest" can be cited as justification for an exemption from the ban on collecting "sensitive" data. Exceptions are case-by-case and would therefore have more limited effects than a legal mandate. This also presupposes that the supervising agency approved the operation as "in the public interest."

4. Explicit consent has been granted.

In theory, these conditions are sufficiently broad to allow the collection of "ethnic" data when required by policy. For this reason, obstacles to processing sensitive data are less a matter of the law *per se* than of a context of an intention to combat racism and discrimination. In most Council of Europe countries, the problem appears to be more a lack of awareness of the potential contributions of statistics to the fight against discrimination than genuine legal obstacles to collecting sensitive personal data. In the case law that guides the authorities responsible for data protection, the public interest and legal obligation are the two most frequently cited justifications for the collection of "ethnic" data.

CHAPTER 2: Expert Opinion From Practitioners and Survey Practices

Patrick Simon, Cecilia Fortunato, and Amparo Gonzalez

Data regarding migrant and ethnic minority populations are not only subject to legal restrictions, but also, for a variety of historical, political, and sociological reasons, their collection is uneven throughout Europe ([Simon, 2012](#)). This explains why the legal framework described in Chapter 1 does not preclude such data collection practices. There are often gaps between formal legal frameworks and how data protection authorities and those who conduct surveys (statistical institutes, survey offices, researchers) implement them. Institutions and researchers may interpret regulatory frameworks either by easing restrictions or, alternatively, by tightening them and restricting data collection. This chapter explores concrete practices used by researchers planning to conduct surveys of EMMs populations.

The chapter addresses the following questions:

1. What is the state of researchers' knowledge regarding data protection guidelines and its impact on the research design underlying the development, administration, and/or analysis of their data?
2. To what extent do existing legal frameworks prevent them from collecting data?
3. What is the exact nature of present-day data collection practices?

In order to extend our awareness of current practices and our understanding of our colleagues' experiences, we addressed a questionnaire (see Appendix) to members of the ETHMIGSURVEYDATA network. 23 responded to the survey. This chapter synthesizes and attempts to interpret responses received from 23 members representing 18 European countries and the United States.

Researcher Awareness of Legal Frameworks and Limitations to Conducting Surveys of EMMs

When asked if the collection of "sensitive data" is prohibited, nearly half of our respondents (9 out of 23) replied that there was no specific ban on the collection of data referring to ethnicity or race. The majority of these unexpected negative responses appear to reflect specific interpretations of the cases in which special regulations in their countries are equivalent to a ban. These respondents did not appear to believe that legal limits on data collection had significant impacts on their work. Indeed, only a single respondent (from France) stated that data protection provisions infringed upon the design of sampling frames.

Other respondents perceived limitations more broadly, but they were not viewed as critical. For example, a researcher from Finland responded, "I am not allowed to make a register that would reveal ethnic identity, belonging, or background of the sample group." Italy pointed out, rightly, that, "Sensitive data have to be handled according to a specific procedure." In Lithuania, "The Law on the Legal

Table 1: *Legal framework and Sensitive Data*

QUESTIONS	Yes	No	Yes and No	Don't know	Didn't answer	Total
Is there a prohibition to collect "sensitive data"?	11	9	3	-	-	23
Does the law prevent you to design a specific sampling?	1	20	-	2	-	23
Does the law prevent you to collect data about sensitive categories? Does the law prevent you to collect data about sensitive categories?	4	16	-	1	2	23

Protection of Personal Data limits collection of data on sensitive categories. The personal data for social research and public opinion surveys may be processed only after the data subject has given his/her consent (except for data that is necessary to contact him/her). There are no specific rules outlined for age groups, vulnerable groups, or research topics when conducting a social research or public opinion survey. The statistical data on sensitive categories (i.e. when a consent from the data subject cannot be acquired) may be processed only after the State Data Protection Inspectorate carry out a prior checking and approves the use of the data."

The respondent from Croatia even expressed opposition to the possibility of restrictions due to the specific status of scientific research: "The Law on Science and Higher Education in Article 20 states: 'Scientific work shall not be subject to any limitations or formal requirements except those resulting from the respect for scientific and research ethics, the protection of human rights as well as the protection of personal and general safety at work.'"

Only 4 respondents reported their impression that the law prevented them from collecting data on so-called sensitive categories. This impression may reflect national practices that diverge from the application of the law in most EU member-states or the fact that these 4 respondents have conducted surveys based on more highly sensitive data categories than country of birth or citizenship. For sampling purposes, the survey suggested that problems do not necessarily stem from data protection provisions but from the existence of information in accessible sources, as in Ireland: "There is no legal restriction in Ireland on sampling as such, but the absence of a population register or migrant register means that getting a representative sample of migrants and/or ethnic minorities is very challenging. There is no way to identify characteristics of the person or household from the most-commonly used sampling frame for household survey (the Geo-Directory)."

Immigrants and Ethnic Minorities: Which Categories?

Previous research has shown that there is no widely accepted definition of ethnicity. Even less agreement has been noted with regard to defining demographic categories. Instead of a unified definition, multiple variables are considered to reflect the various dimensions of ethnicity, including place of birth, nationality, language, religion, place of birth and/or parents' nationalities, self-identification in ethnic groups or ethnic identity, ancestry, race, or color. It is not always possible to convert these dimensions into statistical categories for official or scholarly purposes.

Immigrants are generally defined as foreign-born, i.e., born in a different place than their current place of residence (Eurostat). In some cases (i.e., in France) national definitions combine place of birth with citizenship in order to distinguish between naturalized citizens born outside of French territory and foreigners born in other countries.

Citizenship (or nationality) data are also universally collected, with the exception of the UK, although data regarding multiple citizenships are rarely recorded in censuses or registers.

The same information for individuals' parents is used to identify the so-called second generation, i.e., children of immigrants born in the country of immigration. The nativity of the parents is recorded in Scandinavian countries and the Netherlands, but not in the censuses of other European countries. In Germany, "migration background" is recorded via large-scale surveys like GSOEP or the Microcensus that complement population registers, although they are not comparable in terms of geographical coverage and size of the samples. In France,

second generations are identified in surveys such as the Labor Force Survey or the Housing Survey. However, these data may not be used at the local level and do not cover all areas of social life. An ad hoc module of the EU Labor Force Survey on 'The labor market situation of migrants and their immediate descendants' was developed in 2008 and in 2014 to provide comparative data on "first and second generation immigrants" in the EU 27.

With the exception of the UK and Ireland in Western Europe and Central and Eastern Europe, direct questions about ethnicity are rarely included in surveys or censuses in Europe. In the UK, a question about ethnic group designation is included on the national census as a self-classification item that contains a list of options that combine color or race ("White," "Black British"), ethno-cultural categories ("Arab"), and national/ethnic background ("Asian British," "Pakistani"). In Central and Eastern Europe, when identity data are collected using open questions or a list of categories offered for selection, questions refer to ethnic affiliation or membership in national minorities.

