EUROPEAN COMMISSION

CITIZENS AND GOVERNANCE IN A KNOWLEDGE-BASED SOCIETY

Report on methodological issues

prepared by Michael Jandl, Dita Vogel & Krystyna Iglicka

edited by Albert Kraler & Dita Vogel

November 2008

CLANDESTINO

Report on methodological issues, Deliverable D3 prepared for Work Package 2 of the research project CLANDESTINO *Undocumented Migration: Counting the Uncountable. Data and Trends Across Europe*, funded by the 6th Framework Programme for Research and Technological Development under Priority 7 '*Citizens and Governance in a Knowledge-Based Society*', Research DG, European Commission

Abstract

Estimates of irregular migration are scarce, and a number that appears once has a good chance to develop a life of its own and survive for many years in the media. Such numbers reappear in EU level decision making documents, framing arguments for a substantial expansion of migration control. There is no doubt that the situation could be improved. Social scientist *can* do much better than just telling a number, although estimates of the size of irregular populations may never reach the degree of accuracy that estimates of the size of regular populations achieve. In the project CLANDESTINO, data and estimates for 12 EU countries are collected, classified, and the results exposed to scientific and public dialogue.

This report is a basic background document for this effort. Chapter 1 outlines main features of the problem and different ways to define and conceptualise irregular migration. Chapter 2 indicates the current relatively weak state of aggregate estimates on the EU level. Chapter 3 compiles a wide range of estimation methods and discusses their strength and limitations. Chapter 4 lays out quality standards for estimates.

Beyond the aim of retrospectively assessing the state of knowledge, this methodological report also has a prospective aim. It indicates ways to better information in the future. While high quality estimates may not be possible under all circumstances, we would argue that an increasing number of medium quality estimates would already improve the information situation in the European Union considerably.

Table of Contents

Chapter 4: Classifying the quality of estimates

Dita Vogel

Chapter 1: Introduction to the methodological problem

Dita Vogel and Michael Jandl

1. Introduction

"Precise figures are difficult to obtain but recent estimates of illegal migrants in the EU range between 4.5 million and 8 million, with an estimated increase by 350 000 to 500 000 per year". This statement about the number of irregular migrants in Europe is found in a press release of the European Commission that presents a proposal for a Directive on sanctions against employers of irregular migrants – a directive that may costs billions of Euros to the member states if it should be enacted.[1]

Where do these numbers come from? Our search leads us to a Commission Staff Working Document that refers to an external study by a consulting company, indicating that it constitutes the main support for the Commission report (European Commission 2007). With regard to the yearly increase in irregular migrants, the external report says:

"The number of illegal immigrants in the EU is estimated to increase by 500,000 per annum by Wiener Zeitung (2005), and by 350,000 according to Global Migration Perspectives 2005." (GHK 2007: 23)

We did not follow up where the Vienna newspaper took its estimate of a net increase of 500 000 'illegal' immigrants.[2] It is not important whether it was taken over from old estimates in the 1990s, heard from an unnamed expert or invented by a journalist. Fact is that a number – without any reliable source and specification of time frame – made its way through several documents and finally became a 'recent estimate' in a press release of the European Commission that proposes to devote about five times more resources to controlling employers.

We do not want to blame the Commission or the consulting company, as their approach is typical for use of estimates of undocumented populations. It is difficult and cost-intensive to make estimates of hidden populations. Social scientists are often reluctant to come forward with estimates, because they can not achieve a degree of reliability which they are used to from other fields of study. Numbers are scarce (see chapter 2), and journalists long for numbers to frame their articles on irregular migration. Therefore, a number that is quoted somewhere in the press has a good chance to develop a life of its own and survive for many years in the media where it is quoted like a fact, even if the original estimate was accompanied by disclaimers (Clarke 2000: 21).

There is no doubt that the situation could be improved. Social scientist *can* do much better than just telling a number, although estimates of the size of irregular populations may never reach the degree of accuracy that estimates of the size of regular populations achieve. However, it should be kept in mind that numbers about the size of populations are never accurate. We can take for example Germany, a country with a long tradition of keeping full population records. The idea is to have a full count of the population at all times, but the counting involves problems. After the last census in 1987, the numbers for the foreign population had to be reduced by 10 percent, mainly due to underreporting of outflows.

[1] Reference: IP/07/678 of 16 May 2007.

[2] The figure is approximately in the order of the number of apprehensions in the European Union and it could well be that CIREFI data served as the basis for this estimate. In 2002, some 578 000 aliens were apprehended in the European Union in total (Gédap 2004).

Administrative revisions of population data lead to a similar reductions in 2004 and 2007 (Bundesamt für Migration und Flüchtlinge 2006; Deutscher Bundestag 2007). These statistical revisions indicate that official population data are characterized by a considerable degree of unreliability, but nobody sees this as a reason to stop keeping population data. In this sense, the problem of estimating the size of irregular migrant populations can be seen as a particularly difficult case of making population estimates.

There is a range of methods for these estimates, which necessitate different data. They differ in comprehensiveness, sophistication and problem-awareness on diverse issues that are discussed in detail in chapter 3 of this report. So far, there is no way to assess the quality of estimates of the size of irregular migration. The project CLANDESTINO seeks to fill this gap. It will publish classified estimates that enable policy makers and journalists to obtain concise information on the quality of available estimates. In addition, it will deliver more detailed information for particularly interested persons (such as social scientists in research and administration).

Chapter 4 reviews quality criteria for scientific analysis of qualitative and quantitative data and procedures to ensure research quality and reflects how they can be applied to estimates of irregular migrant populations. For concretisation, it summarizes the most important problems that are specific to the estimation of hidden populations in the field of irregular migration.

Finally, a classification scheme is proposed that aims at fulfilling two conditions: Firstly, it seeks to be a good indicator of the scientific quality of the estimates of the size of irregular migration stocks and flows; secondly, it has to be practicable in the implementation; easily understandable and useful for readers outside the field of social sciences. This does not mean that we want to oversimplify measurement problems, as will become clear from our detailed discussions below. Fact-based administration and policy-making still requires detailed assessments of individual estimates. Therefore our overview will try to expose false claims for accuracy or timeliness and will guide the way to more reliable assessments.

2. Terms and Definitions

We will first reflect on terms and definitions and then describe the connection between stocks and flows with reference to irregular migration.

There are a number of terms and expressions used for persons who enter a country illegally, overstay their terms of legal residence, live in a country without a residence permit, and/ or break immigration rules in a way that makes them liable to expulsion: The adjectives irregular, illegal, undocumented, unauthorized or clandestine are combined with the nouns migrants, immigrants, aliens or foreign nationals, and there are expressions like sans-papiers, clandestinos, shadow persons.

The term *"illegal migration*" reflects, in its broadest sense, an act of migration that is "not legal", or an act of migration that is carried out against legal provisions of entry and residence. The European Union, for example, uses the term "illegal migration" in this sense (cf. Jandl and Kraler 2006a). Various authors have argued against this value-neutral understanding of the term "illegal" and have pointed out that the adverb "illegal" associates this type of migration with criminal or otherwise "illicit" behaviour and should therefore be avoided (Sciortino 2004, p.17). Human rights advocates have long argued that the derived noun "illegal migrant" is discriminatory, since "No human being is illegal", as the slogan of a campaign would put it. In a narrow sense, the term "illegal migration" designates the act of entering a country in contravention to the law and is confined to illegal border crossing (but

not overstaying the terms of visas or residence), referring only to a *flow* and not to a stock of persons.

In recent years, the term "irregular migration" has increasingly replaced the designation "illegal migration" in its broader sense. "Irregular migration" in this context denotes a form of migration that is "not regular", "unlawful" or not according to the rules (without necessarily being "illegal", "illicit" or "criminal" in the legal sense). An "irregular migrant" is therefore a migrant who, at some point in his migration, has contravened the rules of entry or residence. The term "irregular" is widely understood as a very broad concept and can refer to both stocks and flows of migrants (cf. Koser 2005).

Another term often used is "undocumented migrant" with the implicit meaning of "a migrant who is not in possession of the required (and correct) residence papers". The term is often used in its broadest sense equal to that of "irregular migrant" defined above. However, "undocumented migrant" is also used as a synonym for "unrecorded migrant", which excludes persons who are documented but nevertheless unlawfully residing in a country, such as rejected asylum applicants pending deportation, persons with a toleration status, and others. For many authors, the term "undocumented migrants" is more neutral, with fewer connotations, and is thus to be preferred (cf. Pinkerton et al 2004, p.1).

"Unauthorized migrant" refers to people who enter or stay in a country without legal authorization. Of course, not everyone residing in a foreign country needs explicit authorization to do so (e.g. if there are free movement rights like within the EU) and we need to interpret "unauthorized" as "not authorized according to the law".

For our purpose, we find all four terms acceptable for the discussion of the indicated phenomenon. While preferring undocumented or irregular migrant, the expressions illegal and unauthorized may be used as synonym in our research. For measurement purposes the differentiation between stocks and flows is most important.

For this project, irregular or undocumented *residents* are defined as residents without any legal residence status in the country they are residing in, and those whose presence in the territory – if detected – may be subject to termination through an order to leave and/or an expulsion order because of their activities.

Irregular *entrants* are persons who cross an international border without the required valid documents, either uninspected over land or sea, or over ports of entry.

Being aware that the same terms are often used with more or less deviating meanings in the literature, it is most important that we are always clear about what type or sub-type of irregular migration a particular author, a particular method or a particular estimate is referring to.

With regard to stocks of residents, it is of particular importance to understand whether the following groups are included in the definition or not.

First, there is the question of *EU citizens*. One of the four core freedoms of the European Union is the freedom of movement – the right to enter and to reside in another Member States. Therefore, EU citizens enjoy rather high protection under EU legislation. However, mobility rights are not unlimited and maybe waived for reasons of public policy or public security, for example because of criminal convictions or drug addiction or on ground of lack of means if an EU citizen has not yet acquired permanent residence rights and thus the full

protection under EU freedom of movement rights legislation.[3] However, the European Court of Justice has traditionally interpreted the right to freedom of movement extensively and has frequently annulled Member States' decisions on expulsion or residence bans for Union citizens. In practice, therefore, there are very few cases in which residence bans or expulsion orders are issued against EU citizens.[4] Citizens of the new member states, however, are subject to transitional rules in all but three EU-15 Member States. Under these transitional rules, their access to other EU Member States' labour markets and hence their right to settlement may be restricted for up to 7 years (2011 and 2014, respectively).

In many countries, illegal work of EU-citizens from recent accession countries is dealt with under the heading of irregular migration, but it does not make them liable to expulsion. Estimates that include EU-citizens usually refer to irregular work by foreign residents, not to irregular residence. In this project, we are primarily interested in the irregular residence of Third country citizens (irregular foreign residence, IFR) and less in irregular foreign work (IFW). We explicitly differentiate between these two broad definitions and indicate if an estimate includes EU citizens.

The second group concerns persons with *seemingly legal temporary residence status*. Third country nationals (TCN) are by our definition illegally resident if they have no status at all, or if their activities would make them liable to expulsion if detected. The latter concerns particularly persons who conduct irregular work on a tourist visa or during visa-free tourism ("Working tourists").[5] In some countries, "working tourists" are assumed to be the majority of irregular migrants, while in others irregular migrants usually do not have any – even temporary – right to be in the country. In addition to working tourists, migrants with a temporary conditional work permit such as seasonal and contract workers may be liable to expulsion because the terms of their contracts are severely broken, for example because they work longer than permitted or receive payment by unit and not by hours. In many countries, it makes a sizeable difference whether these persons are included in the concept of irregular migration or not. It can be projected that these phenomena will be of growing importance if the European Union introduces new programmes for temporary labour migration. Usually, foreign nationals with a regular residence status that conduct irregular labour are not referred to as irregular migrants. Currently, it is reasonable to restrict the term irregular foreign resident to persons without any residence status and to 'working tourists'.

The third group concerns *persons with forged papers*, or persons who have assumed *false identities with real papers*. They may live a regular life unless they have to show their papers to authorities that can discover the falsification. Their life can be very similar to the life of regular immigrants and citizens, involving regular jobs and free movement in the public sphere, or it can be similar to the life of migrants without any status. If the quality of the papers is so poor that the migrant does not dare to show them anywhere, these papers do not make much of a difference to having no papers at all. Many studies are not explicit with regard to these persons. We assume that the persons are counted as irregular as long as their

[3] Directive 2004/38/EC of the European Parliament and of the Council of 29 April 2004 on the on the right of citizens of the Union and their family members to move and reside freely within the territory of the Member States.

[4] Although EU citizens intending to move to another Member States are required to register themselves within three months, non-registration is not sufficient grounds for expulsion or a residence ban and in any case, can not automatically lead to expulsion (or an ex-lege irregular status).

[5] This should not be confused with the legal immigration category of „working holidaymakers" in the United Kingdom.

residence is not registered officially, even if these persons are in possession of papers that may be presented to employers.

The fourth group concerns *persons whose immigration status is pending*. This group includes persons, whose application for regularization is pending and whose application papers prevent their expulsion, or TCN who have submitted an asylum claim. This group also includes persons who have filed a request for status prolongation but still wait for a decision by the time that their limited residence permit runs out. These categories are *not* considered as irregular residents for the purpose of this report

The fifth group concerns persons who are without residence status in the country, but with knowledge and *toleration of the authorities*. This may involve (1) a formal suspension of enforcement action for a specific period of time as in the case of "toleration" ("Duldung") in Germany which is (in the form of a formal document) communicated to the tolerated person or (2) a (documented) administrative suspension of enforcement action for practical reasons[6], whether or not a enforcement is suspended for a specific time period. What mainly distinguishes (1) from (2) is that tolerated persons are not issued documents proving the suspension of enforcement action. Finally, (3) "toleration" may also be due to simple inaction of authorities without a specific decision to suspend enforcement. Whereas one can expect administrative statistics on the first type of toleration, statistics are usually not kept for the second type; in the third type, authorities are usually not be to identify the number of persons who are thus tolerated "informally".

Technically, toleration – whether formal or informal - is only a suspension of an expulsion order and not in itself a legal status – in other words, toleration does not have a legalizing effect and does not change the unlawful presence of the tolerated alien.[7] On the other hand, there is little difference between formal toleration – the formal suspension of an expulsion order for a specific period of time documented by some sort of permit – and comparable legal statuses such as humanitarian stay or subsidiary protection. We thus don't consider formally tolerated persons as irregular residents for this report.[8]

After discussing stocks, we now turn to flows. Inflows are contributing to the stock of irregular residents, and outflows diminish it. Flows can be differentiated into demographic flows, border related flows and status-related flows.

Demographic flows include birth into illegality. While the United States as the country with the most comprehensive *jus soli* recognizes all children born on the territory as US-citizens, most European states are not as open towards the new-born. Therefore, a baby is usually without status in a country when it is born by a mother without status. Thus, births without status are an inflow to the stock of the irregular residents. If irregular residents die, this is an outflow from this stock.

Geographic flows concern movements over a border of a country in break of migration law. On the one hand, it concerns movements over the green or blue border. Persons try to enter illegally and without inspection. In contrast to the American situation in which each entry without inspection is automatically illegal, entry without inspection is the rule inside the Schengen-area of free movement in Europe. Illegal entries into an EU country concern either entries without inspection over the external borders, or entries from another EU country over

---

[6] For example, limited capacity to organise deportation, unwillingness of the country of origin to issue required travel documents; questionable citizenship status, etc.

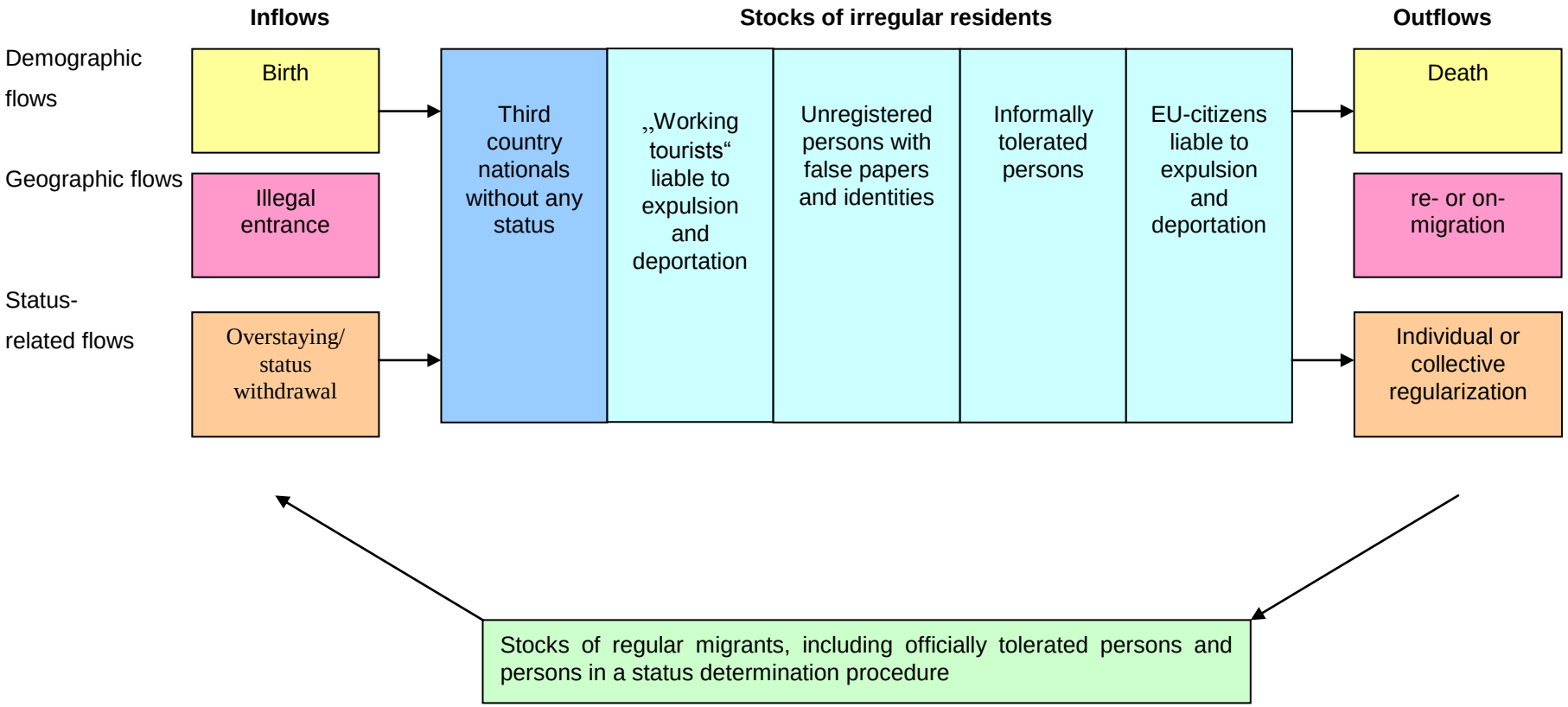
[7] See Sinn et al. 2005 p. 42f on the legal nature of toleration status in Germany

[8] The German official line also seems to suggest that tolerated persons, although not legal migrants, are not seen as illegal migrants either (Blaschke, 2008, 16).

an internal border without the required travel documents. These movements are simultaneously an inflow for one EU country and an outflow for another. Only outflows over external EU borders reduce the number of irregular residents in the European Union, while outflows to other EU countries reduce irregular residence in one country and increase it in another.