Table 2: *Categories Commonly Referred to in Your Country*

	Often	Rarely	Never	Don't know	Didn't answer
Immigrants (foreign born)	12	9	-	-	1
Citizenship/nationality	6	5	-	-	1
Ethnic affiliation or ethnicity	10	8	3	-	1
Race	-	2	17	1	1
Religion	10	10	1	-	1
Descendants of immigrants (second generation)	10	5	6	-	1
If *Other*, specify:	Ethnic minority (2 respondents)	Country of birth; skin color	Undocumented status	Refugees from former Yugoslav republics	

Table 3: *Categories Commonly Included in Surveys*

	Often	Rarely	Never	Don't know	Didn't answer
Immigrants (foreign born)	13	8	-	-	2
Citizenship/nationality	17	4	-	-	2
Ethnic affiliation or ethnicity	9	9	3	-	2
Race	1	-	17	1	4
Religion	9	11	1	-	2
Descendants of immigrants (second generation)	10	5	6	-	2
If *Other*, specify:	Ethnic minority (2 respondents)	Country of birth; skin color	Undocumented status	Refugees from former Yugoslav republics	

We asked our survey respondents to identify the categories typically surveyed in their respective countries. The results indicate that although there is general agreement about the collection of citizenship data, 5 respondents reported that they are rarely collected. Surprisingly, 9 of the 21 respondents stated that immigrants are rarely identified in statistics (including Croatia, Serbia, the UK, Hungary and the Slovak Republic). Conversely, 10 respondents stated that ethnicity data are often included (including Belgium, Estonia, Croatia, Germany, Serbia, Lithuania, and the UK). With the exception of Belgium, where ethnicity and second-generation immigrant status are often identified, there is a clear divide between countries that collect ethnicity data and those that collect data regarding second-generation immigrants. This clearly demonstrates that data regarding second-generation immigrants functions as a substitute for ethnicity, a possible strategy to avoid maintaining explicit ethnicity statistics.

Table 3 shows the categories that our respondents reported as typically included in their surveys. They reveal a strong alignment between categories in official statistics and surveys, more or less mirroring the previous table. To date, race is almost never collected in official statistics or surveys. The absence of racial categories in official statistics, (with the exception of the UK and Ireland) is unsurprising given the colorblind approach adopted by most European states. More intriguingly—and a potential indication of a widespread colorblind research orientation—is the lack of survey data based on ethno-racial categories that address the issues of racism and discrimination as they affect EMMs.

Even when categories are legitimate, however, collecting data for survey purposes is often difficult. Sampling minorities is inevitably complex, and the absence of ethnic identification in some population registers or censuses forces survey researchers to develop alternative strategies. According to our respondent, in Ireland, for example, where the census monitors ethnic categories, the question of availability is crucial:

"There is no legal restriction in Ireland on sampling as such, but the absence of a population register or migrant register means that getting a representative sample of migrants and/or ethnic minorities is very challenging. There is no way to identify characteristics of the person or household from the most-commonly used sampling frame for household surveys (the GeoDirectory)."

Experiences with the Data Protection Framework

As noted above, the majority of our respondents do not believe that the data protection law prevents them from collecting and sampling "sensitive" data. A key question remains, however: What kind of accommodations must survey researchers make to conform to legal guidelines governing sensitive data categories? We asked respondents to describe the specific procedures that they have adopted in conducting their surveys. A selection of their responses follows:

In the Netherlands, a general agreement with the local data protection commission authorized a survey to monitor citizenship in Amsterdam that collected data concerning immigrants and second-generation status as well as ethnic background self-identification:

> *"We have a standard agreement with the Commission in charge of applying the law at the City of Amsterdam for a number of surveys, including this one. In addition, the dataset that we use for analysis is cleared of all personal details by our Data Collection Unit."*

Surveys about immigrants and second- and third-generation status in Portugal were authorized via "informed consent signed by all interviewees."

In Italy, sensitive questions may be authorized when responses are optional: "Surveys by ISTAT (the Italian National Institute for Statistics) which specifically tackle the conditions of foreigners in Italy include questions of religious affiliations. The latter are asked by informing respondents they are not obliged to answer."

The Data Protection Commission in France makes decisions on a case-by-case basis, resulting in an unstable legal framework. Survey takers are always uncertain whether or not they are authorized to collect EMM data and if so, what format they should follow. In 2005 and 2007, the Data Protection Commission attempted to standardize the regulation of data collection on sensitive categories, but the effort was unsuccessful because they were unable to mobilize the chain of actors involved in data collection. There is a widespread belief that data collection regarding ethnicity or race is prohibited in the country, although in reality, the law stipulates the opposite. As a consequence, the number of surveys conducted in France among EMM populations remains limited, and most of them either describe immigrants or second-generation immigrant status but rarely include religion or ethnicity.

CHAPTER 3: Methodological Challenges

Antonino Mario Oliveri

3.1 A Unifying General Framework

Like other research methods used to investigate social issues, surveys are subject to a wide range of error sources. In the field of ethnic and migrant minorities (EMMs) research, these sources of error have specific characteristics with significant implications for data collection and interpretation. Consequently, the methodological challenges relating to EMM research can be addressed by adopting a unified approach to managing and evaluating survey-related errors.

A starting point for this discussion is the well-known Total Survey Error (TSE) paradigm by Groves, et al. (2004) and Groves and Lyberg (2010).

"Total survey error (TSE) refers to the accumulation of all errors that may arise in the design, collection, processing, and analysis of survey data. In this context, a survey error is defined as the deviation of a survey response from its underlying true value" (Biemer, 2010, p. 817).

The central feature of the TSE approach is that the quality of survey statistics depends on factors that include but are not limited to, sample design. Figure 3.1 shows that survey errors are both of sampling and non-sampling origin and that they depend on the measurement process and representation of the population's characteristics.

As Groves and Lyberg observe (2010: 856–

Figure 3.1: *Total Survey Errors. Source: Groves et al., 2004; Groves and Lyberg, 2010)*
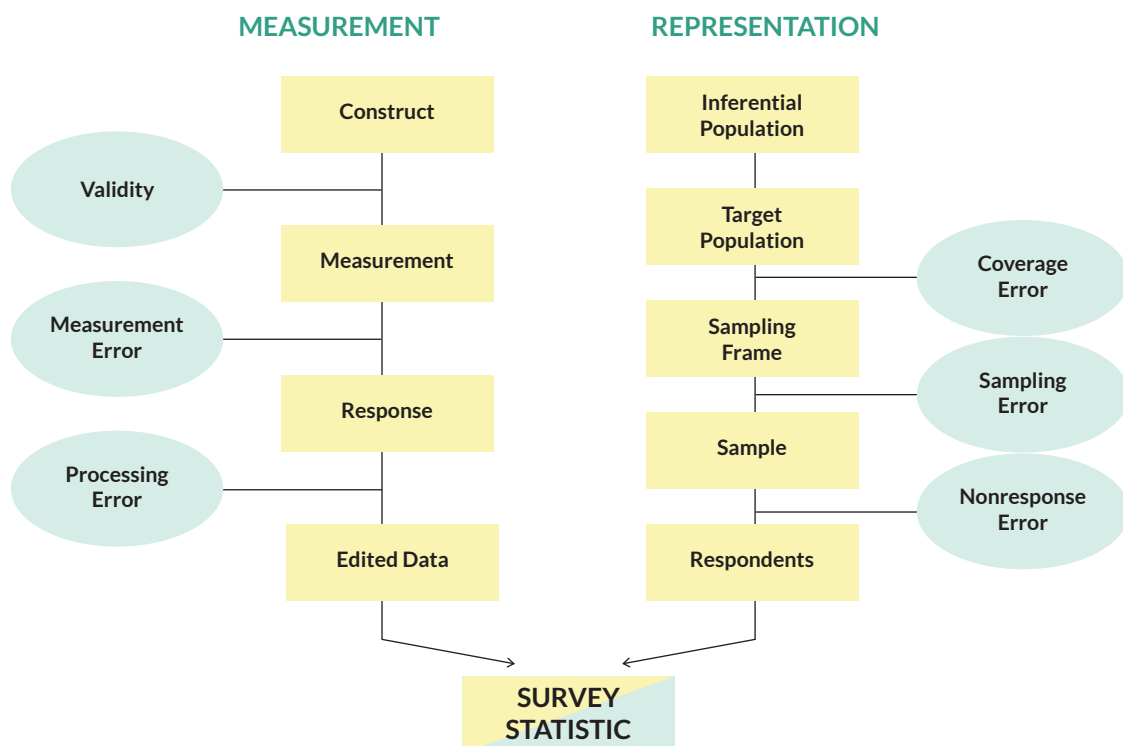
857), "Two separate inferential steps are required in surveys – the first inference is from the response to a question for a single respondent and the underlying construct of interest to the measurement. The second inference is from an estimate based on a set of respondents to the target population. The first of the inferential steps has been the focus of psychometric studies of measurement error, of simple response variance in surveys, and of correlated response variance in surveys. The second of these inferential steps focuses on coverage, nonresponse, and sampling error properties of sample-based statistics."

Figure 3.1 shows that the TSE approach takes the most common error types identified in survey methodology literature into account. Traditionally, the most important distinction made is between sampling errors (SE) and non-sampling errors (NSE).

Sampling errors result from the use of samples to extrapolate findings to broader target populations. Because a sample is not identical to the population, the probable difference between sample estimates and population parameters is known as the estimate error.

Sampling errors are unavoidable in research based on probabilistic or non-probabilistic samples. The optimal solution consists of measuring the (average) impact of sampling errors on sampling statistics by calculating the mean square error, which can be computed depending on (probabilistic) sampling designs. However, it should be noted that SE always represents only a minor fraction of TSE, despite the fact that some researchers show a fixation with SE rather than carrying out wider investigations that include NSEs.

Indeed, most TSE consists of non-sampling errors (NSEs), a general category that encompasses all survey error sources other than SE. NSEs are diverse, ranging from questionnaire construction to questionnaire administration, interviewer-interviewee interaction, and data analysis, as well as other factors.

For these reasons, it seems clear that NSEs could theoretically be avoided because they are the result of flawed survey implementation. However, in practice surveys tend to be heavily affected by errors whose frequency and impact are proportional to the number of interviews.

Table 3.1 shows agreement between various classifications of NSEs proposed over the years although different terms have been used to describe them.

This chapter addresses the methodological

Table 3.1: *Non-sampling Error Source Types. Source: Oliveri (2018).*

Vaccina (1980)	Lessler and Kalsbeek (1992)	Biemer and Lyberg (2003)	Grover and Lyberg (2010)
Identification of the statistical unit	Frame	Frame	Coverage (representation area)
	Non-response	Non-response	Non-response (representation area)
Assessment of the features of the statistical unit		Specification	Validity (measurement area)
	Measurement	Processing	Processing (measurement area)
		Measurement	Measurement (measurement area)

challenges to surveying EMMs resulting from the major error sources described above. Each error category represents a theoretical and practical challenge for scholars interested in designing research with EMM populations.

3.2 Controlling Errors of Validity: Definitions of EMMs and the Problem of Integration

Validity is the extent to which an instrument measures what it is designed to measure. Validity is of particular importance in the context of research topics related to EMMs and EMM integration.

Indeed, assessing validity requires clearly framed operational definitions of the theoretical concepts underlying a particular questionnaire instrument. Concepts should be defined before questionnaires are constructed and sample units are selected.

In other words, measurement strictly depends on theoretical and operational definitions of the phenomena under investigation. The integration of EMMs is no exception to this rule, but it is more complex due to the diverse points of view from which individual studies investigate these populations. The contributions of multiple perspectives—psychological, sociological, economic, and political—significantly enhance the survey research literature.

Because measurement issues such as item selection and scale construction will be reviewed in the WG3 YEAR 2 Report, we focus here on the various components of the expression "Ethnic minority and migrants' integration."

Ethnic. What is meant by the term 'ethnicity'? How and why does the concept differ from "race," "culture," or "civilization"? Similarly, what is "ethnic identity"? And how do ethnicity and ethnic identity differ from one another (Constant & Zimmermann,

2007)? What is the difference between ethnic identity and related concepts such as "national identity" and "cultural identity"? When we use the term ethnic identity, are we referring to a one-dimensional concept or to multiple identities coexisting within a single individual?

In general, ethnic groups may be thought of as communities that identify themselves in terms of differences and similarities based on a series of shared traits or markers (Berthoud, et al., 1997). Blumer (1996) created an exhaustive classification system for the distinctive features of ethnic groups that integrates: real or putative shared ancestry; memories of a shared past; a focus on one or more symbolic elements such as kinship, religion, and language; and a shared territory, nationality, or appearance. An additional important element is an awareness of group belonging.

Minority. This term suggests that people who move from their original country to a host country are fewer in number and weaker in terms of power than the host residents. This can also be a matter of territorial reference, however. In fact, migrants do typically constitute minority groups in host communities, particularly in areas in which migration is a relatively recent phenomenon. Migrants may tend to congregate in certain neighborhoods, however, sometimes rendering local residents a numerical minority. Such expanding migrant communities may reach a point at which they can no longer be considered "minorities." For example, this is what occurred several centuries ago in Southern Italy after a migratory wave of thousands of Albanians escaping Ottoman rule. These people (the Arbëreshë people) are correctly considered an ethnic minority in Italy today, but over time, they are a statistical majority in the cities in which they are concentrated.

Migrants. There are many reasons and motivations that compel people to migrate. Economic motivations drive people to migrate in search of improved material well-being. Political motivation propels people to attempt to escape wars or dictatorial regimes. Cultural migration occurs when migration is a feature of a family's lifestyle. Cultural migration occurs because of lowered barriers to movement between countries, such as the recent expansion of the European Community (Luthra et al., 2014). Clearly, economic immigrants should also be distinguished from refugees, asylum seekers, or Roma and other itinerant peoples.

The concept of "integration" also merits attention. Integration—or the lack thereof—can be economic, social, political, or cultural in nature. These dimensions should be investigated in detail in order to clarify a researcher's biases and research objectives. In general terms, Berry (2005) demonstrates that there may be multiple outcomes of the encounter between the culture of origin and the host culture, including "assimilation," which occurs when groups or individuals leave their original culture behind under the influence of or contact with other cultures. The term integration allows for the possible maintenance of one's original culture even in the presence of interaction with the host culture. Separation consists of maintaining one's own culture of origin while avoiding or minimizing interaction with other cultures. Finally, marginalization describes cases in which a subject has little interest in his/her own culture or interaction with other cultures.

3.3. Controlling Coverage Errors

Coverage errors are a significant problem in EMM research because migrant populations are by definition mobile and elusive and can be difficult to access. An important first step is consequently to carefully define the target population in order to avoid over- or under-representation.

Indeed, survey settings are inherently diverse. For example, interviews can be administered at governmental reception centers for asylum seekers, in which case researchers may be able to gain access to a list of all residents.