Status-related flows concern all flows into irregular residence from a regular status and *vice versa*. The most relevant inflow concerns persons who have entered with a tourist or other temporary visa and overstay the allowed period of residence (visa overstayers). Other persons have lived regularly in a country and stay after their status was withdrawn, for example after the rejection of an asylum application or the withdrawal of a temporary or permanent status after a serious criminal offence. On the other hands, there are status-related outflows from irregular residence. Individuals may be regularized after marriage with a national or as hardship case, or a large number of persons may profit from a collective regularisation programme. The officially recognized toleration that exempts a person from the threat of expulsion for a certain period of time is considered an outflow from illegality, even if the country does not consider this toleration as a regular status. Persons may repeatedly change between regularity and irregularity. Changes of the migration regime can create new statuses or make established statuses available for new groups of people. If irregular migrants make use of the new options, they become regular without being individually regularized. Widened legal options become a functional equivalent to regularisation for them. If migration regimes become more restrictive, formerly legal migrants may continue or repeat their migration practices illegally.

Figure 1: Stocks and flows of irregular migrants



Chapter 2: Past Efforts to Create Estimates on Irregular Migration in Europe

Krystyna Iglicka

1. Introduction

There are enormous difficulties in assessing the scale of irregular migration. These problems are growing when we have to deal with estimates of irregular migration not in one country but in a group of them or in the whole region. In a comprehensive paper on the problems related to the evaluating of illegal migration in the European Union Clarke (2000) stresses that estimating and comparing the numbers of irregular migrations and migrants in EU member states is a very complex and difficult task. Estimating the stock of irregular migrants and the scale of irregular entrances on the level of the European Union is even more challenging. Migrants in irregular situation may be able to move relatively unchecked between countries in the Schengen zone. Border control data cannot be used to the same extent as indicators of illegal migration as in the case of much more intensively controlled Schengen borders (Clarke, 2000: 13). In addition, in the absence of harmonised data collection rules for border control data, significant double counting may occur which makes simple addition of national apprehension figures problematic. Finally, insufficient differentiation between apprehensions of persons leaving versus apprehensions persons entering a country further casts doubt on the quality of the data. Furthermore today's mobility is characterised by new forms and dynamics such as short term or circular migration that are very difficult to be captured by any measures (Salt, 1999).

In this chapter, past efforts to estimate the size of undocumented migration in Europe are reviewed, focussing on those efforts that are often quoted in scientific and political documents and the press. This review shows that the available cross-country estimates are relatively rough and rely on a limited number of sources, which are highlighted in the last section.

2. Past efforts to create estimates for Europe

Due to the enormous difficulties briefly described above, there were few efforts only to estimate the scale of illegal migration into Europe. Past efforts pertained to the estimation of both stocks and flows of irregular migrants. Among the most significant efforts to estimate the scale of irregular migration are:

- Estimates of the stock of non-nationals in Europe by the International Labour Office (ILO, 1991)

- Estimates of the stock of unauthorised migrants in Europe (Papademetriou, 2005a)
- An estimate by Widgren (1994) on annual illegal entrants and proportion of person using smugglers
- Indicators of the flows of illegal migrants to Western Europe (ICMPD, 1994, ICPMD continuous Yearbooks since 1996)
- Estimates of the flows of trafficked migrants from CEE to Western Europe (International Organisation for Migration, 1995, 2000)
- A European Commission staff working document, accompanying the proposal for a directive on sanctions against employers of illegally staying third-country nationals (European Commission 2007) – collecting sources on the numbers of illegal immigrants in EU and methods for these assessments.

There were of course many studies and efforts pertaining to the national levels as well. However, due to the subject of this revision paper, only few of them are quoted here. The studies are presented here in the chronological order in which they became important in the European discussion. However, we have to be aware that some of them relate to flows and others to stocks, and that they include irregular migrants in different and not always fully transparent definitions.

According to the International Labour Office estimates, in 1991 there were an estimated 2.6 million non-nationals in Europe in an irregular or undocumented situation (ILO, 1991). In this *stock estimation* the number of seasonal workers was included as well as those asylum seekers whose applications have been turned down but have not left. The ILO assumption was that irregular foreign stock would be between 10 and 15 per cent of the size of the officially recorded resident foreign population (Clarke, 2000). It is unclear however, how these percentages were derived.

The International Centre for Migration Policy Development (ICMPD) estimated that *inflows* into Western Europe totalled around 250 000- 350 000 in 1994 (Widgren, 1994). According to the majority of experts this still remains the most widely quoted estimate. The figure was calculated on the basis of extrapolation of the number of illegal migrants that reached their destination countries as a reflection of the known numbers of migrants apprehended at the borders (Clarke, 2000). Widgren, on the basis of discussions with border control authorities, multiplied the number of 60 000 border apprehensions by 4-6. He also suggested that about half of the 690 000 asylum applications (approximately 300 000 applications) recorded in 1993 were likely to have been unfounded, thus suggesting that the total number of irregular entries were in the order of 550 000 to 650 000. Generally speaking Widgren's calculations place the scale of illegal immigration (without unfounded asylum cases) at around 20-30 per cent of the legal inflow of foreign nationals. These figures were estimated as *gross* inflows, with no estimate how many irregular migrants leave the European Union in a year, but they are also sometimes used as net inflows. Presumably following the same methodology as outlined in Widgren (1994), an ICMPD report to the Swedish Parliamentary Immigrant and Refugee Commission (ICMPD, 1994: 19) provides a times series for the years 1985

to 1993. Widgren (1994) also provides an estimate on the number of smuggled persons, putting the share of illegal immigrants who have been smuggled at between 15% and 30% while estimating that the share of asylum seekers with unfounded applications (approx. 300 000) who have used the services of smugglers was somewhat higher at between 20% and 40%. In total, he thus arrives at an estimated number of 100 000 to 220 000 smuggled persons in 1993.[9]

A report by Salt and Hogarth (2000) gives detailed sources of information on *trafficking and migrant smuggling* in Europe. According to Clarke (2000: 18) the report provides one of the thorough state-of-the art references on evaluating irregular migration in Europe as a whole. Summing up the issue of estimates of numbers of trafficked and smuggled migrants the authors of the report one year later noted "…rarely is it clear how the estimates have been derived, though in general they rely on assumptions about the ratio between those apprehended at borders and those who succeed in getting through undetected" (Salt, Singleton and Hogarth; 2001:26).

Heckmann and Wunderlich (2000) estimated the number of persons who are trafficked and smuggled into the EU at the level of 400 000 (for the year 1999). The estimate was based on apprehension statistics. For every person caught entering the EU illegally, it was assumed that two pass unhindered. Similar techniques have been used to estimate illegal migration, smuggling and trafficking in Central and Eastern Europe in studies by ICMPD published in yearbooks on illegal migration in CEE since 1996 (ICMPD Yearbooks, various years).

The International Organisation of Migration collects both qualitative and quantitative information on trafficking based on the numbers of trafficked people it assists. However, due to, mainly, the victims reluctance and fear to report crime or testify against traffickers the numbers of traffickers' victims are definitely underestimated. IOM's trafficking database includes information on the number of victims assisted, their country of origin, age, travel route, and the manner in which they were trafficked (Laczko a.o., 2002).

A study by Juhasz (1999) pertaining to the Hungarian case, is worth mentioning here Juhasz suggests that estimates of the proportion of those apprehended in the total number of those engaged in the process of illegal migration vary not only among organisations but even within one organisation. In her research Juhasz observed that at the senior level of the Border Guard there is a strong optimism in methods and tools used to effectively protect frontiers from illegal mobility. However, the situation changes at the bottom and those actually patrolling the border judged their own effectiveness to be only ten percent. Juhasz observation on discrepancies of estimates was confirmed in Ukrainian studies (Klinchenko et al, 1999). For researchers producing estimates of illegal border crossings based on apprehension statistics and a multiplier derived from border authorities this implies a need to carefully consider biases.

Clarke (2000) also stresses the issue of multiple events that may occur in statistics in case of a single person who is arrested at the border, deported, tried again and was caught

[9] It should be noted that Widgren speaks of "trafficked persons" and thus not distinguishes between human smuggling and trafficking in today's sense.

again. The experience proves that quite a big proportion of illegal migrants is involved in repeated illegal border crossing.

Hilderink et al. (2003), published as a Eurostat Working Paper, review existing estimates. In order to estimate the number of illegal migrants in EU in 1997 the authors referred to the number of officially recorded resident foreign population – in 1997 it was 20 million (Hilderink H. at all, 2003: 47,48). However, the topic of estimation of illegal migrant population in EU was only roughly indicated in this paper. It was rather a kind of recommendation for further analysis.

In its estimates of the *stock* of those currently living outside their country of birth, the UN Population Division used its definition of immigrant by which it understand those living outside their mother countries for a minimum one year. According to Papademetriou (2005a) this estimates puts the immigrant stock at 190-200 million, resulting in 3.3 percent of the global population. A rough estimate about the share of unauthorized immigrants in the world's immigrant stock might put it at between 15 and 20 percent of the total (between 30 and 40 million immigrants). According to the same author, continental Europe's share should be little more than 20 percent, what makes the numbers of unauthorised at the level of around 7-8 million. Unfortunately the author did not give any source of the estimation of the total stock for continental Europe.

A commission staff working document (European Commission 2007) contains a review of methods for assessing numbers of illegally staying third-country nationals as well as estimations on their total numbers in the EU. In that paper one finds total numbers on the EU level and on the particular state level – gathered in a form of tables. The EU press release that is quoted in the introduction concludes from this document that the stock is estimated to be between 4.5 and 8 million, with a net inflow of 350 000 to 500 000 persons. However, the staff document quotes stock estimates from 2 to 8 million for Europe, indicating they include to a certain extent EU-citizens from the new members states. With regard to flows, it quotes quotes Global Migration Perspectives (2005) and the newspaper from Vienna for net inflows and aggregates (seemingly) gross inflows from their own tables of 21 countries to a total of 893 000 to 923 000.

The above mentioned efforts undertaken on the European level are tabulated below.

Table 1: Aggregate estimates for Europe

Year	Numbers	Source	Justification
1991	2.6 million non-nationals in Europe in an irregular or undocumented situation	International Labour Office	In this estimation the number of seasonal workers was included as well as those asylum seekers whose applications have been turned down but have not left. ILO assumption was that irregular foreign stock would be between 10 and 15 per cent of the size of the officially recorded resident foreign population (Clarke, 2000). It is unclear however, how these percentages were derived.
1993	inflows in Western Europe totalled around 250 000- 350 000	International Centre for Migration Policy Development (ICMPD); (Widgren, 1994)	The figure was calculated on the basis of extrapolation of the number of illegal migrants that reached their destination countries as a reflection of the known numbers of migrants apprehended at the borders. Widgren, on the basis of discussions with border control authorities, multiplied the number of 60 000 borders' apprehensions by 4-6.
1997	2- 3 million of irregular foreign population, 350 000 net inflow	Hilderink H. at all (2003: 47,48)	About 10 % -15% of foreign population (20 million)
1998	Upper limit of unauthorised migrants in Europe 3 million	Committee on Migration, Refugees and Demography "Conference on the situation of illegal migrants in Council of Europe member states, Paris 13 December 2001)	Illegal migrants are thought to represent between 10-15 % of migrants already present and between 20 and 30 % of inflow
1999	number of trafficked and smuggled into the EU at the level of 400 000	Heckmann *and Wunderlich* (2000)	The estimate was based on apprehension statistics. For every person caught entering the EU illegally, it was assumed that two pass unhindered.
2001	EU 15 (incomplete): 286 000	International Centre for Migration Policy Development (ICMPD)	Estimates on the basis of the total border apprehensions of illegal migrants
2005	7-8 million of un	Papademetriou	continental Europe's share of

	authorised immigrants in continental Europe's	(2005a)	unauthorized immigrant stock should be little more than 20 percent of the estimated world stock
2005	4.5 million in EU 25	Papademetriou (2005b)	Unauthorized immigrants are estimated to be at least 1% of the population of EU 25 and are growing at annual rates that are into mid-hundreds of thousands
2007	893 000 to 923 000 gross inflow	European Commission 2007	Aggregates gross inflow from information about 21 countries, quotes studies mentioned above for net inflows and stocks

Source: Own compilation

3. Conclusion

The review of efforts to estimate the size of irregular migration on a European level has shown that the numbers indicated are based on very rough estimates. Often, we do not know which groups of irregular migrants are in included in a stock estimate, nor we do not know whether a flow estimate is meant to measure net inflows or gross inflows (without substraction of outflows). Older studies are often quoted in newer studies, so that estimates appear to apply to the present, although they were made some years ago. In particular, the estimates that are currently used on the European level are based on estimates before 2004 that include substantial numbers of EU citizens.

With regards to methods, studies usually aggregate data from the national level that is not completely comparable, and without adjusting for potential overlaps. The most frequently used data are apprehension statistics. Some problems of them are already indicated, and more problems will be discussed in detail below in the methodological section. In combination with simple multipliers, apprehension statistics are used to indicate the size of inflows and the size of irregular populations.

The primary data of the Yearbooks originate from the contributions of border services and migration authorities from the countries of the region. The basis for the estimations was the proportion of those apprehended in the total number of those engaged in the migration business. However, one should remember that the existing data on apprehensions as Clarke (2000:20) notes "…often reflect the incidental, local or particular requirements of the agencies collecting the data." This is also why Bijak and Korys (2006) while analysing these estimations stated that although this kind of approach is interesting per se however it does not inform us about the scale of illegal flows in each country and about the stocks of clandestine populations as well.

In the UN study, the total size of irregular migration is estimated on a global level, based on general considerations and indicators, and Europe's share in the total is derived with a simple multiplier applied to the legal population – a very rough procedure as well, so that the often quoted 7-8 million for Europe is a plausibility estimate on very thin grounds.

On the other hand irregular migration grows along with other forms of population mobility. There are some evidences reported of enhanced efforts by traffickers and growing amount of money paid in order to be smuggled into Europe. Indeed, there is a strong need for better data on irregular migration in Europe.[10]

In the following chapter, a comprehensive review of methods indicates that there are more ways to come to estimates for undocumented populations. While they are not all applicable to the European situation, they suggest that there should be ways to improve from the particularly poor state of art on the European level that has been presented in this section.

[10] As Laczko (2002) notes in his critical article on data collection: at the global level, the most widely quoted figure concerns the number of women and children believed to be trafficked worldwide each year across international borders. In 1997, US authorities estimated that 700 000 to two million women and children were being trafficked in this way each year. No explanation has been given as to the assumptions underlying such estimates.

Chapter 3: Methods for estimating stocks and flows of irregular migrants

Michael Jandl

1. Introduction

Despite rapid technological advances and the postulated emergence of an Orwellian system of total and ubiquitous state controls, in the real world there are many social phenomena on which we have only little or at best imperfect information. Public policy areas where knowledge is inherently limited include homelessness, drug abuse, tax evasion, corruption, cigarette smuggling, drunk driving, under-age prostitution, stigmatizing diseases, mental illness, the "informal economy" as well as illegal migration. All these areas involve so-called "hidden populations" or "hidden activities" that are either difficult to observe or, once observed, are difficult to identify as belonging to that population or as performing such activities. For example, persons with certain characteristics (such as being illiterate, paranoid or having HIV/AIDS) or experiences (such as having become homeless or a victim of rape) may wish not to be identified out of fear or shame and often seek to hide it (cf. Chelimsky 1991, p. 685f). Moreover many, though by far not all of these hidden phenomena, involve illicit activities that make an objective counting and description of them even more difficult. However, for various reasons, knowing more about the size, characteristics and social behaviour of such "hidden populations" is of considerable interest to policy-makers as well as social scientists, who are attempting to tackle the challenge of researching such "hard-to-reach" populations.

Generating knowledge on the size and composition of irregular migrant populations is not much different from generating estimates on other "hidden populations" mentioned above and thus it is also plagued by many of the same methodical problems. A full count of the population of "undocumented" migrants is, by definition, impossible. It is also not possible to draw a representative sample from the total population, as the structure of the underlying total is unknown. In fact, without intrusive (police) methods it may even be quite difficult to identify whether a *particular* person is indeed an irregular migrant, even when observed and questioned. Like people involved in other illicit activities, irregular migrants have incentives to deliberately hide from public authorities.