Gaining access to migrant populations in "informal settlements" tends to be far more complex, however (Médecins Sans Frontières, 2016; 2018). Under such conditions, in which lists are unavailable and the number of residents is unknown, the population is elusive since migrants sometimes occupy houses illegally and are highly mobile in terms of time and space.

Yet another research context occurs when a target population of EMMs is part of a larger resident community in a city and can only be found at specific meeting locations.

Examining the problem of accurate representation of migrant populations in public registers, Kraler & Reichel (2010: 68-69) offer a thorough overview of the notion of coverage: "Insufficient coverage of the population of interest, most notably migrants, means over- or under-representation of certain groups. Generally, stocks of migrants are far better covered than flows; and within flows it is mainly outflows/emigrations, which are inadequately measured. This significant problem in measuring migration movements stems from the fact that most emigrants do not register their departures (i.e., de-register at the local authorities). Migration in- and outflows are principally measured on the basis of registrations of residents. However, there are no incentives to de-register at the local authorities and migrants often do not inform about their departure. Besides not informing the authorities about emigration, immigrants might delay the registration of their residence. The delay in registering arrivals leads to biased immigration numbers. These problems concern all countries measuring migration movements with administrative registers. In turn under-coverage of outflows leads to over-coverage of stocks and hence leads to over-estimations of population stocks. Yet, the problem of under-coverage does not only concern migration movements but also other flows of statuses. All registers not auto-

matically updated depend on (often voluntary) information of the persons concerned. For instance, in Austria the naturalizations database is not linked to the social insurance database, this means that changes of citizenship status are not automatically updated. Thus, in case persons do not voluntarily inform about their change of citizenship, the registers keep the wrong status and over-estimate the number of non-nationals covered by social insurance...."

The authors continue, adding, "...In addition to coverage problems in registers, surveys equally have their difficulties in properly covering migrant populations. Most sample surveys, which aim at covering the general population, under-estimate migrant stocks... Besides general under-representation of migrants and minorities in surveys, the sample sizes of the groups of interest such as non-nationals or foreign-born are often too small to allow for statistically significant calculations. Thus, the share of non-nationals in the EU27 was 6.2 percent in 2008, while the unweighted average was 7.7 per cent per country... While the number of observations is still sufficient for broad analytical purposes, breaking down the sample by additional variables such as gender, age and country of origin further reduces the number of observations to the point that no statistical calculations are possible anymore...." (p. 69).

3.4 Controlling for Non-response Errors

Because non-response errors alter the probabilistic nature of samples and extrapolate respondents' opinions to non-respondents, they constitute a significant problem for survey-based research. Non-response may refer to individuals who decline to participate in a study, which is called unit non-response. In other instances, individuals may fail to answer particular questions, resulting in item non-response.

Most of the strategies adopted by survey methodologists to manage such error sources involve data manipulation, including data imputation, data weighting, and the estimate calibration (Chao & Chiang, 2006; Lohr, 2009; Kim & Park, 2010, Lumley, 2010).

Arguably, however, the most effective approach to minimizing this error source involves the quality of interviewer-interviewee interaction. Interviews are ultimately based on interpersonal relations and trust, and the more agreeable the experience is for both parties, the more likely an interviewee will be willing to cooperate. One useful approach to addressing this question is to employ interviewers who belong to the same ethnic group as interviewees (Aronson Fontes, 2008; Adida et al., 2016).

3.5 Controlling Measurement Errors

Measurement errors generally refer to:

1. Issues in questionnaire construction, the formulation of questions, and wording
2. Interviewer behavior and interviewer effect
3. Interviewees' reactions to measurement stimuli.

A major issue faced by EMM researchers resides in how questions are formulated. Indeed, because topics often deal with sensitive subjects, such as religion, personal history, and identity, EMM populations may perceive certain research questions as obtrusive. This per-

ception can easily generate social desirability that in turn prompts untruthful responses.

On the other hand, language barriers and misunderstood questions can further exacerbate the construction and administration of questionnaires. Indeed, carrying out research in multi-cultural, multi-lingual settings—which by definition applies to EMM populations—demands particular attention to how questions and instruments are constructed. Language issues also influence the formulation and eventual translation of questions into target-population languages. Language issues pose specific challenges in terms of the comparability of data and results, particularly when several surveys are conducted at different intervals with the same ethnic group or when the same survey is conducted across a number of ethnic groups.

Kraler and Reichel (2010) have reviewed the comparability of migration data, focusing from national and international surveys that are either general or targeted, as well as the availability and comparability of data and results regarding integration.

Wolf et al. (2016) have explored the intercultural harmonization of survey questions.

Schütz (1999) and Woyde (2001) have used quality metrics to assess the quality of translations in the industrial field. A similar approach could be adapted to EMMs research.

3.6 Controlling Processing Errors

Processing errors typically involve problems in data entry, manipulation, or analysis and weighting procedures. However, it is probable that all surveys suffer to some extent from these errors. Because they are essentially unavoidable, it is unlikely that they have more

negative effects on EMM research.

Particular caution is recommended in EMM research, however, when researchers attempt to use inferential procedures based on non-probabilistic samples. An informed use of complex models is required in such cases to adjust for non-random selection, a process that is frequently questioned and is not generally considered effective (McCreesh et al., 2012).

3.7 Controlling Sampling Errors

This section requires a higher level of coordination with the work carried out by WG1 and WG2. Indeed research groups have adopted a variety of strategies at national and local levels to accommodate the challenges associated with sampling mobile, hidden, and hard-to-reach populations (See McFall, et al., 2015).

Indeed, sampling migrants represents a genuine challenge, particularly for the construction of probabilistic samples. The most relevant issue, as noted in section 3.3, involves sizing and identifying the frame. Guidelines for sampling hidden populations have been made available, in particular to support the preservation of randomness. Kish (1991) and Kalton (2001) have both reviewed general problems in surveying hard-to-reach populations, while Verma (2013) offers extensive guidelines for sampling elusive populations in the allied field of child labor.

Blangiardo (1996), Pratesi & Rocco (2002), & Baio, et al. (2011) suggest using center sampling.

Other well-known probabilistic sampling designs for hard-to-reach or mobile populations include adaptive cluster sampling (Brown, et al., 2013) and Time-Location Sampling (TLS) (Kalton, 1991; 2001; 2009; McKenzie & Misti-

aen, 2009).

Because sampling designs are complex, computing estimator variance is not necessarily *straightforward*. However, researchers who have introduced or used these methods offer feasible solutions (Mecatti & Migliorati, 2003; Karon & Wejnert, 2012).

In many cases, however, achieving randomness merely remains wishful thinking and non-probabilistic sampling designs remain more feasible options. Jandl, et al. (2008) suggest snowball sampling (Natale, 1998) and respondent-driven sampling (Heckathorn, 1997; 2002), for example. In both cases, efforts have been made to build estimators and perform inferential statistics. Vehovar, et al. (2016) provide an overview of issues and opportunities in inferential procedures involving non-random sampling.

3.8 Communicating about Findings

This section refers to common mistakes and/or omitted information in how survey results are presented. Professional associations such as the American Association for Public Opinion Research–AAPOR (2010) offer recommendations of best practices to help avoid commonly made errors. As Moy and Murphy (2016: 24) have reported, "This information includes (a) who sponsored the research, who conducted the study, as well as who funded it; (b) the exact wording and presentation of questions and responses; (c) a full description of the population under study and sampling frame used to identify it; (d) a description of the sample design, and method of respondent selection (i.e., using probability or nonprobability sampling); (e) sample sizes and a discussion of the precision of the findings, including

estimates of sampling error and weighting or estimating procedures; and (f) the method(s) and dates of data collection. … In their review of nearly 500 survey-research-based articles published in four mass communication journals. Between 2001 and 2010 using the TSE framework, Ha et al. (2015) found the lack of information on response rates and survey limitations to be common problems."

Although guidelines for every type of survey research exist, they appear even more crucial for EMM research. This is because topicality and related ideological and political commitments can cause the distribution of erroneous information when survey error or response rate estimates are not explicitly reported or when inference is performed without due caution.

References

ADIDA, C.L., FERREE, K.E., POSNER, D.N., & ROBINSON, A.L. (2016). Who's Asking? Interviewer Coethnicity Effects in African Survey Data. Comparative Political Studies, 49(12): 1630-1660.

AMERICAN ASSOCIATION FOR PUBLIC OPINION RESEARCH–AAPOR. (2010). Disclosure standards. Accessed June 24, 2018 from https://www.aapor.org/Standards-Ethics/AAPOR-Code-of-Ethics/Survey-Disclosure-Checklist/Disclosure-Standards.aspx

ARONSON FONTES, L. (2008). Interviewing clients across cultures. New York: The Guilford Press.

BERRY, J. W. (2005). Acculturation: Living successfully in two cultures. International Journal of Intercultural Relations, 29(6): 697–712.

BERTHOUD, R., MODOOD, T., & SMITH, P. (1997). Introduction. In: Modood, T., Berthoud, R., Lakey, J., Nazroo, J., Smith, P., Virdee, S., & Beishon, S. (Eds.), Ethnic Minorities in Britain Diversity and Disadvantage, London: Policy Studies Institute, pp. 1-18.

BIEMER, P.P. (2010). Total survey error design, implementation, and evaluation. Public Opinion Quarterly, 74 (5): 817-848.

BIEMER, P.P. AND LYBERG, L.E. (2003). Introduction to Survey Quality. Hoboken, NJ: Wiley.

BLANGIARDO, G.C. (1996). Il campionamento per centri o ambienti di aggregazione nelle indagini sulla presenza straniera. In AA.VV., Studi in onore di G. Landenna. Milan: Giuffrè, pp. 21–29.

BLUMER, M. (1996). The ethnic group question in the 1991 Census of Population. In: Coleman D., Salt J. Ethnicity in the 1991 Census. London: The Stationery Office, pp. 33–62.

BAIO, G., BLANGIARDO, G.C., & BLANGIARDO M. (2011). Centre Sampling Technique in Foreign Migration Surveys: A Methodological Note. Journal of Official Statistics, 27(3): 451–465.

BROWN, J.A., SALEHI, M., MORADI, M.M., PANAHBEHAGH, B., & SMITH, D.R. (2013). Adaptive survey designs for sampling rare and clustered populations. Mathematics and Computers in Simulation, 93: 108–116.

CHAO, C.T., CHIANG T.C. (2006). A comparison between ratio estimation and post-stratification. JSM 2006, The Section on Survey Research Methods of the American Statistical Association, (OIC, DBS). Accessed June 26, 2018, from: https://ww2.amstat.org/sections/srms/Proceedings/y2006/Files/JSM2006-000620.pdf

CONSTANT, A.& ZIMMERMANN, K.F. (2007). Measuring Ethnic Identity and Its Impact on Economic Behavior. IZA Discussion paper, 3063, 1–23.

GROVES, R.M. & LYBERG, L. (2010). Total survey error past, present, and future. Public Opinion Quarterly, 74 (5): 849–879.

GROVES, R.M., FOWLER, F., COUPER, M., SINGER, E., & TOURANGEAU, R. (2004). Survey Methodology. New York: Wiley.

HA, L., HU, X., FANG, L., HENIZE, S., PARK, S., STANA, A., & ZHANG, X. (2015). Use of survey research in top mass communication journals 2001-2010 and the total survey error paradigm. Review of Communication, 15:39–59.

HECKATHORN, D.D. (1997). Respondent-Driven Sampling: A New Approach to the Study of Hidden Populations. Social Problems, 44(2): 174-99.

HECKATHORN, D.D. (2002). Respondent-Driven Sampling II: Deriving Valid Population Estimates from Chain-Referral Samples of Hidden Populations. Social Problems, 49(1): 11–35.

JANDL M., VOGEL, D., & IGLICKA K. (2008). Report on methodological issues, CLANDESTINO report. Accessed June 24, 2018 from: http://clandestino.eliamep.gr/wp-content/uploads/2009/10 /nclandestino_report-on-methodological-issues_final12.pdf

KALTON, G. (1991). Sampling flows of Human mobile populations. Survey methodology, 17(2): 183-194.

KALTON, G. (2001). Practical methods for sampling rare and mobile populations, Proceedings of the Annual Meeting of the American Statistical Association, August 5-9, 2001. Accessed June 24, 2018 : http://www.asasrms.org/Proceedings/y2001/Proceed/00454.pdf

KALTON, G. (2009). Methods for oversampling rare subpopulations in social surveys. Survey methodology, 35(2): 125-141.

KAPPELHOF, J.W.S. (2015). Surveying ethnic minorities: the impact of survey design on data quality. The Hague: Netherlands Institute for Social Research.

KARON, J.M. & WEJNERT, C. (2012), Statistical Methods for the Analysis of Time–Location Sampling Data. Journal of Urban Health, 89(3): 565–586.

KIM, J.K. & PARK, M. (2010), Calibration Estimation in Survey Sampling. International Statistical Review, 78 (1): 21–39.

KISH, L. (1991). Taxonomy of Elusive Population. Journal of Official Statistics, 7(3), 339-347.

KRALER, A. & REICHEL, D. (2010). Statistics on Migration, Integration and Discrimination in Europe – PROMINSTAT Final Report, International Centre for Migration Policy Development. Consulted June 24, 2018: http://research.icmpd.org/fileadmin/Research-Website/Project_material/PROMINSTAT_File_Exchange/PROMINSTAT_FINAL_REPORT.pdf

LESSLER, J.T. & KALSBEEK, W.D. (1992). Non-sampling Error in Surveys. New York: Wiley.

LOHR, S.L. (2009). Sampling: design an analysis. Pacific Grove, CA: Brooks/Cole Publishing Company.

LUMLEY, T. (2010). Complex Surveys: A Guide to Analysis Using R. Hoboken, NJ: Wiley & Sons.

LUTHRA, R., PLATT L., & SALAMOŃSKA J. (2014). Migrant diversity, migration motivations and early integration: the case of Poles in Germany, the Netherlands, London and Dublin, LEQS Paper No. 74/2014. Consulted June 26, 2018: http://www.lse.ac.uk/

european–institute/Assets/Documents/ LEQS–Discussion–Papers/LEQSPaper74.pdf

MCCREESH, N., FROST, S.D., SEELEY, J., KATONGOLE, J., TARSH, M.N., NDUNGUSE, R., JICHI, F., LUNEL, N.L., MAHER, D., JOHNSTON, L.G., SONNENBERG, P., COPAS, A.J., HAYES, R.J., & WHITE, R.G. (2012). Evaluation of respondent–driven sampling. Epidemiology, 23(1): 138–147.