Thus, the number of irregular migrants that are documented (in one way or another) at any one time is inevitably only a subset of the total population of irregular migrants while the "true number" of irregular migrants can never be known with certainty or a high degree of precision. However, public policy needs to be guided by evidence, often in the form of facts and figures. This leads to calls for estimating the "*dark figure*" of official statistics, a term that designates that part of the irregular migrant population that is not documented in the data but is likely to constitute the major part of it. As we will see in the following sections, there are various methodologies for estimating irregular migrant populations (stocks and flows) but it is important to note that, ultimately, all serious methods to estimate the *unknown* part of such populations must be based on some form of "hard data" on *known and reported* cases. This is also true for detecting and

interpreting trends in the characteristic and size of the hidden population. Thus, while the estimation of irregular migration ultimately aims at the description of undocumented and unobservable persons or events, any description of the nature and extent of the phenomenon has to rely on the availability of observed and registered indicators. From the observed statistical indicators conclusions on the dimensions of the unobserved phenomena (illegal entries, illegal residence, illegal employment, etc.) are then drawn, using a postulated or otherwise inferred relationship on the linkage between the two sets of variables.

As this postulated or inferred relationship between observed and unobserved variables is at best imprecise and at worst simply wrong, all estimates of irregular migration phenomena are prone to large margins of error, whether they are considered "scientific" or not. However, "scientific" methods do present a considerable advantage in that they specify their methods of calculation, the field covered, the hypotheses used and their statistical biases. Only if the method of estimation of a hidden phenomenon is transparent can we evaluate whether or not the resulting estimate can be considered more or less reliable. This chapter aims at giving a comprehensive overview over methods that are currently used, or could easily be used, to estimate the size of irregular migrant populations, discussing their key idea, necessary preconditions and limitations.

2. A classification of methods for the estimation of irregular migration

In analogy to data on legal migration, the fundamental distinction between all estimates on irregular migration is that they refer to one of two distinct statistical concepts: *stocks* (e.g. of undocumented/illegal residents or irregular migrant workers *at a point in time*) or *flows* (e.g. of illegal entrants or migrants "overstaying" over a certain *period*). While theoretically the two concepts are linked (see chapter 1), in practice the quantitative consistency of the two variables has proven to be elusive. Moreover, given the highly volatile nature of migration flows, the scarcity of reliable indicators on illegal migration flows, and the dearth of appropriate methods for estimating such flows, most efforts have concentrated on estimating stocks of undocumented migrants rather than flows. Accordingly, the following discussion focuses mostly on available methods for estimating the size of stocks of undocumented residents before taking up the matter of irregular migration flows and their inter-linkages.

Methods for the estimation of stocks of illegal residents can be divided into direct and indirect approaches.[11] Direct approaches are based on data that "capture" the subject of research (i.e. illegally resident foreigners) directly, while indirect approaches do not rely on such data. Data sets that contain (a subset of) the target population directly as

[11] It should be noted that the classification proposed here has some similarities but also significant differences to the typology proposed by Delaunay and Tapinos (1998, p.36ff) and taken up in the excellent overview provided by Pinkterton et al. (2004, p.33ff) due to differences in the criteria used for classification. For example, in the classification suggested here, the Delphi-method is not classified as a direct but an indirect method. Moreover, over the past decade, there have also been new methods developed and described in the literature that have not been available before.

(identified) illegal residents are immigration enforcement data (e.g. apprehended illegal residents), administrative records (e.g. data on regularization of unauthorized residents) and survey data (e.g. illegal residents identified through snowball sampling techniques). On the other hand, residual methods based on the difference between, for example, the total population represented in census figures and some estimate of the legally resident population (see 5.1.1 below) are classified not as direct but indirect approaches, because at no point do they identify illegal residents directly in their counts.

Direct estimation approaches can be further classified into multiplier methods, methods of self-identification and snowball sampling (referral by others) methods. For each of these categories further sub-categories (for example, among the multiplier methods, there are simple multiplier models, capture-recapture models, models using a comparison of administrative registers and random effect mixed modelling) will be identified and illustrated by an actual or hypothetical example. Among the indirect approaches, we can identify residual methods, demographic methods, subjective estimations/indicators methods, econometric methods on the size and structure of shadow economies, comparisons of immigration and emigration statistics, flow-stock methods and methods based on indirect inferences as the most import estimation techniques. There are also several combined approaches that use a combination of data sources and estimation techniques. Table 2a and Table 2b provide an overview of this classification scheme for stocks and flows.

Table 2a: Methods for Stock Estimates

Approach	Data sources	Method	Estimation technique	Main Idea of Calculation
Direct approaches	Based on immigration enforcement data	Multiplier methods	Simple Multiplier	Estimation of total with a simple multiplier based on derived or estimated ratio of "dark field" vs. "clear field"
			Capture-recapture/ Repeated capture	Estimation based on probabilistic function derived from multiple recaptures of individuals in sample
			Matching of registers	Estimation based on implied non-registrations in two or more individually matched registers
			Random effect mixed modelling approach	Estimation using statistical regression model assuming comparable apprehension rates of legal/ illegal residents with statistical adjustment for random effects
	Based on administrative statistics	Methods of self-identification	Evidence based on regularisation data	Inferences on the size and composition of irregular migrant stock prior to regularization from data on applications for and grants of regularizations
			Using data on status adjustments over time	Inferences derived from data on changes in residence status after a period of irregular residence
	Based on surveys	Survey methods	Direct survey methods	Reconstruction of a "random sample" of regular and irregular migrants through a re-weighting of the probability of contacts
			Snowball sampling methods	Estimation using chain referral methods to obtain a sample of persons not registered vs. persons registered
			Respondent driven sampling	Recruitment of interviewees through peers and incentive system leads to equilibrium sample of respondents after several recruitment waves which is independent from original sample and can be analysed statistically

Table 2a: Methods for Stock Estimates (continued)

Indirect approaches	Based on census/registers	Residual methods	Differences census results – legal immigration data	Indirect estimation based on the calculated difference between census data and data on legal immigrants
			Simple comparison of various registers	Indirect estimation based on a comparison of two or more registers with data on the same target population
		Demographic methods	Use of birth/death rates	Inferences on demographic subgroups based on the comparison of real and expected birth or death rates
	Based on census/registers/ demographic data	Expected population methods	Comparison of census/emigration data and immigration statistics	Indirect estimation of illegal resident population from comparison of emigration estimates with data on legal immigrants at destination
	Based on administrative statistics	Flow-stock methods	Calculating the stock through flow figures	Using estimated inflow- and duration of stay indicators to estimate steady-state stock of illegal residents
	Based on complementary data sources and estimates	Indirect inferences	Using information on correlated phenomena as basis of calculation	Making inferences on subgroups of irregular foreign residents on the basis of indirectly related phenomena and estimates such as irregular work, sector-specific demand for irregular services, school attendance or health services (e.g. inference of share and size of irregular foreign workers from econometric estimates on the shadow economy)
	Based on surveys of "key informants"	Subjective Estimation/ Indicators Methods	Expert surveys	Survey of key informants on their assessments of sizes, ratios and characteristics of target population
			Delphi surveys	Anonymous multiple round survey of key informants mediated by researcher to attain convergence of opinion

Table 2a: Methods for Stock Estimates (continued)

Combined approaches	Based on small scale surveys	Window/Postal code method	Small scale study + use of regression analysis	Extrapolation of estimates derived from intense local study with regression analysis
	Based on expert opinions	Localized Delphi	Delphi method + use of regression analysis	Extrapolation of estimates derived from localized Delphi study with regression analysis
	Adjustment to surveys/ census data	Non-threatening survey design	Direct survey method + randomized response/ 3-cards method + residual method	Statistical inferences from employer survey using randomized response method and inferences about share of irregular migrants in sample survey using non-threatening survey questions combined with residual estimation results

Table 2b: Methods for Flow Estimates

Approach	**Data sources**	**Method**	**Model**	
Direct approaches	Based on border apprehension data	Multiplier methods	Simple multiplier	Estimating total illegal border crossings by applying estimated or derived multipliers on border apprehension data
Indirect approaches	Based on stock estimates	Differential methods	Net differences in stocks	Deriving estimated annual net increase of irregular migrants through changes in estimated stocks
	Based on entry-exit statistics	Residual method	Double entry card system	Deriving estimated number of overstayers through individual matching of entry-exit records

Source: Own compilation

3. Direct Approaches

3.1. Multiplier Methods

Among the methods used for the estimation of stocks of undocumented migrants, a significant share is based on the "multiplier principle". This method starts from the proposition that the size of the unknown total can be directly calculated from the size of a known subtotal by use of an appropriately estimated multiplier (for example, that the stock of irregular migrants in a country at a given time can be derived by use of a multiplier on the number of detected irregular migrants). Once this proposition is accepted (and the size of the subtotal has been established with an acceptable degree of accuracy), the problem is redefined as finding the "right" multiplier (Vogel 2002).

The use of multipliers to derive the size of a hidden population from the size of a known subtotal of that population is probably the most common method of estimating the "dark field" of an unknown population in this as in other fields. For example, the United Nations Office on Drugs and Crime uses various multipliers to derive the size of problem drug users in a country: "If a survey among heroin addicts reveals, for instance, that one quarter of them was in treatment in the last year, the multiplication of the registered treatment population with a multiplier of four provides an estimate of the likely total number of problem heroin users in a country" (UNODC 2007, p. 266). An alternative estimate that can be compared to the first one is derived by applying a multiplier on arrest data: "...if a survey among heroin addicts reveals that one out of five addicts was arrested in the previous year, a multiplication of the persons arrested for heroin possession by the multiplier (five) provides another estimate for the number of heroin users" (ibid).

This description of a simple technique for producing a widely used estimation of a hidden population (drug users) contains several lessons for the proper use of multipliers in similar contexts: First, the underlying indicator (e.g. heroin users in treatment, apprehended illegal residents) has to be sufficiently reliable and correlated to the estimated variable (total heroin users, total illegal residents). Second, the definition of the indicator determines the definition of the estimated total. Third, the multiplier should be derived by an appropriate and reliable method (e.g. a broad-based survey). Fourth, no single estimate and multiplier is likely to be completely accurate and reliable, thus the combination (by using the mid-point, median or average) of several multipliers and estimates yields a better outcome than a single method alone.

Why did we dwell so long on the estimation of total heroin users, rather than irregular migrants? First, for illustration purposes and to gain from the experience of others working on hidden populations. Second, because it is surprisingly difficult to find genuine examples of the application of the multiplier principle when it comes to irregular migration. This has to do both with the difficulties of obtaining an appropriate underlying indicator and the problem of deriving a reliable (sample based) multiplier.

The sections below describe efforts to estimate the size of hidden populations with simple multipliers, but also sophisticated estimation techniques that take derived multipliers as starting point for statistical calculations.

3.1.1. Simple Multiplier Methods

For obtaining a relevant underlying indicator on illegal residence, most countries in Europe collect data on persons found to be illegally present on their territories (even if these statistics are often not publicly available). Definitions vary from country to country and some

aggregates are simply not suitable for estimating the number of illegal residents at a point in time. For example, Eurostat (the statistical office of the European Commission) maintains a regularly updated database (called CIREFI)[12] that regularly compiles statistics on illegal migration provided by EU Member States. There are three main types of data linked to illegal migration and published from the CIREFI database, namely statistics on "refusals of entry", on "removed aliens" and on "apprehensions of aliens illegally present". [13] There is no precise standard definition of any of these categories and EU Member States simply supply those statistics they already collect for their own administrative purposes using their own definitions. Thus, data on "apprehensions of aliens illegally present" in the CIREFI database generally do not distinguish between apprehensions of aliens illegally present *inside* the country and *at or near the borders* of the country that aliens have unsuccessfully tried to cross illegally; nor do they distinguish between persons and events (Jandl and Kraler 2006b). These data can therefore not be used for estimations of stocks of illegal residents or flows of illegal entrants. Whether available data on apprehensions of irregular migrants in individual EU Member States are suitable indicators for producing estimates needs to be decided on a case-by-case basis after careful scrutiny of the data.

A further problem related to the underlying indicator, data on "apprehended aliens illegally present in the country" within the course of one year do not, strictly speaking, refer to stocks of aliens illegally present at any one point in time. And second, more than other apprehension data, such data are dependant on the intensity of law enforcement efforts during the course of one year. To simplify, one could say that data on apprehensions within the country can easily be "produced": If higher numbers are desired, a few additional police raids on suspected work sites or living quarters of illegal foreign residents will lead to a higher number of apprehensions. This means that the (estimated) multiplier has to take account of both the time and the risk dimension of the indicator.[14] For example, *ceteris paribus*, a higher risk of apprehension that leads to a higher share of illegal residents apprehended should be reflected in a lower multiplier so that the estimated total number remains the same.

A good example of using a sample of the total population for estimating a derived multiplier is provided by Burgers (1996a and b cited in Pinkerton 2004, p. 14f) who uses a record and survey-based method of determining the number of illegal foreigners in Rotterdam at a particular moment in time. First, Burgers uses the number of apprehended foreigners over a six year period to gain a measure of how many *criminal* illegal foreigners were captured. Then he uses in-depths interviews with a (non-random) sample of illegal migrants (145 interviews) to determine the share of migrants involved in criminal activities. This proportion was then applied to the apprehension number derived before to produce an estimate of the total illegally resident population in Rotterdam. Finally, the share of the estimated population in the total population (1.8 %) was used to extend the estimate, first to the four largest cities in

---

12 CIREFI stands for "Centre for Information, Discussion and Exchange on the Crossing of Frontiers and Immigration" and was established as a confidential information-sharing mechanism between EU Member States during the 1990s. Eurostat has started to regularly publish data from this database only in 2004 and these data then usually date several years back. The are available as a subset of statistics in the Annual Reports on Asylum and Migration at:

http://europa.eu.int/comm/justice_home/doc_centre/asylum/statistical/doc_annual_report_2001_en.htm

13 The CIREFI database includes two more categories of data on enforcement measures relating to irregular migration, namely data on apprehended facilitators (by citizenship) and apprehended facilitated aliens (by citizenship and type of border), see: UN Statistical Commission, UN Economic Commission for Europe & Eurostat 2003

14 In the estimation of the multipliers for drug users mentioned above, this is done by establishing what share of a sample drug users was in treatment (or apprehended) over the course of the year.

the Netherlands (Amsterdam, Rotterdam, The Hague and Utrecht) and second to the whole country (In extending the estimate from the four largest cities to the whole country, a lower share of illegal residents in rural areas was assumed; this spatial concentration of irregular migrants in large cities and in specific rural area is an important aspect of the demographic composition of illegality in European countries (cf. Leerkes et al 2007).

A more recent example of applying a multiplier method to the estimation of the number of persons without legal residence is provided by a recent study on irregular migrants in Belgium (van Meeteren et. al. 2007). In this study, the authors explicitly apply the approach of Burgers (1996a and b) on the basis of police data on apprehended criminal foreigners combined with data from their own in-depth interviews with 120 irregular migrants. The authors use police data on criminal acts committed by foreigners without legal residence over a five year period and combine these data with the calculated "crime rate" among the 120 foreigners without legal residence interviewed for the study. In doing so, the authors adjust the periodicity of the data to the average length of stay of their respondents.[15] In addition to the application of the multiplier method on crime data, the authors also apply the method on administrative records on the health system. In Belgium, health clinics are compensated by the federal government for emergency medical care treatment dispensed to illegally resident foreigners. To claim these benefits, detailed records on emergency medical care are kept and forwarded to the government. The study now applies the ratio of foreigners without legal residence status who have used such emergency medical care services at least once during their stay in Belgium, derived from their sample of 120 respondents, to the administrative health records. Adjusted for the periodicity of the data and the likely double counting over the average period of stay, the resulting estimate is surprisingly close to the estimate based on the crime data but looks much less solid than the previous estimate.[16]

The authors of the study go to some length to defend the plausibility of their estimate and argue that the average of their two estimates is likely to represent a lower bound, or minimum estimate, of the likely range. In fact, the quality of the method depends on at least two basic assumptions. First, that the crime rate derived from the interview sample is a good indicator for the actual crime rate among the target population. An underreporting of criminality would result in an overestimation of the target population. Here, the authors argue that the results are fair because long in-depth interviews with sufficient trust established between interviewer and respondent did provide realistic answers. It is argued that this is even more realistic for the health usage rate. Second, the estimate is based on the assumption that the reported crime figures are a good indicator for total crime figures. This is most likely not the case, as crime is generally not detected in all cases by the police and the resulting estimate is therefore likely to be an underestimate. On the other hand, the authors argue that the undercount should be partly offset by the fact that the data are on number of suspects, not on "proven" crime. All in all,

[15] Over the five year period 2001-2005 foreigners without legal residence were registered in a total of 24,871 criminal cases committed by 13 628 persons. From the (non-random) interview sample a total of 10 persons out of 120 had admitted to implication in criminal acts in Belgium. The average duration of (illegal) residence in the Belgium of all respondents was 3.3 years. Thus, we get 13 628 divided by 5*3.3 = 8 966 and 10/120 = 0.083 for a multiplier. Combining the two we get 8 966*100/8.3 = 108 024 (van Meeteren et. al. 2007, p.280).

[16] Health data were available only for 2004, when 7 252 persons without legal residence status had received emergency medical care. Out of the sample of 120 respondents, 13 persons had received such emergency medical care at least in one year, 3 persons once, 4 persons two times or less and 5 persons three times or less. Adjusting for the average period of stay of 3.3 years, we thus get 7 252 + (3/13)*7252 + (4/13)*7252 + 0.33*(5/13)*7252 = 12 077 persons for a period of 3.3. years. The derived health care ratio from the interviews is 13/120 = 0.1083. Combining the two yields an estimate of 12 077*100/10.83 = 111 514 (van Meeteren et. al. 2007, p.282).

the authors state that theirs is a "conservative estimate" of the likely total (van Meeteren et. al. 2007, p14ff).