MCFALL S., NANDI A., & PLATT, L. (2015). Understanding society. UK household longitudinal study: user guide to ethnicity research. Colchester: Institute for Social and Economic Research. Consulted June 26, 2018: https://lucindaplatt.files. wordpress.com/2016/01/6614__ ethnicity_ guide___2ndedition.pdf

MCKENZIE, D.J. & MISTIAEN, J. (2009). Surveying migrant households: a comparison of census–based, snowball and intercept point surveys. Journal of the Royal Statistical Society A, 172(2): 339–360.

MECATTI F. & MIGLIORATI S. (2003), Center sampling: a strategy for elusive population surveys. Statistica, LXIII (3): 537–560.

MEDECINS SANS FRONTIÈRES (2016). Fuoricampo. Consulted June 24, 2018 http://fuoricampo. medicisenzafrontiere.it/Fuoricampo.pdf

MEDECINS SANS FRONTIÈRES (2018). Fuoricampo. Consulted June 24, 2018 http://fuoricampo. medicisenzafrontiere.it/Fuoricampo2018.pdf

NATALE, M. (1998). Determining irregular foreigners in the Italian population. In: Delaunay, D. & Tapinos, G., La mesure de la migration clandestine en Europe. Volume 2: rapport des experts, Luxembourg: Eurostat Working Papers, Series 'Population et conditions sociales', 3/1998/E/N°7. Consulted on June 24, 2018 http://edz.bib.uni– mannheim.de/www–edz/pdf/eurostat/01/ KS–AP–01–007–FR–I–FR.pdf

MOY, P. & MURPHY, J. (2016). Problems and Prospects in Survey Research. Journalism & Mass Communication Quarterly, 93(1): 16– 37.

OLIVERI, A.M. (2018). Administering Face–To– Face Structured Questionnaires In Tourism Research. In Nunkoo, R. (Ed.), Handbook of Research Methods for Tourism and Hospitality Management, Cheltenham: Edward Elgar, pp. 259–270.

PRATESI, M. & ROCCO, E. (2002). Centre sampling for estimating elusive population size. Statistica, LXII (4): 745–757.

SCHÜTZ, J. (1999), Deploying the SAE J2450 Translation Quality Metric in Language Technology Evaluation Projects. Translating and the Computer 21 – Proceedings of the Twenty–first International Conference on Translating and the Computer, 10–11 November 1999, London. Consulted June 24, 2018: at http:// mt–archive.info/Aslib–1999–Schutz.pdf

VEHOVAR, V., TOEPOEL, V. & STEINMETZ S. (2016). Non–probability Sampling. In: Wolf, C., Joye, D.Smith, T.W., & Fu Y. The SAGE Handbook of Survey Methodology, London: SAGE, pp. 329–345.

VACCINA, F. (1980). Evitabilità e inevitabilità dell'errore. In Trentini, G. (Ed.). Manuale del colloquio e dell'intervista. Milan: ISEDI, pp. 9_1–9_34.

VERMA, V. (2013). Sampling elusive populations: Applications to studies of child labour. Geneva: International Labour Organization.

WOLF, C., SCHNEIDER, S.L., BEHR, D., & JOYE, D. (2016). Harmonizing Survey Questions Between Cultures and Over Time. In: Wolf, C., Joye, D., Tom W Smith, T.W., & Fu Y. The SAGE Handbook of Survey Methodology, London: SAGE, pp. 502–524.

WOYDE, R. (2001). Introduction to the SAE J2450 translation quality metric. Language International, 13(2): 37–39.

CHAPTER 4: An Overview of Major Immigration Integration Indices Used in Europe

Esther Lopatin

Introduction

This section provides an overview of three widely used immigration integration indices that reveal measurement issues in efforts to measure integration success in Europe. The indices discussed here originated with the [Council of Europe](#), [the Migration Integration Policy Index (MIPEX)](#), and [the OECD](#). An immigration integration index is useful to the extent that it captures the criteria that policy makers believe are important. Consequently, this section analyzes the criteria measured by these indices measure and discusses whether the criteria adequately reflect the needs of policy makers. Following a review of the literature associated with immigration indices, we briefly describe the major attributes of the three indices before comparing them.

Literature Review

As Schlueter, et al. have shown, immigration policies often determine host-country attitudes towards immigrants, and not, as could be believed, vice versa. Specifically, they demonstrate a correlation between the permissiveness of a country's pre-existing integration policies and local citizens' subsequent openness towards immigrants ([Schlueter, Meuleman, & Davidov, 2013](#)). More importantly, the authors used each country's

MIPEX ranking as the independent variable, and citizen's feelings towards immigrants (as revealed in responses to a questionnaire) as the dependent variable. This enabled them to show that MIPEX is intended as a metric for assessing integration policy, although it does not assess the degree of success in terms of integration. The authors do not mention efforts to measure the extent to which the immigrant community subscribed to democratic norms.

Caselli evaluated an integration index (based on responses to a questionnaire to by immigrants aged 18 and over) developed by the ISMU Foundation in Italy that resembles the OECD index in several respects (reviewed later in the paper). The ISMU defines integration as follows: "The multi-dimensional process leading to peaceful cohabitation, within a particular historical and social setting, of culturally and ethnically different individual and groups, founded on mutual respect for ethnic-cultural differences, provided that these do not impair fundamental human rights or endanger the democratic institutions" ([Caselli, 2015](#)). The author notes that although this definition explicitly mentions "mutual respect" between host and immigrant populations, the index itself does not assess host-population attitudes towards immigrants. This suggests that surveys should include questions about immigrant perceptions of local population attitudes towards them. It should be noted that the ISMU index,

which is based on immigrant responses to a questionnaire, includes certain unique features that are not present in the MIPEX or OECD indices. For example, it asks whether immigrant respondents have a sense of belonging to Italy, agree with the Italian lifestyle, would approve of their child's marriage to an Italian, and are interested in Italian current affairs. The author suggests that the indices should also include an indicator reflecting political participation.

While the index does capture a number of relevant aspects of immigrant attitudes towards host country values, it appears to focus on cultural values and does not explicitly assess respondent acceptance of democratic values.

Wong and Tezil developed an integration index that examines social rather than economic integration in Canada. The index is based on these four axes: Social and civic participation, political participation, sense of belonging in Canada, and experience of discrimination. Drawing on earlier survey data from the 2002 Ethnic Diversity Survey, the authors used factor analysis to assess the extent to which 19 variables captured these dimensions, rejecting 11 variables. Interestingly, they also concluded that degree of participation in civic and religious activities and experience of discrimination should not be included in the index (Wong and Tezil, 2013). Most of the variables that they retained measured attitudes including trust in neighbors and sense of belonging, which is perhaps unsurprising since the index was intended to assess social (as opposed to economic) integration. Their emphasis on using a statistically rigorous approach to select variables is also worth noting.

The authors then administered the index to several populations, finding few differences between the scores of native-born Canadians and foreign-born immigrants. They also found that race significantly influenced scores, with Filipinos (with a score of 9) ranking highest and Koreans ranking lowest (with a score of 12.7). Especially noteworthy was the fact that the scores of Canadian-born, non-Caucasian minorities were lower than those of foreign-born racial minorities, which suggests a significant downward trend in the rate of integration of immigrants during the transition from first to second generation. The authors conclude that this finding is consistent with other cited studies. Finally, the authors observe that this negative trend does not reflect exclusively the experience of discrimination because, as stated earlier, it was found to be a non-significant variable.