While there are few examples of similarly sophisticated applications of the multiplier principle, simple multiplier techniques are still widely used for estimating irregular migration phenomena, with the multiplier often resulting from "expert opinions" or other subjective methods. The strength of the multiplier principle is that it can be used simply and easily on a number of different indicators – for example on apprehended illegal residents, on illegal foreign workers or even on asylum or return statistics. The problems regarding definitions, data and multipliers, however, remain the same. Some examples can serve to illustrate this point.

For Slovakia, Divinský (2008, p.19) relates an earlier estimate of the stock of irregular migrants in Slovakia that was derived "as a proportional figure out of all estimated unlawfully employed persons in the country (both domestic and foreign) according to the structure of those being discovered by labour inspection". The proportional figure (the multiplier) is thus derived from labour inspection data and the underlying assumption is that natives and foreigners are randomly controlled and proportionally detected according to their actual share in irregular work. This is, of course a heroic assumption, as irregular foreign workers are known to be concentrated in certain branches that tend to receive more attention from labour inspectors (such as construction, agriculture and catering). Their likely overrepresentation in control data will thus lead to an overestimated multiplier. A second problem with this method is that it also needs to rely on an estimated total of "unlawfully employed persons in the country (both domestic and foreign)" and this may be no less difficult than the estimation of the (foreign) target group alone. A third problem is the equation of irregular foreign workers with illegal residents, an implicit assumption that is often made particularly in so-called "new immigration countries", where labour migration makes up the bulk of all immigration.

For the Czech Republic, Drbohlav and Lachmanová (2008, p. 21f), explain in more detail the method and underlying reasoning of several estimates put forward and widely quoted by Drbohlav in the late 1990s. For the estimate on the number of undocumented foreign workers, the author makes use of a rough-and-ready multiplier. Based on the author's ongoing survey work of migrant workers in the Czech Republic, he found that around 40% of the respondents were repeatedly illegally working migrants, which led to the conclusion that the ratio between legal and illegal workers might be around 1:1. Thus, the multiplier of one is applied to the number of registered legally economically active foreigners. Apart from noticing that this estimate was carried out in a time when the share of irregular migrant workers in the Czech Republic was likely to be relatively high (and thus not as rare and elusive as in most other contexts), we should also note that the derived multiplier is applied to a relatively well-known figure (the registered foreign workers) and not to another estimated figure as in the previous example.

An example where some of the shortcomings of labour inspection data for the application of a simple multiplier method are explicitly taken into account is provided by Cyrus (2008, p.41f). Using data on the total number of persons controlled by labour inspectors and, out of these, the fraction of persons who have been detected as working without the required work permit and residence permit, respectively, Cyrus calculates two multipliers – one for irregular work and one for illegal residence. Applying these multipliers to the total labour force, the author is careful to note that the derived figures can only be interpreted as "maximum estimates" for

undeclared employment in companies[17] under the assumption that irregular residents and irregular foreign workers are *at least as likely* to be found in the labour inspection data as in the labour force. Given the selectivity of controls and the fact that branches with a presumably higher involvement of irregular foreign workers (such as construction) are specifically targeted by labour inspectors, this assumption is plausible and can help to establish an upper boundary of an estimated range. However, without further assumptions or more information on the selectivity of controls, the data cannot tell us what the lower boundary or a central estimate would be.

Another example of the application of a simple multiplier provided by Cyrus (2008, p.38) refers to Berlin before the 2004 enlargement of the EU – thus a period when the number of irregularly working foreigners in Berlin was probably much higher than today, Cyrus assumed that the number of unauthorized immigrants from Poland was equivalent to the number of officially registered Polish immigrants (thus a multiplier of 1). At the same time, Cyrus explains that this ratio and the implied number of unauthorized immigrants from Poland is not to be seen as a stock figure, but rather as a pool of irregular migrants living in Berlin *and* Poland, as it included many persons who regularly commuted between Poland and Berlin for carrying out undeclared work. Thus, in this as in the other examples, the problems of the data, the derivation of the multiplier and the definitions remain in the foreground, even though the actual calculations of the estimates are straightforward.

3.1.2. Repeated capture methods

Capture-Recapture methods have a long tradition in population biology and were originally developed to estimate animal populations in the wild (Petersen 1896). Today, capture-recapture methods are widely used in epidemiology and in the estimation of hidden populations.

In their simplest form, capture-recapture methods are a sort of multiplier method where the multiplier is developed through repeated sampling of the same population. To illustrate, consider the following application of the principle to the estimation of the stock of fish in a pond. First, capture 1,000 fish, mark them, and let them free again. Then, capture another 1,000 fish and examine them. If 100 of them are marked (recaptured), you can deduce that the 1,000 marked fish statistically make up 10% of the total, so there are – presumably – 10,000 fish in the pond (cf. Jandl 2004).

There are few studies that apply this classic capture-recapture technique in a migration context.[18] However several studies apply the principles of this method in an advanced form to the estimation of the stock of illegal residents. The repeated capture method and the matching registers method rely on the capture-recapture idea.

Based on the above principle, a mathematically more demanding variant of the capture-recapture method has been applied for the first time with regard to illegal residents in the

[17] As labour inspections rarely inspect private households and public service, the estimate would reflect only undeclared employment in companies.

[18] However, the method has recently become used in the estimation of the number of victims of human trafficking (both internal and international). The widely cited ILO estimate of global human trafficking victims is based on a) an estimation of reported cases worldwide by way of a capture-recapture technique that seeks to establish the number of *reported* cases through comparing two samples of reported cases in order to estimate all reported cases not captured in the samples; and b) a further extrapolation of this estimated number by a factor of 10 to arrive at the ILO global minimum estimate of forced labour and, as a subgroup, the ILO global minimum estimate of human trafficking. See Belser, P. et al (2005).

Netherlands. Van der Leun, Engbersen and van der Heijden (1998) use the so-called "*repeated capture method*" for estimating the size of the illegally resident population in the four largest Dutch cities mentioned above.[19] The "*repeated capture method*" is based on one single data set (7 000 files related to all apprehensions of illegal immigrants in 1995 in the four largest Dutch cities) in which illegally resident foreigners may appear more than once. By analysing the files, it is determined who is captured once, twice, three times and so on.[20] Using the counts of persons captured and re-captured 1, 2, 3 and so on times, it is postulated that the number of appearances follows a probabilistic distribution, the Poisson distribution. On the basis of the available data, the crucial parameter determining the Poisson distribution (usually denoted by the Greek letter Lambda) can be estimated, which is then used to calculate the probability that an individual is *never* (zero times) caught by the police. Adding up this estimated number with the number of illegal residents actually "captured" various times in the police files produces an estimate of the total illegally resident population (cf. van der Heijden, Bustami et al 2003, van der Heijden, Cruijff et al 2003).[21]

The repeated capture method, as originally applied in the Netherlands depends on a number of crucial assumptions. First, the chance of getting caught over the period of study (one year in this case) remains constant and this probability does not change after a previous capture.[22] Second, the total population under consideration remains constant over the period (no inflows or outflows). Obviously, this is not the case at least for those apprehended illegal residents who are subsequently removed from the country, therefore a separate estimation is performed for the group of effectively expelled persons (van der Heijden, Bustami et al 2003). Finally, when applying a uniform probability of being caught for all persons, the underlying assumption is that the total population is homogeneous with respect to their risk of being caught. However, van der Heijden, P. (2006) demonstrates that this assumption can be relaxed when differences in the risk of being caught are included in the model by explicitly estimating different Poisson parameters for different groups of illegal residents.

To illustrate the model and the type of data needed to perform the calculation, Table 3 reproduces the police records used by the Dutch researchers for their calculations.

[19] The study was part of a wider study known as the "Unknown City Project". Their estimation results (a minimum number of 40 000 illegal residents in Amsterdam, Rotterdam, The Hague and Utrecht together) are surprisingly close to the estimate provided by Burgers (1996).

[20] An important precondition for applying the method is the ability to observe the same persons at least twice. As only few people will be captured twice and even fewer more than twice, the results of the estimation depend crucially on a correct identification of persons "recaptured" and a good matching of records, which is, however, made more reliable through the use of fingerprinting.

[21] For a short description of the method and results, see Pinkerton et al. (2004) and Sikkel et al. (2006).

[22] This also means that changes in policy or law enforcement efforts that could affect this probability are only insignificant over the period.

Table 3: Illegal immigrants data: observed frequencies for the three groups

Group	*f*1	*f*2	*f*3	*f*4	*f*5	*f*6	Total
Not effectively expelled	1645	183	37	13	1	1	1880
Effectively expelled	1999	33	2	1	1		2036
Other missing	430	41	5				476
Total	4074	257	44	14	2	1	4392

Source: van der Heijden, Bustami et al 2003, p.308

The sophistication of the method and the fact that it is based on data already available (police files) have made it the method of choice for the estimation of illegal residents in the Netherlands. Since the first study using the method was published in 1998, several follow-up studies have been carried out in the Netherlands. For example, Engbersen et al. (2002) apply the technique on police apprehension files from 25 Dutch police regions in the period 1997-2000. A specificity of this estimate is that they divided the data into those on illegal residents from third countries and a group of illegal residents from Eastern European accession countries.[23] In a recent update of this estimation, van der Heijden et al (2006) again use the repeated capture method for producing an estimate of the number of illegal residents in the Netherlands. In their estimation, they provide a breakdown into "European" and "non-European" illegal residents and additionally, in view of the then-imminent entry into the EU of Romania and Bulgaria, an estimate on the number of illegal Romanians and Bulgarians.[24]

3.1.3. Matching of Registers

The matching of registers technique uses the fact that illegal residents sometimes appear in a certain registry (e.g. police apprehension files, aliens registry) and can be identified and re-identified when appearing in another registry as well. If the probability of appearing in one register is independent of that appearing in the other one, the identification of the same individual in both registers constitutes, metaphorically speaking, a "recapture" of the same individual and the total can be derived from the sizes and capture rates of the two samples.

[23] The authors estimated that every year between 65 000 and 91 000 illegal residents from third countries but without people from Eastern Europe were in the Netherlands. In addition every year between 47 000 and 72 000 illegally resident Eastern Europeans were in the country, giving a total illegally resident population of 112 000–163 000.

[24] This method results in an estimate of 88 116 illegally resident non-Europeans, whereby the total is between 62 320 and 113,912 with a reliability of 95%. For European illegal residents, the estimate is 40 791 with a 95% reliability interval of between 12 000 and 70 000. The estimated number of illegally resident Bulgarians is 15 403 and the estimated number of illegally resident Rumanians is 6 782.

Table 4: Contingency table of two independent data sources on the same population

		ZAR	
		YES	NO
ZAS	YES	n_{11}	n_{12}
	NO	n_{21}	n_{22}

Now, if ZAR and ZAS are truly independent data sources, then n_{22} is given by

$n_{22} = (n_{12} \times n_{21}) / n_{11}$

and the total number of workers (including those working illegally) n is given by:

$n = n_{11} + n_{12} + n_{21} + n_{22}$

Sheldon (2002) illustrates the method with the use of a so-called contingency table. Suppose ZAS and ZAR are two independent data sources (registries) on the number of foreign workers (or any other variable of interest) in country X (ZAS could be a registry of work permits and ZAR a registry of persons with social security). The table lists the number of foreign workers in country X that can be found in both data sources (n_{11}), only in one but not the other of the two sources (n_{12} and n_{21}) and in neither data source (n_{22}).

Like the other capture-recapture methods, the simple application of this estimation technique is based on a number of crucial and problematic assumptions. First, the method assumes a closed population for the period the estimates relate to and no linking errors (i.e. a person can be correctly identified as appearing in both registers or only one register). Two further assumptions are that of homogeneity (all persons must have the same probability to be "captured" in ZAS and ZAR) and independency (the probability to be captured in one of the two registers is not influenced by the fact of being captured in the other registry). However both these latter assumptions can be corrected for with advanced statistical methods (cf. Sikkel et al (2006)). The assumption of homogeneity can be relaxed if heterogeneity in the probability of being captured can be modelled with a set of co-variates (nationality, socio-economic variables). Independency can be relaxed when there are more than two relevant registries available (in which case the two-dimensional matrix above will become a complex multi-dimensional contingency table). Van der Heijden (2006) explores the application of this method for the estimation of persons from the Dutch Antillean who are not registered in the Dutch population register and overstay their terms of residence in the Netherlands. However, for the method to be readily applicable in practice with three or more linked registers more statistical tests need to be carried out.

3.1.4. A Random Effects Mixed Modelling Approach

This approach tries to estimate the size of the illegally resident population from non-EU countries through statistical modelling from data on registered legal residents (N) and data on

apprehended legal (n) and illegal (m) residents to estimate the number of non-registered illegal residents (M). Thus, while the proportion of apprehended foreigners is known among the legal residents (i.e. n/N), it is unknown for the illegal residents (i.e. m/M is not known). The model realistically assumes that there are a number of reasons why these rates must differ and without knowing the difference one cannot estimate M directly. However, the model introduces a random effect that is able to accommodate the perceived heterogeneity in the relative apprehension rates for individuals and makes the estimation of M possible. To illustrate, let

$i = 1,\ldots t$ be the index of subpopulations by nationality, then

(1) $\log (M_i/N_i) = \Theta + v_i$

gives $M_i=N_i.e.^{\Theta}.e^{vi}$

That is, there is an overall proportional relationship between the two sets of stocks and a random effect (e^{vi}) to allow for heterogeneity.

Next, we have

(2) $\text{Log} (m_i/M_i) = \beta.\log(n_i/N_i) +e_i$

which gives $m_i= M_i(n_i/N_i)^{\beta}.\exp(e_i)$

That is, there is also an overall proportional relationship between the two sets of proportions (on the log scale) and a random effect (e_i) to allow for heterogeneity.

Without spelling out the full model (see Zhang 2008 for a description of the model and further extensions), combining the two equations yields a function with parameters that can be estimated on the basis of available data on 1) the numbers of foreign-born persons legally resident in Norway and registered in the Central Population Register at the reference date (N_i), 2) the numbers of (legally and illegally resident) foreign citizens who face criminal charges over a certain period (n_i) and 3) the number of apprehended unauthorized foreigners registered over a comparable period (m_i).

The basic features of the model are, first, that equation (1) assumes that the ratio of the illegally resident to the legally resident subpopulations are a *random* deviation from a global proportionality coefficient (which is by definition true) and, second, that equation (2) gives us the "catch-rate" mi/Mi of the illegally resident population as a function of the "catch-rate" of the legally resident population, including a random effect (which, given the latter, is by definition also true). The question is whether the model can be sensibly estimated and whether the nature of the data allow for the desired interpretation.

In applying the model to data on registered and apprehended foreigners in Norway, Zhang (2008) finds that the regression of the model to the data results in a good fit and thus allows for an estimation of the illegally resident population within reasonable confidence intervals. The model further allows a differentiation between the number of illegal residents who previously were (rejected) asylum-seekers (an important issue in the Scandinavian countries) and those who had never applied for asylum in Norway.[25]

[25] The expected total irregularly resident population with non-EU origins is estimated to be 18 196 by 1.1.2006. The estimated lower and upper bounds of a 95% confidence interval are 10 460 and 31 917, respectively. Of the estimated total irregular residents, 12 325 were previous asylum-seekers and the rest, 5 871, were persons who had never applied for asylum in Norway (Zhang 2008, p.ii).

3.2. Methods of Self-identification

Methods that rely on the self-identification of members of "hidden populations" are often mistaken for counts of the total. Regarding irregular migration, in practice the best known example of an "estimation method" that relies on self-identification of undocumented migrants is the count obtained from large-scale regularizations. These data are often interpreted as an approximate size of the illegally resident (or illegally working) population with the implicit assumption that all members of the hidden population would be able and willing to take advantage of an uncertain chance to obtain a regular status in return for revealing their identity and (work or residence) status to the authorities. As in most cases no explicit estimation of the underlying total is made, it is better to use the term "evidence" rather than "method".

3.2.1. Evidence based on Regularization Data

Large scale regularizations (or amnesties) are often carried out to deal with the perceived problems of dealing with a large number of irregular migrants or migrant workers. Almost as a by-product they also provide information on the illegally resident population (cf. Pinkerton 2004, p.39f).

When looking at regularization data it is important to clearly establish the nature of the data available, which is partly determined by the terms of the amnesty on offer. These terms can vary widely between countries and between successive amnesties in the same country over time. The only thing they have in common is that the persons regularized have previously been in some form of "irregular situation" pertaining to their residence or work status, or both. General amnesties are usually open to most persons in such an irregular situation while limited amnesties can be restricted to certain subgroups of the irregular population (e.g. former asylum seekers, family members of regular migrants, minors, etc.). However, even the terms of general amnesties usually involve some defined cut-off date specifying the required length of residence in the country, appropriate means of proof for this and other conditions (e.g. no criminal record).

Regularizations that pertain to the work status of irregular migrants usually involve a number of additional conditions such as proof of employment and a work contract. There are also important differences as to who can apply for the amnesty (the worker or the employer) and how the application can be filed.