Although this index does not explicitly measure attitudes about democratic norms, it appears to capture an important related phenomenon among second- and third-generation immigrants: A tendency for the descendants of immigrants to hold less positive attitudes toward their host countries than previous generations.

Helbling uses correlations between several existing integration and citizenship indices in order to evaluate and compare then. The purpose is to demonstrate what the indices truly measure while illustrating the current state of research. Helbling's comparison shows strong positive correlation between the indices (CPI, Barrier index, and Koning) in terms of citizenship policies, indicating that most of the indices (with the exception of a low correlation between the Barrier Index and CPI) measure similar phenomena (Helbling, 2013: 563). This raises the question of the utility of seemingly similar indices.

Goodman developed a civic integration index in order to measure the civic requirements

for settlement naturalization and migration gradually introduced by 15 European countries. Her study reflects a general but variable increase in civic integration scores in all 15 countries. In countries such as Germany and Denmark, the change was significant, whereas in Portugal, Belgium, and Greece, there was little observable change and no measurable change in Italy or Spain). In my opinion, Goodman's method of weighting and ranking different policies lacks clarity. For example, why do requirements imposed on family members wishing to enter the country rank 0.5, the same ranking as taking an oath? Because Goodman does not explain the rational for these ranking decisions, the validity of her findings is limited.

K. Burkin, et al. comprehensively reviewed the use of indicators to evaluate refugee integration ([Burkin, Huddleston, and Chindea, 2014](#)). Although the study notes that refugee integration presents additional challenges compared to the integration of other immigrants groups, there is considerable overlap in the two populations' needs and in the legal and policy infrastructure required to support the integration process. The authors present the results for the "Integration Evaluation Tool" (IET), an integration index developed under the auspices of the European Refugee Fund Community Actions used to collect data in a group of Western (Ireland, France, Austria, and Sweden) and Central European Countries (Poland, Bulgaria, Romania, and Slovakia). The present discussion is restricted to the results for Central Europe because they clearly demonstrate the challenges to integrating refugees that host countries confront.

The authors note that, apart from specific national concerns associated with the Common Immigration and Refugee policy, the Hague Program calls for EU member-states to develop indicators to monitor refugee integration in their own countries. Differentiating between a Common Immigration and Refugee Policy and evaluation methods (and the extent to which policies are implemented) is an initial challenge in achieving success in immigration and refugee integration across Europe. The UNHCR, which has been tasked with ensuring refugee protection, recognized the need for an integration index that is specifically for refugees, which led to the development of the IET (an online data-collection tool) and the Migration Policy Group and Central European Stakeholders. The purpose was to assess the extent to which refugees are provided with the rights to which they are entitled.

The IET evaluates integration outcomes (e.g. achievements in employment, education, housing), policy indicators (e.g. how well policies are implemented by states) and inputs (how much a country invests in the policy). Field-based experts typically complete the IET, although in some cases refugees themselves contribute data. A significant quantity of data has been collected: Thirty pages of data were required for the 2013 results. This substantial mass of data allowed the authors to draw several less-than-obvious conclusions. For example, they found important discrepancies between countries in terms of the amount of time required before refugees are able to obtain work permits and the length of time between asylum applications and permission for applicant's children to enroll in school.

The IET also showed that a number countries lacked adequate data concerning how many asylum-seekers' children even needed schooling. The index also revealed that certain local housing regulations violate international standards that are guaranteed for asylum

seekers. These irregularities confirm the need for a geographically broad-based, comprehensive index.

Although it discussed the general results, this report did not provide the detailed statistical analysis that might have supported understanding of the correlation between policies and outcomes.

Indices

According to the Council of Europe (CE), successful integration of immigrant populations is based on "a common framework of legal rights [...] freedom of choice of religious and political beliefs, cultural and sexual affiliation, within the framework of basic democratic rights and liberties [...] immigrant groups will cease to be distinctive in culture [...] minority and majority groups learn from each other and take aspects of each other's culture" (Council of Europe, 2015). The CE defines the three central components of successful integration as: 1. Immigrants should share the basic values of the host country, 2. Immigrants should cease to have a distinctive culture, and 3. Immigrants and nationals should benefit from each other. The CE stresses the importance of interaction between immigrants and host societies. One might therefore expect the indicators contained in the Council's index to reflect this emphasis, but this is not the case. In fact, the actual indicators as listed below represent the following categories: Participation in the labor market, income, housing, health, civic engagement, discrimination (against immigrants), education, and the law (i.e., criminality rates). Although such indicators can be readily quantified, they shed little light on the criteria that the CE considers important for successful integration.

The OECD simply states that, "Successful integration means equal opportunities for immigrants, ensuring they become an integral part of society" (OECD, 2015). This quotation refers to two key topics: 1. The host country is responsible for ensuring that immigrants are provided with the rights and opportunities that they need to succeed, and 2. Immigrants' economic success is identical to successful integration. These assumptions naturally lead to the conclusion that measuring economic success is one way of measuring immigration success. The OECD immigrant questionnaire attempts to measure integration outcomes (i.e., are the immigrants showing adequate economic progress?), but in fact inquires about very few indicators (e.g. naturalization rate and level of discrimination felt by the respondent) that might indicate the extent to which governments are protecting immigrants' rights or providing opportunities.

The two indices described above concentrate on outcomes. By contrast, the MIPEX, which is based on evaluations by country-specific experts, is solely concerned with measuring the extent to which countries respect immigrants' rights. According to MIPEX, "Integration [is] the concept of equal opportunities for all. [...] In civic terms, all residents can commit themselves to mutual rights and responsibilities on the basis of equality." By juxtaposing the concept of equal rights with residents' responsibilities, MIPEX authors assume that immigration success depends on the degree to which immigrants achieve equal rights. Accordingly, the questionnaire exclusively measures the extent to which laws and policies meant to facilitate integration are implemented in a given country; indicators of outcomes are not included. For example, the index measures access to

citizenship, labor market mobility, and opportunities for family reunification, but it does not assess the extent to which labor market mobility improves the economic circumstances of immigrants. Nor does it assess the target community's genuine commitment to equal rights and equality, despite the fact that that this is an explicit goal of integration.

The table below compares the indicators used by the OECD and CE indices. Although a number of broad indicators are included in every index, MIPEX evaluates them differently. For example, for the category "labor," MIPEX asks questions such as, "Can legal migrant workers and their families access and change jobs in all sectors like nationals?" and "Are foreign residents able to accept any private-sector employment under equal conditions as nationals?."

Discussion

The two indices that measure outcomes, the OECD and CE, clearly focus on what could be called "dry statistics." Although both the CE and the OECD confirm the importance of a "common framework of legal rights, freedom of choice of religious and political beliefs, cultural and sexual affiliation, within the framework of basic democratic rights and liberties" and "ensuring they become an integral part of society," they do not include indicators that provide evidence of the extent to which immigrants share these values with their host country. In other words, although sharing common values would appear to be an important component of successful integration, the subject is not addressed by any of these indices. The values that Europeans consider to be important are well known—as shown by a sample survey below. The important question is whether current indices can be considered complete if they fail to indicate whether immigrants also share these values.

Measurements of Indicators – A Comparison of OECD, the Council of Europe and MIPEX	OECD	CE	MIPEX
LABOR MARKET			✓
Employment rate	✓	✓	
Activity rate (active/inactive in work force)	✓		
Unemployment rate	✓	✓	
Share of inactive who wish to work	✓		
Working hours	✓	✓	
Self-employmet	✓	✓	
Share of employment in public services	✓		
Proportion in dangerous/dirty jobs		✓	
Proportion in key professions		✓	
Vocational/professional training	✓	✓	
INCOME			
Poverty rate	✓		
In-work poverty rate (number of working people living in poverty)	✓		
Share of households with bank account/overdrawn bank account	✓		
HOUSING			✓
Overcrowding	✓	✓	
Housing quality		✓	
Proportion in public, rented/self-owned housing	✓	✓	
Proportion claiming benefits, child benefits, maternity, or state pension		✓	
Housing cost overburden rate	✓		
Share in substandard dwellings	✓		
Average household size	✓		
Household composition	✓		
HEALTH			✓
State of people's health	✓		
Share of people with unmet medical needs	✓		
Share of people who have not seen a doctor	✓		
Proportion of immigrants affected by major diseases/cause of death		✓	
Birth rate		✓	

Chapter 4: An Overview of Major Immigration Integration Indices Used in Europe

Measurements of Indicators – A Comparison of OECD, the Council of Europe and MIPEX	OECD	CE	MIPEX
Inter-ethnic marriage		✓	
CIVIC ENGAGEMENT			✓
Naturalization rate	✓		
Voter participation rate	✓	✓	
Participation in key institutions and organizations		✓	
DISCRIMINATION			✓
Share of immigrants who feel they've been discriminated against	✓		
Share of people who think their areas are a good place for migrants to live	✓		
Perceived economic/social impact of immigration	✓		
EDUCATION			✓
Participation in pre-school education		✓	
Results school leaving certificates		✓	
Take up of adult language training		✓	
Distribution in types of school relative to area of residence		✓	
JURIDICAL			
Comparative data on arrests, convictions, acquittal rates		✓	
Data on racially violent crimes/racial harrassment		✓	
Data on complaints of discrimination and convictions		✓	

CONCLUSION

Despite harmonization of the legal frameworks governing data collection within the EU, national and local practices for surveying EMMs across Europe continue to show dramatic variations for a number of reasons, including:

1. Different interpretations of the legal framework
2. National narratives about EMMs
3. Strategies of colorblindness in some European countries
4. Nationally and locally specific configurations of diversity
5. The research agenda in quantitative studies

Although the mapping exercise presented here was relatively limited in scope, it is intended as a partial overview of the contents and of some key surveys. This introductory overview will be supplemented by the metadata collection of existing surveys to EMMs and by the creation of a data hub.

The central objective was to provide background information about the legal and methodological constraints on and significant gaps in survey research on EMM populations. Future segments of the project will address how the information gathered in surveys is used. The next report will consider the challenge of indicators of integration. As the conclusion of this report clearly shows, designing indicators is not straightforward, largely because they depend upon coherent definition and accessible variables. Only when these are achieved can the problem of interpretation be addressed. This will be the subject of further discussion.

APPENDIX: Questionnaire

COST ACTION 16111 ETHMIGSURVEY DATA - WG3:
Questionnaire on the Legislative Framework, Ethical Principles and Practices Relating to "Ethnic" Data Collection

This questionnaire is intended to gather information on your experience of collecting "ethnic" data in surveys, according to your country's legislation and its concrete implementation. It will take you less than 15 minutes to complete it.

By "legislation", the questionnaire refers to the laws governing the collection and dissemination of statistical data, meaning on the one hand the general laws on data protection and on the other hand the laws that regulate statistical activity.

By "ethnic data", the questionnaire refers to all data alluding to national, ethnic or racial origin, religion, language or nationality (i.e. citizenship). Population categories might refer to migrants, immigrants, refugees, asylum seekers, descendants of immigrants, ethnic or national minorities, etc.

Thank you for taking the time to answer the questions. Answers will be made anonymous in their treatment. Questionnaire to be sent to simon@ined.fr and amparo.gonzalez@cchs.csic.es

COUNTRY:

NAME OF ORGANISATION:

RESPONDENT'S IDENTITY:

We will first collect some information about the legal context on data collection in your country.

Does the law on data protection contain a definition of "sensitive data"?

Yes ☐ No ☐ Don't know ☐

If so, give a list of these "sensitive data":

Is there a prohibition to collect these sensitive data?

Yes ☐ No ☐ Don't know ☐

Are there exemptions (conditions for authorization) to the prohibition that makes it possible to collect these data anyway?

Yes ☐ No ☐ Don't know ☐

Does the law prevent you to design a specific sampling?

Yes ☐ No ☐ Don't know ☐

If yes, elaborate:

Does the law prevent you to collect data about sensitive categories:

Yes ☐ No ☐ Don't know ☐

If yes, elaborate:

WE WILL TALK NOW ABOUT ONE OR SEVERAL SURVEYS YOU HAVE CONDUCTED ABOUT MIGRANTS OR ETHNIC MINORITIES (WITH AN EXTENSIVE DEFINITION), DUPLICATE THESE PAGES IF YOU HAVE SEVERAL SURVEYS TO REPORT ON. IF YOU HAVE DONE MORE THAN ONE SURVEY BUT THE CONTEXT WAS SIMILAR, JUST FILL IN ONE ANSWER.

SURVEY 1

NAME:

DESCRIPTION OF YOUR SURVEY:

PLEASE GIVE BASIC INFORMATION ABOUT THE SURVEY:

- Date of fieldwork

- Size of the achieved sample

- Geographical coverage (Cross countries, National, Local)

- Type and source of sampling (pop. Registers, censuses, telephone book, random route in selected neighbourhoods, etc)

- Use of listing with names/addresses/emails

- Method of data collection (face to face, telephone, internet, etc)

- Average Duration of the questionnaire

- Non response rate

CATEGORIES USUALLY REFERRED TO IN YOUR COUNTRY, AND CATEGORIES YOU HAVE USED IN YOUR SURVEY:

Please mention if these categories are often, rarely or never collected for statistics in your country:

	Often	Rarely	Never	Don't know
Immigrants (foreign born)				
Citizenship/nationality				
Ethnic affiliation or ethnicity				
Race				
Religion				
Language				
Descendants of immigrants (second generation)				
Other, specify:				

Other (specify):

Which categories stated above did you collect in your survey?

FOR SAMPLING:

IN YOUR DATASET:

FOR ANALYSIS:

How did you proceed to deal with the data protection law?

- ☐ I was not aware I had to comply with the law
- ☐ I did not need to comply with this law
- ☐ I had to apply and fill in a form

If you did not have to comply, explain why:

What kind of commitments you had to take to fulfil your legal obligations?

Examples: Commitment to destroy the personal identifiers after X months; Commitment to store the data in a secure server; Specific rules about anonymization of the data; Lack of fine geographical code

If you had to go through a process of authorization, can you explain how it worked, and especially if you had to claim exceptions to the legal framework to collect sensitive data:

If you had to go through a process of authorization, can you explain how it worked, and especially if you had to claim exceptions to the legal framework to collect sensitive data:

Are there information you were not allowed to collect or use because of their sensitivity?