Depending on the terms of the regularization on offer, the number of persons applying may be a good indicator of the total number of persons in the relevant category (illegal residents, irregular foreign workers, etc.) but there are also good reasons to expect significant deviations from the total number of persons in that category. First, not all eligible persons will apply for a regularization of their status. In most cases, the number of regularizations granted is significantly lower than the number of applications and not everyone will be willing to take the chance. Second, the number of applications is sometimes inflated when the same persons apply more than once to increase their chances. Thus the raw application data have to be cleaned of double-entries.[26] Third, some migrants who possess a legal residence permit that does not grant them the right to work may apply to obtain a different kind of residence permit that does grant this right (Cangiano 2008, p. 103). Fourth, when the regular status (residence or work) is offered only for a relatively short period of time (e.g. one year), once regularized

[26] Pinkerton et al (2004 p. 40) cite the example of the French amnesty 1997, where the number of applications was thus reduced from 180 000 applications to 144 000 applicants.

migrants may eventually fall back into irregularity (thus showing up repeatedly in successive amnesty programs.[27]

Finally, instead of pertaining only to a defined group of irregular foreigners in the country the regularization is carried out, an unknown number of (irregular) foreigners from other (usually neighbouring) countries may also file an application.[28]

A recent comprehensive table on European regularization programmes is presented in Pastore (2008) building on the work of De Bruycker (2000), Levinson (2005), OECD (2004) and (2007). Where available, it details both the number of applications and the number of regularizations granted. While in some Southern European countries, regularization programs for undocumented migrants are carried out in more or less frequent intervals, in other countries regularizations are implemented on a more restricted and discretionary basis. All in all, some four million *regularizations* have been granted over the period under review. However, for the reasons mentioned, it is unclear to how many *persons* this figure relates to.

3.2.2. Using Data on Status Adjustments

The regularization of the residence status of foreigners who were previously in an irregular situation in many countries takes place not in widely announced special (mass) amnesty programs but in more discrete programs of ongoing regularizations on a case by case basis. While governments are usually reluctant to publish detailed statistics on such programs, making them one of the most under-researched migration topics of our times, when such statistics become available they can also shed some light on previous illegal residence of those regularized. For example, in a comparative paper on France and the UK Thierry and Rogers (2004) take the number of first contacts 1999 – 2002 with the prefectures in France in order to obtain a regularization after a period of illegal residence as a minimum estimate of new illegal residences since the special regularizations in 1997 and 1998. Most individuals regularized had been in France for many years.[29]

Another method of self-identification that has, however, not been applied in practice so far, was proposed by Jandl (2007). The proposed method would use the fact that at certain moments in time the applicable immigration and residence requirements for certain groups of persons may change. In particular, residence may become legal for a group of persons where it was illegal previously (The situation is thus similar to that of a mass amnesty). Using data on the number of persons who had previously resided in the country undocumented but who use this opportunity of "legalization" to register themselves, an estimation of all illegal residents in the country at a certain moment in time can be performed.

---

[27] Depending on the terms of the regularization it is likely that the extent of this phenomenon varies. For example, in the case of Italy relatively few are thought to fall back into irregularity while in the case of Greece the majority of beneficiaries of recent regularization programs are thought to have fallen back into an irregular status (cf. OECD (2004), p.70). For Italy, the migrant surveys carried out by ISMU (see also Section 3.3.1) have shown that in 2006 only around 2% of the interviewees were individuals who had fallen back into irregularity after having been regular (Sciortino 2008 as cited in Fasani 2008, p.47).

[28] For example, in Italy it has been observed in past regularizations that a number of irregular migrants normally residing in France tried to obtain a regular status by applying for the regularization in Italy moving to Italy just to file their application. Sciortino, G. (2003): Regularization of Foreigners in Italy, unpublished manuscript, Communication to Michael Jandl on 3.4.2003, p.2

[29] This number of new first-time contacts with the authorities was 13 000 in 1999, 14 000 in 2000, 18 000 in 2001 and 29 000 in 2002. 71 % of those regularized in 1999 – 2001 had arrived before 1998. For the same period, and for all third-country-nationals combined, an average of 11% of the adult foreigners admitted for residence in France had previously been in an illegal situation. See: Thierry and Rogers (2004), p.654ff

In recent European migration history, two notorious moments come to mind that may be used for this type of estimation. The first is May 2004, when the full freedom of movement for the new EU citizens from 10 new EU Member States went into effect and the second is January 2007 when full freedom of movement for the new EU citizens of Bulgaria and Romania came into effect in the 25 other EU countries. For persons from these countries, who were already in the "old" EU countries prior to those dates without the authorization to stay, one could say that the new regime constituted a kind of *de-facto* regularization. It was further observed that the number of "new" EU citizens who registered themselves after those dates grew strongly, giving rise to the assumption that part of the newly registered had resided in the country already before the lifting of residence restrictions. For example, for the United Kingdom it was estimated that from some 232 000 applicants to the Worker Registration Scheme from the new EU member states between 1 May 2004 and 30 June 2005, up to 30% or 70 000 may have already been in the UK before May 2004 in an irregular situation (Home Office 2005, p.1 and 8). The estimate of the *share* of persons who had been in the country prior to EU enlargement without residence rights is thus crucial to the method to hold and it is not clear how reliable the cited share for the UK really is. However, a good estimate could be easily obtained by drawing a sample of newly registered persons and establishing their length of residence by confidential interviews.

The resulting (minimum) estimate on the number of new EU citizens who had illegally resided in the country prior to EU enlargement can then be used for extrapolation to other nationalities. For example, assume that the estimate was performed on the number of illegally resident nationals from X and Y in 2006. These numbers can be extrapolated using a very basic form of capture-recapture inference: Drawing from different but complementary samples[30] of "captured" X and Y citizens in 2006, one can estimate the share of X and Y citizens among all unauthorized residents during that period and calculate the implied total number of illegal residents.

3.3. Survey Methods

Survey methods for the study of hidden populations are dominated by two types of techniques: Location sampling techniques are suitable when the target population is geographically concentrated (overall or at certain times), while chain-referral sampling techniques are suitable when members of the target population know each another and are interconnected (Heckathorn 2002).

3.3.1. Direct Survey Methods

The use of survey instruments among immigrant populations to gain insights about those immigrants who are in the country on an irregular basis usually presupposes a significant share of the irregular component among the total immigrant population. Moreover, for any meaningful sampling certain conditions have to be fulfilled (such as minimum sample size and the confidence that respondents do not systematically misrepresent their status) that make the technique extremely time- and resource intensive. For example, a regular survey of both regular and irregular immigrants in the Italian city of Milan has been carried out by the Foundation ISMU since the early 1990s. In 2001 the survey was extended to the Italian region of Lombardy and the monitoring activity consolidated in a new Regional Observatory for Integration and Multiethnicity. The annual survey encompasses interviews with 8 000

[30] These samples could consist of police statistics of apprehensions, labour statistics of foreign workers without permits, health statistics of undocumented foreigners in clinics, etc. - all by nationality.

migrants (since 2006: 9 000 migrants) and provides a detailed picture of the life and work conditions of both regular and irregular migrants in the Lombardia region. In 2005 the survey technique was extended to all of Italy with a sample of 30 000 migrant interviews in 40 Italian provinces to measure the effects of the 2002 regularization. [31]

The survey technique applied by ISMU is the so-called *centre sampling technique*, which aims to arrive at a random sample of the universe of regular and irregular migrants residing in the sampling region. The basic rationale and assumption of the technique is that each foreigner in the sampled area at the time of the survey by necessity keeps a set of contacts with one or more centres or gathering places located in that area (migrant institutions, places of worship, health care, entertainment, meeting places, etc.). Once a sufficiently wide set of "centres" is identified any foreigner could be described by his list of contacts with various centres (one or more centres frequented). This information is then derived from the interviews with migrants contacted at various centres. Questioning the interviewees for all the meeting places/centres frequented by them, a matrix of all such centres visited by each interviewee can be drawn up. As the total number and characteristics of (regular and irregular) migrants is not known in advance, the sample cannot randomly draw migrants. However, once all the centres in the sampling area are known the centres can be selected randomly (with replacement) and then within each centre interviewees can be chosen among the persons frequenting it. To make the sample representative two statistical adjustments are carried out. First, the number of interviews in a certain centre will depend on its size, with smaller centres given a smaller probability to be chosen than larger ones. Second, after each interview the responding migrant is assigned a profile of all the centres she or he frequents. His or her probability of inclusion is higher, the higher the number of centres frequented and the lower the number of people frequenting each centre. Thus, each individual receives a weighting coefficient that corrects for this probability so that the weighted sample has the same representativeness as a random sample and the share, characteristics and size of the irregular resident population can be calculated (Baio, G., Blangiardo, G.C. and Blangiardo, M. 2008).

3.3.2. Snowball Sampling Methods

Identifying (potential) irregular migrants through a first set of contacts that in turn lead to further contacts is another way of building a sample of respondents. However, the selection process is clearly non-random and consequently snowball sampling as a means to estimate the size of a hidden population is fraught with many difficulties and potential sample biases. Natale (1998) reports on various examples where the method has been tested in Italy and distinguished between different types of snowball sampling. The method proposed by himself starts with the extraction of an initial sample from a list of registered foreigners and asks each interviewee to indicate another foreigner (who is either in the initial sample or not) to obtain a second sample and so forth. Taking all the interviewed persons *not* in the initial sample together one can obtain an estimator of the probability of not belonging to the initial list and from there an estimate of the non-registered foreigner population. Natale admits that the method has several defects and introduces two variants of the method. The first is the so-called *habitation snowball method*, which takes designated (fixed) houses for immigrants as sampling places for foreigners, collecting also information on co-habitants not present. The second method proposed is the so-called hospitality centre or *meeting place method*, which is similar to the centre sample technique described in the previous section (see Natale 1998, p.10).

[31] For more information, as well as current statistics and annual reports, see: www.ismu.org/ORIM/

The *snowball method* described above is a non-random sampling method that relies on a number of initial respondents for locating additional waves of respondents of a hard-to-reach population (such as drug users or illegal residents). This method is conventionally used for identifying a larger number of respondents (of hidden populations) for qualitative interviews.[32] There are also attempts to apply the method to quantify the size of a hidden population through statistical inferences from a single wave of respondents. Zhang (2007) provides an overview of the data requirements and the calculations for producing a statistical estimate of the so-called *single-stage link-tracing sampling*, the simplest form of quantitative snowball sampling. In this model, initial respondents are asked to nominate other members of the target population (one round only). Nominees do not have to be known by name but instead the requirement is only to identify whether they are in the sample of initial respondents or not. The higher the (relative) number of persons nominated that link back to the initial sample of respondents, the smaller the predicted total of the target population. A formula can be used to calculate the estimated total.

While the data requirements may not seem to be very high, in reality they are likely to work only for relatively small target populations. Zhang (2007) provides the example of the study of Frank and Snijders (1994) who used the method to estimate the number of cocaine users in Rotterdam, a relatively small hidden population.[33] Given these constraints, the method has so far not been used for sizing relatively large irregular migrant populations. However, the method may be fruitfully applied in combination with other estimation techniques, e.g. methods suitable for extrapolating from smaller to larger areas (see Section 4.5 on combined approaches).

[32] It may often also be the only method available for identifying a sizeable sample of a hidden population for the purpose of estimating a multiplier, as in the study of Burgers (1996 cited in Pinkerton 2004, p. 14f).

[33] From an initial sample of 34 persons, the number of nominations was 311 of which 15 pointed back to the initial sample. This yielded an estimated total of 685 persons in the target population.

3.3.3. Direct surveys using network characteristics

Respondent-driven sampling (RDS) is a different form of chain referral technique that explicitly tries to reduce the bias introduced through the initial (non-random) sample of subjects from which the sampling begins. Moreover, RDS differs from snowball sampling in that subjects are not asked to identify their peers to the research but to "recruit" them into the study. To do so, respondents are offered two types of incentives: a (financial) award for participating in the interview plus a reward for recruiting others into the study (Heckathorn 1997). Those recruited may decide for themselves whether to participate or not thus reducing possible "masking" of other peers. Analytically, the RDS technique produces a stochastic process in which each recruiter's social characteristics (e.g. being black or white) affect the characteristics of her or his recruits with a clear tendency towards in-group recruitment. The probabilities associated with these recruitment patterns (e.g. black recruits black, black recruits white, etc.) can be described as a Markov Process. Under certain assumptions, such processes quickly lead to an equilibrium state, that is to say, the characteristics of the sample of recruits after a certain number of recruitment waves approach the "true" equilibrium state independently from the original sample. When applying the technique for obtaining representative samples of injection drug users for HIV/AIDS prevention programmes it was shown that a rapid convergence typically occurs within five to six recruitment waves (Heckathorn 2002).

While the RDS technique introduces some ingenious innovations in sampling hidden populations it relies, like any chain referral sampling procedure, on the crucial assumption that sampling subjects have traceable contact patterns. Thus the method is not suitable for drawing national or even regional samples across large geographic distances (Heckathorn 1997). However, as for the previous technique mentioned, it may be a valid basis for an intense small-scale study.

4. Indirect Approaches

There are great differences between the methods subsumed under "indirect methods". What they have in common is that at no point do they use counts of irregular migrants (such as apprehensions etc.) directly for their estimations.

4.1. Residual Estimation Methods

4.1.1. Comparison of Census Results with Legal Immigration Figures

This method is based on the proposition that a total count of the immigrant population in a country would capture both legally and illegally resident immigrants at the same time, while not allowing a distinction between them. Subtracting the total number of legally resident immigrants from this total count then gives the estimated number of illegally resident immigrants.

This method is most commonly applied in the United States, where the Census (carried out every ten years) is thought to give a good count of the total foreign born population, including the illegally resident population (Pinkerton 2004, p.20f). Undercounting in the census in this model is estimated and the data are adjusted accordingly (cf. Hoefer et al 2007). The major difficulty in the US model is to calculate the number of legal immigrants (legally resident foreign born) from the sum of the constituent immigration components as there is no current

population registry available. The basic formula for computing the residual foreign-born population R (which is equal to the illegally resident foreign-born population) is thus:

(1) FB = (L – (M+E)) + T + R

Where FB is the total foreign-born population at the time of the census, L are legal immigrants until the time of the census, M is the estimated number of death of legal immigrants in the country, E is emigration of legal immigrants and T is the number of temporary legal migrants.

From (1) it follows that

(2) R = FB – (L – (M+E)) – T

An application of the method on the 2000 Census data produced an estimate of some 8.5 million undocumented immigrants lived in the United States in 2000 (Passel, 2002).[34] This figure represents an increase of about 5 million over the estimates for 1990, leading to the conclusion that the annual (net) increase during the 1990s averaged about 500,000. The latest estimate gives a total number of unauthorized immigrants of 11.6 million in January 2006 (Hoefer et al 2007), implying growing illegal (net) migration over the past 5 years (see Section 5.2).

The use of "residual" estimation methods based on the differences between census data and other registries of immigrants is not common in European states, mostly because the censuses are thought to be seriously undercounting irregular migrants in Europe (cf. Lederer 2004 p.197). The two major exceptions are the UK and Spain where the residual method has recently been applied for the estimation of the size of the illegally resident population.

In the UK, the Home Office commissioned a report that provides an estimate of the foreign-born population living there illegally based on a residual method that is summarized in Woodbridge (2005). The methodology involves the same calculations as in the US case above but adds another category of migrants who are in the UK on a "quasi-legal basis" (Q), for example asylum-seekers or refused applicants with leave to remain. Thus, the total foreign-born population can be calculated as

(1) $FB = (L – (M_L+E_L)) + (T – (M_T+E_T)) + (Q – (M_Q+E_Q)) + R$

with M_L, E_L … denoting group-specific mortality and emigration rates.

The total foreign-born population is obtained from the census of 2001, thus the estimate of the residual R pertains to April 2001. As no explicit undercount estimates are available for the foreign-born population FB, a range of undercounts (lower-middle-upper) is applied that results in a range of estimates of R.[35] An additional complication and uncertainty is given by the fact that in the UK no registered stock figure for legal permanent immigrants L is available – the stock is thus obtained by adding the number of migrants granted settlement each year since 1970 adjusted for likely emigration and death.[36]

[34] The estimate combines data from Census 2000, the Immigration and Naturalization Service, the March 2000 Current Population Survey, the Census 2000 Supplementary Survey, and previous estimates.

[35] The central estimate of the unauthorized resident population in the UK in April 2001 was 430 000 while the lower and upper range estimates were 310 000 and 570 000, respectively.

[36] The number of legal immigrants granted settlement in the UK before 1970 was estimated with an alternative method.

4.1.2. Simple Comparison of Registers

Another exception to the general assumption that illegally resident foreigners are not well covered in the census or other official registers is provided by Spain. Here the law allows foreigners – whatever their legal status – to be included in the municipal register (*Padrón*) if they can provide evidence of residence in that municipality (Aparicio and Ruiz 2008). Irregular migrants, particularly since 2001, have been encouraged to register to obtain health benefits and because they can use this as proof of residence for later regularization. On the other hand there are hardly any disincentives connected to registering as these data are not utilized for removing unauthorized residents from the country. Consequently, the difference between the figures in the municipal population register and those of the register of foreigners with residence permits has been used as an indicator for the number of irregular foreigners in the country.[37]

Despite the apparent simplicity of comparing the two registers as an estimation technique, there are also several problems connected with using this method. First, the municipal population register itself has many flaws, particularly with regard to the number of registered persons who have already left the country but did not de-register (as commonly observed in population registers) – which would result in an overestimation.[38] A related problem is posed by persons who have never lived in Spain but have been registered in the *Padrón* by their relatives – an overcoverage which would result in an underestimation when comparing registers. Second, the register of residence permits is also imperfect, that is the number of resident permits counted is not necessarily the same as the number of legal foreigners in the country (due to emigration, death, naturalization, etc.). And third, the simple comparison of registers does not allow the estimation of foreigners in the country who are neither recorded in one nor the other register.[39] The last problem is known to be particularly relevant for EU-citizens residing in Spain without any official registration. Recognizing these difficulties, González Enriquez (2008, p.27ff) provides a more careful estimation of the number of non-EU illegal residents in Spain for early 2008 based on a comparison of the *Padrón* with the resident permits database that also takes account of the number of foreign students (with temporary stay permits) and asylum seekers.

In addition to the residual methods discussed above that try to make use of the census or (population) registers where *all* foreign residents should be represented, there are also attempts to apply residual methods on registers that contain only certain subgroups of the target group. For example, Biffl (2001) put forward an estimate of the number of illegally resident school children (6 to 15 years) that uses a residual estimation technique by comparing the stock data of foreign students in grade 1 to 10 derived from school statistics with demographic data on the foreign resident population by age groups derived from the central population register. However, the estimate has been criticized on the ground that the high margins of error of in the population data make any residual estimate unreliable (see Kraler et al, 2008, p. 23ff). Another example of using school attendance data for the estimation of a

---

[37] Between 2000 and 2004 the difference between the two registers has rapidly increased to more than a million foreigners. This compares with about 690 000 foreigners in an irregular situation who had applied for regularization in the regularization campaign of 2005 (see Section 4.3.1 above).

[38] Since a reform of the Law on Aliens in 2003, third country citizens without permanent residence permits must renew their individual registrations in the *Padrón* every two years. The first process for clearing the register in this way started in December 2005 and resulted in the removal of about 300,000 third country citizens, who did not renew their registration, from the *Padrón* (González Enriquez 2008, p.26).

[39] A specific technique to estimate persons not registered in any of the available registers would require an individual matching of records (see Section 3.1.4)

subgroup of illegal residents is demonstrated by Maroukis (2008, p.48ff) for the case of Greece.

In all the above examples of applying residual estimation techniques in a comparison of existing registers, the estimations are strongly affected by either over- or under coverage of the target groups in the registers. The simple comparison of registers is thus generating rather rough estimates which could nonetheless be improved, and which are useful for triangulation with other methods.

4.2. Demographic Methods

So-called "demographic methods" are based on the idea that over the course of one's life each individual is exposed to certain "risks" of experiencing a demographic event like giving birth, dying or being hospitalised. Statistically these events are well documented and age, sex and group specific birth rates, death rates and hospitalisation rates are widely available for comparison. In principle, illegal residents are exposed to many of the same demographic "risks" as legal residents, which should result in traces of these events in demographic statistics (births, deaths, accidents). Hence, a comparison of (an apparent surplus of) recorded demographic events in civil registries with those normally expected for a given age, sex and group cohort could theoretically be used for deducing an estimate of the illegally resident population (cf. Wanner 2002, p.11ff).

There are some examples of the application of demographic methods in practice. Delaunay and Tapinos (1998, p.65) present an Italian study (Natale 1990) that analyzes the growth in birth rates of foreigners and compares it with the growth of legal foreign residents between 1984 and 1988. The assumption is that the growth in birth rates would equal the growth in the foreign population and that the excess growth in the former would be due to the presence of illegal residents. The reference year for the birth rate of legal foreigners in Italy was taken to be 1971 when the presence of illegal foreign residents was assumed to be negligible. However, as Delaunay and Tapinos (1998, p.65) note, the assumption that the birth rate stays constant for such a long time period is questionable.

Another example cited in Lederer (2004, p.195) and Delaunay and Tapinos (1998, p.88) is an older study by Robinson (1980) on age and death specific death rates in 10 US States. The study relates the "surplus" in the recorded (age and sex specific) death rates of the registered legal immigrants (compared to the native population) to a statistical error that results from an undercount of the statistical basis for the death rates – namely the presence of a significant number of undocumented foreigners which are not factored into the denominator of the death rates. Delaunay and Tapinos (ibid) question the reliability of the death records for illegal residents as well as the validity of equating death rates of migrants and natives, while Lederer (ibid) critically notes the fairly wide range given by the estimate.[40]

The advantage of demographic methods (where applicable) lies in the ready availability of the required data without the need for generating new data sets. The basic assumption of these methods is that certain demographic events, like births or deaths by accident are hard to conceal from the authorities and are usually registered comprehensively (cf. Vogel 1999, p. 172).[41] However, the application of the method in practice is hindered by several obstacles.

[40] The stock of illegal foreign residents in the 10 US states was estimated at between 570 000 and 4.7 million persons.

[41] The Netherlands provide one example, where in principle data on births given by illegal residents could be available from midwives. In the Netherlands, pregnant women tend to go to midwives and if women do not have health insurance, the midwives would likely know their residence status as well. According to Dutch law,

One difficulty is that "demographic risks" for illegally resident foreigners are likely to differ from those of natives and legally resident foreigners as migration is highly (self-) selective. For example, births tend to occur in families while irregular migrants often move and live alone. Another problem is that migrants are (almost by definition) more mobile and often choose to return in case of impending demographic "risks" (giving birth, chronic illness, death), particularly in case of illegal residents. If this is frequently the case than the registered birth, death or hospitalization rates will not indicate the actual statistical base of these demographic events. In fact, given the detailed demographic statistics required for applying the method in practice (such as death rates by age, sex and nationality), the number of registered events may not be large enough in most of the smaller European countries for reliable statistical calculations. Still, a further exploration of certain demographic statistics (such as fatal road accidents) that suffer from less bias through self-selectivity, return migration, non-registration etc. may produce some useful results for the estimation of stocks of illegal residents (cf. also Lederer (2004, p. 240f) and Wanner (2002, p. 14)).

4.3. Comparison of Immigration and Emigration Statistics

In an overview of some non-European studies on the size of illegal migrant populations Lederer (2004, p. 192ff) describes several studies that attempted to estimate the number of undocumented migrants in the US from Mexico based on demographic data from Mexico. One of the methods applied used the Mexican census results (1960 and 1970) to calculate the difference between the actual and the expected population living in Mexico as an approximation of actual emigration. Assuming that most emigration is to the US and subtracting legal emigration, the authors arrive at an estimate of undocumented Mexicans in the US. Another method used observed and expected sex-ratios in Mexico to estimate the number of undocumented migrants in the US ("missing males").

More recently, attempts were made to apply the method in the European context. Delaunay and Tapinos (1998, pp. 42-56) detail their own calculations on emigration data from Morocco and Tunisia to estimate the number of emigrants from these countries living illegally in the main European destination countries. The authors use both data from Moroccan censuses in 1982 and 1994 and Tunisian censuses in 1984 and 1994 and also test the sex-ratio method on the Moroccan case. However, the authors admit that their results for Morocco[42] suffer from serious data deficiencies in the census data available and that the estimation for Tunisia was not possible due to a lack of reliable data.

These observations are in line with general observations on census and emigration data in European countries of origin: emigration data are undercounting and census results are often unreliable. Moreover, the method is complicated by the fact that – unlike the Mexican-US case – migrants originate from a large number of countries and disperse across a larger number of receiving states in Europe. However, despite these general caveats, emigration data could still be a valuable complementary source of information on irregular migrants in European countries. For example, a study on emigration from Ukraine presents estimates on the number of Ukrainian migrants working abroad (often illegally), based on information

---

illegally resident pregnant women are eligible for pre- and post-natal health care where costs are covered by a special fund. However, according to NGOs and midwives many pregnant women do not know about their rights, which would result in an undercoverage in such statistics (cf. van der Leun and Ilies 2008, p.32).

[42] The clandestine population from Morocco in Europe was estimated by the expected population method to be around 370 000 and by the sex-ratio method to be between 600 000 and 700 000 persons.

received from embassies and consulates in the reception countries (Malynovska 2004).[43] Comparing this information with the number of legal immigrants from Ukraine registered in the most important destination countries provides an alternative indicator for estimating the number of illegal residents from certain countries of origin and destination. However, neither the quality of the underlying data nor the method for estimating the number of nationals over and above those registered at their embassies and consulates abroad are generally available, making it impossible to evaluate the quality of such indicators.

Another possible indicator is obtained from surveys in emigration countries. An example is provided by the estimated number of Moldovans in the Czech Republic as related by Drbohlav and Lachmanová (2008, p. 42f). The authors base their estimate on (2003-2005) survey data from Moldova that indicate that around 1% of the Moldovan workforce works on a long-time basis in the Czech Republic. Comparing the implied numbers of Moldovans in the Czech Republic with the number of legally registered Moldovans in the Czech Republic in 2007, a residual estimation of the number of illegally residing Moldovans is obtained. However, the results are affected both by the differences in time periods between the survey and the register data and the difficult definition of "long-time basis", when many Moldovans are expected to work also on a short-time basis abroad. This last observation points to a particular methodical problem in any estimation based on a comparison of survey and register data, namely the difficulty in measuring the periodicity of any particular migration episode. Thus, migration may be short-term or circular, in particular when short-distance migration is involved. This type of migration pattern often remains unrecorded in most official migration statistics, even when any single (short-time) period of residence remains well within legal limits (e.g. up to 3 months on short term visas or during visa-free stay). Similarly, it is likely that such complex back-and-forth movements between countries are not well captured in general emigration surveys. Thus, it is possible that the number of emigrants captured by such surveys captures the total number of persons involved in (short-term and circular as well as long-term) migration projects, a substantial part of whom would, in fact, not concurrently reside abroad. Thus, the stock estimates derived from emigration surveys may be overestimating (irregular) emigration to an unknown extent.

On the other hand, survey data carried out in countries of emigration can provide additional information on socio-economic activities of (regular and irregular) emigrants. For example, Iglicka and Gmaj (2008, p.18) cite data from a survey conducted by the Ukrainian National Institute for International Security in 2003. According to this information, most emigrants working illegally in Europe are concentrated in a few economic sectors: 38.4% work in construction, 15.1% in trade and 14% work as housekeepers or workers in private households.

4.4. Flow-Stock Methods

In principle, starting from a given base, the stock of illegal residents in a country is determined by the number of migrants entering the status of illegality (through illegal entry or overstaying) and those leaving the status of illegality (e.g. through leaving the country or by regularisation). Tapinos (1999, p. 232) observes that: "The stock of clandestine immigrants results from the difference between their inflow and outflow and depends therefore on the average length of time spent in the state of clandestinity". Thus, in principle the stock of

[43] "According to the information from Ukrainian embassies, Ukrainian labour migration has the following structure in terms of countries of destination: in Poland there are 300 000 labour migrants, in Italy and the Czech Republic, 200 000 (each), in Portugal, 115 000, in Spain, 100 000, in Turkey, 35 000, in the USA, 20 000. The number of Ukrainians who work in the Russian Federation is estimated to be 1 million people." (Malynovska 2004, p.14)

illegal residence at a given point in time could be estimated if we know the size of inflows and outflows and the average time spent in an illegal situation in the country.

A more formal model for this estimation is provided by Sheldon (2002, p.22ff). From physics, we get the formula for calculating the (steady-state) equilibrium of a (replenishing) stock as

(1) Stock = Flow per Period x average Duration of Stay within the system

This relationship between stocks and flows is valid only in equilibrium, that is that situation where new entries into the system equal exactly the number of exits. For example, if 1,000 rejected asylum seekers per month "disappear" into illegality and remain hidden for an average of 6 month until their apprehension, it follows that the stock of disappeared asylum seekers is 6,000.[44] However, as Sheldon (2002, p.23f) further explains, the average duration of stay to be used for this calculation should, strictly speaking, measure the average *completed* duration of stay in the system (the average time between entering and leaving the system). The average duration of stay of persons *within* the system at any point in time (measured for example through surveys or apprehensions) is not enough, as it could differ from the completed duration (it could be both longer or shorter).

The difficulties of obtaining reliable data on both new entries into the system and on their average duration of stay within the system have so far prevented a widespread application of the method. However, the method could usefully be applied for certain categories of entrants into illegality where good base-line data are available. For example, Wanner (2002, p.14f) suggests the application of the method to "disappeared asylum seekers", a group of persons that often ends up in an illegal residence status but is well documented up until the moment of their disappearance.

4.5. Methods based on Indirect Inferences

In addition to the many methods surveyed in this paper, a large number of estimation techniques exist that are more or less suitable for estimating stocks of illegal residents in European countries. Such methods often stem from the availability of particular sources of data for specific sectors or points in time and may not be easily transferable to other countries.

A widely used technique for estimating the number of irregular foreign workers involved in certain sectors of the economy starts from calculating the likely labour demand in a certain sector or branch and then making indirect inferences on the involvement of irregular (foreign) workers. A good example is provided by a detailed study on seasonal labour demand in German agriculture by Hess (2006 as cited in Cyrus 2008, p. 44). Based on data of floorspace, yields and output, the total volume of working hours for a sample of 126 special cultures was calculated. Next, an average working time of 50 hours per person and week and an average contract duration of 10 weeks for seasonal workers was assumed. This resulted in an estimate of the overall number of seasonal workers required in German agriculture for each year. Comparing these estimates with the numbers of officially placed seasonal workers in 1994 and 2001, Hess (2006) concluded that the number of irregularly employed foreign workers in agriculture had significantly declined, mainly due to a large increase in official quota places for seasonal agricultural workers.

Other estimates relying on demand side considerations are based on surveys on the use of certain services. For example, Iglicka and Gmaj (2008, p.20ff) develop their own estimate of

[44] The example is taken from Sheldon (2002, p.22) who has modified a suggestion by Wanner (2002, p.14) for applying the method to estimating the size of illegal residence in Switzerland resulting from rejections of asylum applications.

the number of irregular foreign workers in domestic services (cleaning, nursing, gardening, etc.) on the basis of a household survey that indicated the number of (native and foreign) domestic service workers employed by Polish households. Using additional information from research on the type and frequency of services procured by 101 households that employed foreign domestic service workers, they first calculate the number of irregular foreign workers, who work full-time for one household only. In addition, by using certain assumptions on the number of households served by other foreign domestic workers working in more than one household, as demonstrated in Kraler et al. (2008, p. 16f) for the Austrian case, the authors arrive at an indirect estimate of the total number of irregular foreign domestic workers in Poland between 2001 and 2003.

Evidence on the presence of illegal residents can also be based on estimates of the size of the informal sector of a given country and the participation of (irregular) migrants in it. In this method estimates of the size of the irregular foreign population are based on indirect estimates of the participation of migrants in the informal sector of a given country (i.e. the amount of illegal foreign employment). Estimates on the size of the informal economy can in turn be based on the amount of cash in circulation or the amount of electricity consumed.

In his estimations on the sizes of the shadow economies[45], economist Friedrich Schneider uses econometric models that attempt to estimate unobserved variables (GDP, output) with the help of observed variables (e.g. cash in circulation). The so-called currency-demand approach is based on the idea that services in the shadow economy are usually paid in cash and that the size of such transactions can be estimated with properly specified currency-demand equations, while the so-called DYMIMIC (dynamic multiple-indicators multiple-causes) model links the unobserved variable (the size of the shadow economy) with a set of indicators through a number of structural equations (cf. Schneider und Enste 1999 and 2000).

Schneider (2002) estimates that the (absolute and relative) size of the shadow economy in Austria has grown considerably over the past 30 years from less than 2% of official GDP in 1970 to almost 11% of official GDP in 2002. However, later estimates indicate a stagnation between 10 – 11% after the year 2000 (Schneider 2005) and a slight decrease to some 10% by the year 2006 (Schneider 2006). Based on the estimations of the shadow economy output, then, estimates on the extent of „black labour“ are provided: For 2006, these are estimated to involve 716 000 native Austrians (full-time equivalents, a highly fictional number of a much larger number of Austrians involved part-time) and 98 000 illegal foreign workers (full-time equivalents, ibid p.18).

It should be noted that the econometric estimation methods (and results) of Schneider are not uncontested. For example, in a comprehensive study on the economic integration of foreigners in Austria, Gudrun Biffl (2001), the leading labour market economists in Austria presents her own methodology and estimation results on foreigners in the informal economy. According to her estimates there are only about 47 000 foreigners working full-time in the informal sector (equivalent to an actual number of foreigners engaged in (full and part time) illegal foreign employment of 50 000 – 70 000; Biffl 2001, p. 360ff).[46]

Summarizing, it can be said that the indirect estimation of the size of the illegal foreign population through economic indicators is a useful addition to the repertoire of estimation

[45] The shadow economy is defined as „all unregistered economic activities that contribute to the officially calculated GDP“, see Schneider and Klinglmaier (2004, p. 4).

[46] Biffl (2001) bases her approximation on the estimation that the black economy contributes about 8.7% to total value creation in Austria, that the actual labour force deployed to the black economy corresponds to 10% of the total labour force in Austria and that the share of foreigners in the informal sector does not exceed their ten percent share in overall employment.

techniques available but that it is hardly the panacea. Moreover, given the wide range of estimates resulting from the same methodological corner, it is save to repeat a previous evaluation of such methods: "There are … no indications that the resulting estimations are prone to smaller margins of error than those derived from other estimation techniques and the resulting estimates on illegal foreign residents display even greater variations than estimates derived through other methods."[47]

Besides such indirect inferences from data and estimates on irregular economic activities, some indirect methods try to make use of yet other data sources. Lederer (2004, p.194) discusses a study on remittances from the US to Mexico and ponders the use of this method for Germany. However, as the author of this paper experienced himself over the past years, it is simply impossible to gain access to meaningful data from the main supplier of money-transfer services in Europe (Western Union), despite the fact that these data exist and are used for elaborate and sophisticated business planning within the firm itself. Other indirect "methods" that are conceivable or have been suggested in this context may simply not be suitable for use in the estimation of illegal residents.

A potentially more promising source of data on irregular migrants could be data on health, sickness, accidents or medical attention. For example, the French government bases its estimate on the illegal resident population on the number of people who have received medical aid from the state (Coste 2008).[48] However, in basing the estimate on the stock of illegal residents in France on the raw data of persons who have received medical aid, no serious attempt is made to evaluate and adjust the existing database for its shortcomings. First, no adjustment is made for the number of persons who do not fulfil the eligibility criteria for benefiting from medical aid.[49] And second, no explicit adjustment is made for the share of illegal residents not claiming any medical aid during their residence in France (Courau 2008).

4.6. Subjective Estimation/ Indicators Methods

All estimations methods that rely on the collection and aggregation of subjective opinions and estimations of "experts" will necessarily result in subjective indicators, no matter how large the sample or sophisticated the information processing. In the following we will distinguish between (one round) expert surveys and (multi round) Delphi surveys.

4.6.1. Expert Surveys

In this context "experts" (also called "insiders" or "key informants") are individuals who, through their direct engagement with the research subject, are assumed to avail of privileged knowledge of the target population (e.g. on its size and structure, at least in their immediate surroundings). The validity of such methods is thus largely dependent on the validity of this

---

47 Jandl (2004), see ibid for further examples of economy based estimation methods in the Netherlands. See also Pinkerton et al (2004, p.29) on Belgium and ibid (p.31) on Italy.

48 According to a Law of 25 July 1999 the Medical Aid of State -*Aide Médicale d'Etat*-, provides medical care to illegally resident persons (among others). The government noted a sharp rise in those who benefited from this medical assistance between 2001 and 2003, from 139 000 to 170 000. Using such numbers, the Prime Minister Office estimates the illegally resident population to lie between 200 000 and 400 000 people. (Coste 2008).

49 There are at least three eligibility criteria that are likely to result in the exclusion of a significant number of illegal residents from the system: being able to produce an identity card from a list of acceptable documents; having an income of less than € 606 per month; and showing proof of residence in France for at least three months (for example through a rental contract or an electricity bill).

assumption, namely that the "expert" providing his subjective estimation does have at least an approximate grasp of the absolute or relative size of the population he is commenting upon, for example through direct interaction with that population. Given the limited scope of contacts any individual person can have, it is clear that the accuracy of any subjective estimate is higher the smaller the (geographical or otherwise) scope of her or his estimate extends to. A local NGO working with irregular migrants in distress, for example, may have a reasonable idea of the scope of its target population at the local level but can hardly be expected to provide a national estimate. Similarly, "insiders" from the target population itself or employers of irregular migrants are likely to have a better grasp of the environment they interact with often than unconcerned "outsiders".[50]

A single subjective estimation does not constitute a reliable estimation *method*. For arriving at more sophisticated estimates with a wider scope (extending to a sector, an economy or a nation, for example) both a suitable *sample* of experts as well as an appropriate *aggregation* method are needed. The sample should (at least potentially) cover experts with insights into the whole target population, while the aggregation method should (at least theoretically) lead to a representative estimate of the expert sample. Both components should be illustrated with the following methodical examples.

Lederer (2004, p. 237ff) relates the method and results of a survey on irregular migrants in Germany carried out by Caritas Germany in 2005. The survey involved the dissemination of a written questionnaire to 1,000 key informants from NGOs, law firms, churches, etc. working with (irregular) migrants. While the survey produced valuable qualitative insights into the needs and service requirements of irregular migrants in Germany, no conclusive information on the size of the population could be collected.[51] Cyrus (2008, p. 24f) discusses this and other surveys asking for irregular migrants among the clients of established charity services in Germany. The main problems he identifies in interpreting such data are that the surveys are restricted to charities specialised on immigrant counselling, which are by nature highly selective and non-representative (e.g. focused on certain nationalities, certain gender-related issues, etc.). In addition, the organisations face the same problems in identifying non-legally resident persons among their clients as everyone else (which results in undercounting), while it cannot be excluded that other clients are repeatedly making use of such services from several organisations (which would result in double-counting). The data are thus not suitable for estimating the size of the target groups concerned.

In contrast, Piguet and Losa (2001) have deliberately carried out survey of key informants in Switzerland in order to estimate the extent of illegally employed foreigners. The authors made an anonymised survey of 821 Swiss employers in particular branches of business, asking them about the estimated percentage of illegally employed foreigners in their own branch of business. Using the survey results, they then calculate the estimated number of illegally employed foreigners in Switzerland. It is noteworthy, however, that the results depend to a

[50] Again, it is instructive to look at methods developed to research other hidden phenomena than irregular migration. Research on corruption, for example, distinguishes in its surveys between the simple "perception of corruption" and the "experience of corruption" (cf. the work of Transparency International at http://www.transparency.org). However, the psychology of perception itself is to some extent shaped by experience: It is a common observation that somebody driving a blue car will notice more blue cars on the street than those driving cars of other colours (many thank's to Peter van der Heijden for pointing this out to me). It is this selective perception that gives insiders (such as irregular employers in surveys of irregular employment) the qualification of "experts".

[51] The suvey produced 310 answers (31% return rate) that reported on a total of 5 700 contacts with illegally resident foreigners. No attempt to arrive at an estimated total is reported by the author.

large degree on the choice of statistical aggregation.[52] Another employer survey in the four most notorious sectors illegally employing foreign workers was carried out in the Netherlands (Visser and Van Zevenbergen as reported in van der Leun and Ilies 2008, p. 17f). The authors came into contact with 2 235 firms (out of a sample base of 6 000) employers, among which the response rate was 36%. Based on a non-response analysis, the authors concluded that the reliability of the results was not threatened by (selective) non-response. Using information from the Labour Inspectorate on the average number of foreign workers illegally employed by non-compliant employers, the authors estimated the total number of foreigners illegally employed in each of the four investigated sectors and the Netherlands as a whole in 1999.[53]

Other expert surveys carried out in the Netherlands (Zuidam and Grijpstra 2004 and Dijkema et al. 2006 as reported in van der Leun and Ilies 2008, p. 18f) attempt to estimate the number of illegally employed migrants hired through intermediary employment agencies. Based on the opinions from experts in employment agencies, government and social partners the authors estimate the number of *malafide* employment agencies and the number of foreign workers illegally hired through them to arrive at a final estimate of the stock of illegally employed migrants.[54] However, the authors of the studies do not sufficiently explain the methods used to extract and aggregate the information obtained from their informants (ibid).

Pinkerton et al (2004, p.16f) report on an expert survey on the agricultural sector in Greece carried out by Lianos et al (1996). The authors selected four regions in northern Greece, where they selected key informants for their survey of illegal foreign workers in agriculture. The results for the regions were extrapolated to all regions and economic sectors in Greece using a different methodology to yield a total that is called "highly questionable" by Pinkerton et al (ibid, p.17).

Another important example that falls into the group of "expert" surveys is the estimation carried out in the 1990s on the city of Leipzig by Alt (1999 as related by Cyrus 2008, p.32). For this small-scale estimate, Alt made use of publicly available statistical data and carried out expert interviews with key informants as well as with irregular migrants. Based on a broad range of estimates indicated by the former group of experts, Alt finally settles on an estimate towards the higher end, which was supported by the statements of interviewed irregular migrants who indicated that they themselves knew several hundred irregular migrants in Leipzig.[55] Applying the same approach to the estimation of the size of the irregular migrant population in Munich (Alt 2003), Alt put forward the suggestion that the share of the illegally resident population may range between 1% and 3% in large German cities, which much influenced all subsequent estimates on the total number of illegally resident foreigners in Germany. However, such a generalization stands on weak theoretical and empirical grounds. The problem with extending the small-scale estimates from Leipzig and Munich to a general

[52] Applying the (adjusted) average of the survey results, the calculated number of illegally employed foreigners was 182 556. However, applying the median of the survey results, the calculated number is only 73 100. The authors concluded cautiously that the likely range was 70 000 – 180 000. Sheldon (2002, p.17) observes that the average is increased by the high estimates provided by relatively few key informants while the majority of respondents cluster their answers around the lower median, which would make the average value less reliable.

[53] The authors estimated that there have been potentially 128 000 illegally employed foreign workers in the Netherlands in 1999.

[54] The authors concluded that in 2003 there were an estimated 80 000 illegally employed migrants in the Netherlands, while in 2006 there were between 54 000 and 97 000 illegally employed migrants in the Netherlands.

[55] From an indicated range of between 4 000 to 10 000 irregular migrants, Alt finally settled at the estimated level of 8 000 irregular migrants in Leipzig.

share of irregular residents in all German cities is precisely that it does not take relevant differences between cities into account (e.g. settled migrant populations, the structure of the economy, income levels, housing markets, geographical location, etc.).

4.6.2. Delphi Surveys

The Delphi method involves a large number of independent experts in an interactive process of exchange through the use of written questionnaires, designed to foster convergence and consensus. [56] Experts are given the opportunity to state their views on a given topic and react to the (anonymized and consolidated) views and assessments of other experts in the second round. In order to allow an exchange, revision or refinement of opinions expressed by experts, the survey must include valid feedback mechanisms. Therefore the Delphi method includes several consecutive stages (written questionnaires) but at a minimum two rounds of questionnaires. An example that has already been mentioned in Section 4.2.2 above is provided by a Swiss study on *sans papiers* which uses the Delphi method in combination with regression analysis to estimate stocks of illegal residents in Switzerland (see also 5.5.2 below). Here small-scale estimation results are gained from an interactive expert survey with 10 key informants in six cantons in different parts of the country (for a total of 60, see GFS 2005).

In a recent application of the method in a two round Delphi-Survey in Austria (Bilger et al 2006), participants were asked to provide estimations on the extent of irregular migrant work in the most notorious economic branches in increasing detail.[57] A parallel Delphi-Survey on migrants' irregular economic activities was carried out in Hungary by Juhasz (Futo 2008, p. 31f) and in the Czech Republic by Drbohlav and his research team (Drbohlav and Lachmanová 2008, p.20). While the results produced some valuable insights in the likely (relative) importance of irregular migrant work (a broader concept than illegal foreign employment and much wider than work of illegal residents) in certain branches[58], the quantitative results must still be seen as subjective and of low reliability. Consequently, in the English summary of the Delphi methodology (Jandl et al 2007, p.12) the authors conclude: "The value of a Delphi-Survey lies not primarily in the (quantitative) representativeness of its findings but in the qualitative input of a diverse collection of experts that helped to elucidate a broad view of the research subject." This view is shared by many others who question the widespread application of the Delphi method (or one round expert surveys) for quantitative estimation. Lederer (2004, pp. 207f), for example, sees the value of the Delphi method in the context of sizing irregular migrant populations mainly in the validation of various estimates and as a valuable addition to a system of multiple indicators on irregular migration.

4.7. Combined Approaches

This chapter has described and illustrated a number of distinctive approaches to the estimation of stocks of irregular migrants. Besides such methodologically "pure" methods it is fairly

[56] General information on the Delphi-Method is provided in Linstone and Turoff (1975). A good overview is provided in Masser and Foley (1987). For a recent application of the method in migration research, see: Lachmanová and Drbohlav 2004.

[57] The questionnaire was answered by 37 experts in the first round and 22 experts in the second round.

[58] On average, the expert panel estimated the extent of irregular migrant work as a percentage of total employment to be highest in construction and catering/tourism (around 15%) as well as in agriculture (13%). The average estimate for trade and industry (5.2%) was only slightly higher than that for the share of irregular migrant work in total employment in Austria (5%).

common that a combination of estimation techniques, or combined approaches, for sizing stocks of irregular migrants are used in practice. Below a number of possible combinations that have recently been used, are currently explored, or could easily be used in future, are described.

A number of such combined approaches use regression analysis as part of their estimation technique. Regression analysis can be used in the estimation of hidden populations in various settings. Strictly speaking, the above presented random effects mixed modelling approach (Section 4.2.1) uses regression analysis to estimate various proportions (including a random effect) of the apprehended illegally resident population (by subgroups) within the total illegally resident population. More commonly, however, regression analysis is used in combination with other methods to extrapolate a "given" quantity from a smaller sample (e.g. a smaller area) to a larger sample (e.g. the whole country), using the calculated correlation of the "given" quantity to a set of known variables for extrapolation. In this context, the application of the method assumes that the size of the "given" quantity (e.g. any "hidden population" or the illegally resident population) in a smaller area can be derived by the use of other methods and that the results can then be further extrapolated to a larger area (e.g. the whole country).

As an example, a study on sans papiers in Switzerland (GFS 2005) uses the Delphi method (see also 5.5.2 below) to estimate the number of illegally resident foreigners in six cantons in different parts of the country. The results are analysed for correlation with known variables – in this case, the variables "share of work force in agriculture", "share of registered foreigners", "income level" are found to be positively and plausibly correlated with the "given quantity" determined before (the higher any of them, the higher the share of illegal residents in the canton).[59] Taking these calculated effects, the results for the six cantons are then extrapolated to the whole country to yield a central estimate of 90 000 sans papier in Switzerland (ca. 2005).

A further application of regression analysis from a small-scale estimate to a large-scale one is currently tested in the Czech Republic. In this project, the so-called "windows method" (see 5.5.1 below) is used to determine the (presumable) size of the illegally resident population in several areas of Prague. The results are subsequently extrapolated to a larger area with the use of regression analysis (Drbohlav and Lachmanová 2008, see also 4.7.1. below).

Of course, any results obtained through the extrapolation of a "predetermined" quantity from a smaller to a larger area, depend crucially on the quality of the small-scale estimate. However, the use of regression analysis for the extrapolation should normally result in a better national estimate than the simple extrapolation by one variable only (e.g. by multiplying with the population size).

4.7.1. Combination of Window Method and Regression Analysis

The window method, also called *postal code area method* tries to establish the size of a hidden population in a well defined and restricted area (such as a small postal code area in a larger country) and to use the resulting figures to extrapolate the result to a larger area (e.g. the country as a whole). The advantage of using well specified areas is that often detailed statistics on specific socio-economic characteristics within this area are available that allow for the possibility to use the postal code characteristics for efficient estimation at the national

[59] The variable "degree of urbanization" is found to be positively correlated in a simple regression model, while it is found to be negatively correlated in a multi-variate regression model. It is therefore excluded from the further calculation.

level. If these characteristics (e.g. purchasing power) are statistically linked to the size of the hidden population present, then the use of these characteristics in extrapolating micro-level results to macro-level results (by multiple regression analysis) should improve the estimation (cf. Sikkel et al 2006, p.153f). This "reduces" the problem of estimating a national total of a hidden population (e.g. homeless, drug addicts, prostitutes, illegal residents) to one of estimating the size of a hidden population on the micro-level. Possible methods for establishing an approximation of the size of the hidden population at the micro-level include the Delphi method (see Section 5.5.2), the 3 card method (see Section 5.5.3) or a "full count" by going from door to door and establishing the residence status of the inhabitants. The latter application of the window method is currently tested in an ongoing research project in the Czech Republic (Drbohlav and Lachmanová 2008, p.25f). In the project, a survey is to be implemented (via observation and interview method) that tries to establish the actual number of inhabitants, and of foreigners specifically (regardless of their legal status), in specific localities in Prague. The observed numbers will then be compared with official numbers for the specific localities to derive the number of illegal residents as a "residual" for each locality. Using socio-economic variables, the micro-level results will then be generalised to other areas of Prague. The feasibility of the method is currently tested and at the time of writing results were not yet available.

4.7.2. Combination of Delphi Surveys and Regression Analysis

A recent study on *sans papiers* in Switzerland (GFS 2005) uses a combination of Delphi surveys and regression analysis to extrapolate the result to the whole country. As the two methods used have already been described in Sections 5.3.2 and in the introduction to this chapter, it should only been mentioned that the combination of the subjective Delphi method with the statistical regression method results in an estimation that must still be considered as subjective, even if the extrapolation is based on statistical analysis. Moreover, the subjective estimates of the responding experts on a relatively small area (their own cantons) are likely to be better founded than corresponding subjective estimates on a larger (e.g. national) scale (see also Section 5.3.1).

4.7.3. Randomized Response and Residual Method

Randomized response is a method designed to make surveys of sensitive topics (like drug use, sexuality or illegal residence) where respondents may feel uncomfortable to answer sensitive questions. Therefore, the survey is designed in a way that answers to sensitive questions are partly determined by chance but can statistically be used for calculating aggregate results. One particular variant of the method is the so-called "forced response" method. The respondent is asked to throw two dice and to answer a question on the basis of the count of the dice. On counts 2,3 and 4 is must answer "yes", on 11 and 12 he must answer "no" while in all other cases he is asked to answer truthfully. While on an individual basis nobody knows whether an obligatory answer was given, in the aggregate the statistical distribution of the true answers can be deduced (cf. Sikkel et al 2006, p.153f).

A randomized response method for surveying employers of irregularly employed migrants has been applied in the Netherlands by Regioplan Policy Research for 2004 and 2006 as reported by van der Leun and Ilies (2008, p. 19f). The research team concluded that 19% of the employers they had surveyed had used illegal workers for one or more days during 2004. To arrive at an estimate of illegally working foreigners in the Netherlands, they use this figure and further assume that each of these company had hired roughly 3 irregular workers and furthermore that each irregular worker in turn had held 3 to 4 such irregular work relations

with Dutch employers.[60] Using the same randomized response method, a 2007 Regioplan report estimated that 9% of the total number of Dutch employers had hired irregular workers in 2004 while in 2006 it amounted to 11%. The resulting estimates of the total number of irregular foreign workers in the Netherlands was comparable to the previous estimates (van der Leun and Ilies 2008, p. 20).

A new alternative to randomized response surveys is the so-called 3 cards method, developed by the US Government Accountability Office (GAO) to improve the estimation of the illegal alien population in the United States (see Larson and Droitcour 2003). It asks the respondent to choose one of several mutually exclusive answer categories on their immigration status (including the sensitive undocumented status) from an answer card. There are three slightly different answer cards, each of which arranges the various answer categories together in three boxes. The sensitive answer category is never alone in one of three boxes but appears always in combination with others (thus, respondents never answer the sensitive question directly). By dividing all respondents into three random samples, information on the sensitive answer category can be estimated indirectly by statistical inference. While the 3 card method apparently needs further testing before being administered on a large scale, it seems to have the potential for improving existing estimation techniques if applied in combination with other methods (such as the residual method in the US).

5. Methods for the Estimation of Irregular Migration Flows

As this paper demonstrates, a considerable body of work has been devoted to the development of appropriate methods for the estimation of the size of illegal resident stocks. However, much less methodological progress was made in developing better methods for estimating the flow of illegal immigrants. This is particularly true for Europe, where only rudimentary data and methods exist to analyse illegal border crossings, visa overstayers and gross and net illegal migration.

5.1. Estimating Illegal Border Crossings with Multiplier Methods

The only methodology explicitly used for estimating flows of illegal migrants to Europe is the extrapolation of border apprehensions with simple multiplier, using an estimated ratio of border apprehensions to (non-detected) illegal entries. This is done, for example, by Heckmann and Wunderlich (2000) and Widgren (2002).[61] Using a more complete set of border apprehension data, the estimation procedure was repeated by Jandl (2003) to produce a revised figure.[62]

---

[60] Based on the results of the survey the authors estimated that 19% or around 89 650 companies had violated Durch Aliens Employment Law in 2004. Thus, they arrive at a range of 89 000 x 3 persons / 3 to 4 jobs = 66 750 to 89 000 irregular workers in the Netherlands in 2004.

[61] Heckmann and Wunderlich (2000) used an assumed ratio of 1:2 border apprehensions to illegal entries to the EU (on the basis of some 260 000 border apprehensions), and estimated annual illegal (gross) immigration of over 400 000. A more elaborate description of his own methodology that includes the same 1:2 ratio of border apprehensions to illegal entries is outlined in a communication by Widgren (2002), who applied the ratio of asylum seekers who arrived irregularly in European countries as a rough multiplier and estimated some 500 000 illegal entries to the EU for 2001.

[62] For 2001, the total volume of illegal migration flows to Europe was estimated at 650 000 for the EU-15 and at 800 000 for the (later) EU-25. See: Jandl, M., (2003).

While the estimates based on this method have assumed much weight (not least due to the absence of alternative methods), several problems should be pointed out here. The first one is the quality and availability of apprehension data in Europe. No valid generalisations about this issue can be made, as much of the statistical material collected by border enforcement agencies is kept secret and only few countries in Europe regularly publish their border apprehension data.[63]

The second methodological problem associated with the use of border apprehension data for the estimation of illegal migration flows is the fact that border apprehension data refer to *cases* rather than to *individuals*. Thus, if an individual is apprehended more than once while trying to cross an international border illegally, he or she will be counted more than once in apprehension statistics – the estimation thus refers to illegal border *crossings* rather than illegal border *crossers*. However, there is some evidence that this conceptual problem could be less significant than often assumed.[64]

The third methodological problem that should be taken up here is the estimation of the "correct" multiplier. Conventionally, the most commonly used multiplier for the estimation of actual illegal crossings on the basis of apprehension figures has been two ("two pass for each one caught").[65] However, it is easy to see how one can derive various multipliers for different countries and different nationalities. To illustrate, consider the following calculations for Germany by nationalities. In 2001, for example, there were 17 000 asylum-seekers from Iraq, of which 3 000 had contact with the police before lodging their asylum application. This means that about 14 000 Iraqi asylum seekers had managed to come to Germany undetected and stay undetected for at least a while before lodging an asylum application. Thus, in the case of the Iraqis in Germany in 2001, the multiplier could be as high as five, while for other nationalities, the "correct" multiplier is likely to be much lower. In other cases, applying a multiplier of two or higher would almost certainly be misleading due to the high mobility of a large part of the irregular migrant population, who may frequently leave the country to return days, weeks, months or years later (so-called "circular migration").[66]

The fourth methodological problem in using border apprehensions data for the estimations of illegal migration flows is that apprehensions are influenced not only by actual volumes of

---

[63] In principle, the European Union has been collecting data on border apprehensions and illegal migration since 1995 in the context of its CIREFI (Centre for Information, Discussion and Exchange on the Crossing of Frontiers and Immigration) working group, but the data from this database on "apprehensions of aliens illegally present" do not distinguish between apprehensions of aliens illegally present *inside* the country and *at or near the borders* of the country that aliens have unsuccessfully tried to cross illegally. These data can therefore not be used for estimations based on border apprehensions.

[64] For example, data provided by the border police of the Czech Republic indicate that the share of persons repeatedly apprehended in 2003 was only 11.5%, slightly lower than in 2002. See Futo and Jandl, (2004).

[65] A formal model for deriving estimated multipliers based on the probability of apprehension is presented by Massey and Singer (1995). This model is then applied to an estimated probability of apprehension of 0.35 derived from a survey in 22 Mexican communities in Mexico and the US and the recorded apprehensions by the INS between 1965 and 1992. The survey involved a random sample of over 4 000 Mexican households and a smaller non-random sample of Mexican out-migrants identified by snowball sampling. A total of 1 229 undocumented migrants recounted 3 207 trips to the US. In 71% of the trips the number of apprehensions was zero, on 18% one, on 6% two, on 3% three and on 1% four.

[66] Greece provides a good example due to the high mobility of her Albanian migrants (constituting around 75 % of all apprehended illegal migrants), who may frequently leave the country to return days, weeks, months or years later (so-called "circular migration"). One indicator of the circular nature of illegal migration to and from this country is the extraordinary high level of expulsions of foreigners not in possession of valid residence documents. Between 1992 and 1995, these numbered over 225 000 a year. See Baldwin-Edwards, M. (2001, p.11).

illegal border crossing attempts but also by a number of external factors, such as stricter border controls.[67]

A final, practical, problem lies in the fact that estimates become quickly outdated and would need constant updating through new data. For example, the latest evidence on border apprehensions in Central and Eastern European countries indicates that the number of border apprehensions in 20 Central and Eastern European states has decreased by more than half between 2000 and 2006, at the same time as efforts of border control have greatly increased in many of these countries (Futo and Jandl 2007).

To summarize the discussion, there are a number of conceptual problems associated with the estimation of illegal migration flows on the basis of border apprehension data but to date there are no reliable alternatives available for application in the European context.

5.2. Estimating Net Illegal Migration with Differential Methods

While the estimated volumes of illegal border crossings (Section 6.1) provide some indication on the total volume of *gross* illegal migration flows, the *net* flows of illegal immigrants cannot be estimated on this basis (but is likely to be much lower). Net migration is defined as the difference between immigration and emigration. However, as with legal migration, net migration can also be measured as the (residual) difference between stocks (adjusting for mortality and fertility). Thus, theoretically, if the estimates on the stocks of illegal immigrants at two different points in time are reliable approximations of the "true figures" of these stocks, one could calculate the net flows of illegal migration, adjusting for regularizations, fertility and mortality. The logic of this approach of estimating (net) flow through stock data is thus the inverse principle of estimating stocks through (net) flow data (see Section 5.5.5).

This relatively simple calculation is regularly made for the United States (see, for example, Passel (2002 and 2006) and Hoefer et al (2007) for the latest estimates), but there are only few examples for Europe to date. One exception is a recent calculation by de Haas (2007) who estimated net irregular migration from West Africa to Europe on the basis of estimated yearly increases of West African migrant populations in France, Italy, the Netherlands, Portugal and Spain. De Haas (2007, p. 45f) first calculates the number and share of the increase in legal migrants from West Africa in Europe falling on Spain alone (20%). He then applies an assumed share of irregular to regular migrants from West Africa in these population increases that he takes from the Spanish case (i.e. a comparison of the municipal registry with the number of registered residence permits) and then multiplies the total with 5 for all of Europe. His results for irregular net migration flows from West Africa to Europe are significantly lower than others commonly cited in the literature.[68]

Moreover, taking a ratio of 1/3 illegal entries (compared to two thirds of overstayers) found in surveys carried out in West Africa by Eurostat already some years ago (Schoorl et al 2000),

---

[67] Stepped-up border enforcement efforts could result in more apprehensions in the short term and lower apprehensions in the long term by frustrating further attempts of illegal crossings. Other external factors include the removal of border controls within the Schengen area; visa-exemptions or, alternatively, new visa obligations for certain nationalities; and last but not least new – as yet undetected – modus operandi of human smugglers that result in decreased probabilities of apprehensions.

[68] De Haas (2007) estimated net irregular migration from West Africa to Spain on 15 000 each year. Spain's estimated yearly increase of West African migrant populations (20,500) accounts for about 20 percent of the estimated total yearly increase in Europe. This leads to a maximum estimate of 75 000 (15 000 / 0.2) irregular immigrants per year. See: De Haas (2007, p.45)

De Haas (2007, p.45) further ventures to calculate the total number of successful irregular entries of West Africans to Europe each year.[69]

5.3. Estimating the Number of Visa Overstayers with Residual Methods

When analysing the genesis of illegal residence, it is likely that even more so than illegal entry, overstaying the terms of one's visa (or entry conditions in case of visa exemptions) is the major route to illegal residence in European and other industrialized countries. However, estimates of illegal migration flows made on the basis of border apprehensions data are likely to seriously undercount so-called visa-overstayers[70] and few other methods for estimating the prevalence of "overstaying" have been applied in practice. One exception is the Italian Home Office, which provides detailed estimates that classify the unauthorized residents by type of entry into Italy. These data indicate that the overwhelming majority (between 60% and 75%) of the stock of unauthorized migrants in the period 2000 – 2006 was constituted by overstayers (Fasani 2008. p.60). However, while the data constitute some evidence of the relative importance of overstaying, it is not clear, how exactly these calculations were done and no clear methodology is specified. If, for example, data on regularisations or in-country apprehensions were used for this estimate, the underlying assumption would be that overstayers and illegal entrants would have the same probability to be represented in these data, an assumption that is far from certain.

Thus, the statement of Delaunay et al. (1998, p. 62f) seems to have continued validity: "No country in Europe manages to count or estimate visa overstayers, as the records on foreign visitors departing are incomplete". In fact, with the progressive expansion of the Schengen area, and the possibility to travel from one country to another on the same Schengen visa, it may be even less feasible today than only 10 years ago. Not even the UK (another island State which is not in the Schengen area and thus collects continuous, if somewhat rudimentary, border crossing data from its Passenger Survey on entry) succeeds in providing an estimate of their visa overstayers.

Examples from other countries indicate that estimating the number of visa overstayers from entry-exit records is not, in principle, impossible but depends on the specific migration and travel situation of the country and on the quality of the data available. An early example is mentioned by Morita and Sassen (1994), who calculate the number of visa overstayers in Japan for the period 1980-1992. Another example is provided by Australia, which claims to possess detailed knowledge at the number of visa overstayers (and consequently illegal residents in the country, as the number of clandestine entries is assumed to be very low) at any point in time (Australian Government 2007).[71] In both countries, the method applied to estimate the number of visa overstayers is the use of a double entry card system, in which persons entering the country fill out a card in two copies, one of which is retained upon entry and the other is retained upon departure. The number of aggregate non-matches (adjusted for mortality and status changes) represents the estimated number of visa overstayers and, hence, unauthorised immigrants (cf. also OECD 2006, p.44f.)

---

[69] Based on a maximum estimate of 75,000 additional irregular immigrants per year, De Haas (2007) reaches a provisional estimate of 25 000 successful irregular entries of West Africans to Europe each year.

[70] Visa-overstayers are counted only, when detected at borders when trying to leave a country.

[71] The number of visa overstayers in Australia at 31 December 2005 was estimated to be just under 46 400. This is a significant decrease on the estimate of around 49 400 at 31 December 2004 and around 59 300 as at 31 December 2003. Source: Australian Government 2007.

Both Japan and Australia are islands, a geographical fact that facilitates this type of control over their borders and the use of this estimation technique. However, even the Immigration and Naturalisation Service of the US, with its long border with Mexico and Canada has attempted to estimate visa overstayers among its undocumented immigrant population (cf. Lederer 2004, p. 198). While the method involves complex computations from large datasets, a congressional body noted that the reliability of the results is still lacking and needed further improvement (General Accounting Office 1995).

6. Conclusions

This chapter has attempted to provide an overview of existing and applicable methods to the estimation of irregular migration phenomena in Europe. The emphasis was on methodological clarification and categorisation of different estimation methods and not on a full coverage and evaluation of existing estimates. The basic logic of methodologies was laid out in a commonly accessible language leaving statistical intricacies and mathematical formulas out in the references for access by the specialists. A critical evaluation of the described methods and a discussion of any special data requirements or data problems were included as a guide to the description, categorisation, evaluation and classification of existing estimates found in European countries.

In the conclusions of their comprehensive review of existing estimation methodologies in Europe, Delaunay and Tapinos (1998, p.72) had stated that "the most obvious and disappointing finding which emerges … is that we have practically nothing that is well-founded". The next review of methods by Pinkerton et al (2004) examined the applicability of estimation methods to the UK case and sounded already much more optimistic, resulting eventually in the first official estimate on the UK (Woodbridge 2005 uses the residual method). This review, only 10 years after the first one by Delaunay and Tapinos (1998), could add again a number of new studies, techniques and estimation methods that may or may not be fruitfully applied in the European context. The question is, then, whether this multitude of existing methods has actually been used in practice or whether most existing estimates are not much more than subjective "guesstimates".

Two questions that may emerge "Why so many methods?" and "Which one is the best?" should be briefly addressed here. These questions are linked. As there is no "best" method (only methods that are more or less suitable to the specific data and migration situation in a given country) it is also not enough to have one or only a few methods. The applicability of any given method depends much on the migration context, the available data base and the resources and time frame available. Some are more suitable in a particular context than others and some may not be applicable at all due to the lack of data or other considerations.

Chapter 4: Classifying the quality of estimates

Dita Vogel

1. Introduction

Chapter 3 gave a comprehensive overview over methods to estimate the size of irregular foreign resident populations and irregular entrants. It became clear that there is no easy way to asses the quality of estimates by looking at the methods. Therefore, the next section develops a classification for estimates to assess their quality. With this classification, we want to label estimates that we collect in the project CLANDESTINO and thus achieve more transparency on the state of knowledge concerning this question.

Section 2 introduces and discusses quality criteria in quantitative and qualitative research. Section 3 discusses key problems in the application of quality criteria to the problem of estimating the size of irregular migration. Section 4 applies quality criteria to this problem. Finally, section 5 proposes a classification scheme.

2. Quality criteria in social research

Our research question belongs to descriptive population statistics in the sense that we seek to assess the size of a population, seeking rather to describe how things *are* than to explain *why* a phenomenon occurs. Thus, we are mainly concerned with data collection and data quality, while the problem of causal inference that dominates a lot of the methodological discourse is of only limited relevance. However, it comes in through the back door as estimation from limited data implies some assumptions about causal links.

The estimation of the size of undocumented population thus belongs to the field of quantitative social research, but as it also concerns a hidden and hard-to-reach population. For such populations, qualitative studies are much more frequent than quantitative studies, and quantitative studies require at least additional qualitative research work because otherwise the data quality cannot be assessed. For example, an understanding of administrative data may require reading official documents, observing administrative procedures and interviewing experts – a well-known problem in criminology that is "haunted by the dark figure" (unregistered offences in comparison to registered offences) (Coleman and Moynihan 1996). Therefore, it is useful to look for quality criteria of social research both in quantitative and qualitative research.

In this section, quality criteria and procedures to ensure quality in social research are presented and discussed in view of this research problem.

2.1. Quality criteria

Reliability and validity are elementary criteria for the quality of research methods (Lewin 2005:216). They seem to be so elementary that many textbooks of statistics often do not care to mention them although all the presented techniques serve at nothing else than ensuring the reliability of results. *Objectivity* is often added as an additional criterion that aims at ensuring that results are not dependent on the investigator (Lamnek 2005:172), but objectivity can also be subsumed under reliability because a result that cannot be reproduced by other researchers is not reliable.

Reliability refers to the accuracy, stability and consistency of measurements. A procedural definition of reliability for quantitative research calls a result reliable if the results are confirmed in repeated tests. A result is relatively unreliable if it has large error range, is not stable over time, and is not robust, i.e. reacting sensitively to slight variations in the method (e.g. in the framing of a question in a questionnaire).

Validity refers to whether or not the measurement collects the data required to answer the research question. There is a wealth of different concepts of validity and of procedures to ensure validation of results both in qualitative and quantitative research (Teddlie and Tashakkori 2003:13), but they all address the overarching question of the adequacy of the method for the research question. A method can be useless when it is reliable but not valid because something irrelevant is measured in an accurate and replicable way. On the other hand, an unreliable method cannot produce valid results, because a high degree of unreliability implies that it does not adequately measure what is intended to measure.

Internal validity concerns the credible establishment of relations *within* a study, *external validity* concerns the generalizability and transferability of results to other persons, contexts and situations (Campbell and Stanley 1966 according to Seale 1999: 38-41).

Validity and reliability in a more general sense are applicable both to quantitative and qualitative research questions, although means to ensure them differ between the fields and tend to be more technical in quantitative analysis and more procedural in qualitative analysis.

This view is not undisputed among qualitatively working scientists, some of whom argue that a different research paradigm and style requires different quality criteria (Lincoln and Guba 1985). Qualitatively and quantitatively working researchers sometimes analyse the same type of questions with different means, but often simply choose different research questions. While qualitatively working researchers emphasize that all social realities are constructed and analyse these diverging constructions, quantitatively working researchers tend to subscribe to a particular measurable construction of reality and analyse it in detail. We agree with Brühl and Buch (2006: 8) that quality criteria as such are not all that different beyond preferences for other terminologies. Seale argues into the same direction: “Factuality may be a fiction, but it appears to be a useful one, and attention to how (constructed) facts relate to (constructed) claims and theories is a widely recognized hallmark of good-quality research” (Seale 2004:411). Both research paradigms subscribe to general criteria for good social research, although often using different terms and necessarily relying on different means to establish them. Table 5 gives an overview over quality criteria in the quantitative and qualitative paradigm.

Table 5: Quality criteria in social research

	Quantitative paradigm	**Qualitative paradigm**
Internal validity	extent to which variations in outcome can be explained by variations in independent variable, usually by controlling or randomizing other factors	extent to which the outcome of a qualitative inquiry can be made credible, usually by different types of triangulation of data and methods, analytical induction and communicative validation
external validity	adequacy of the study design for studying wider social questions, possibility to generalize across different types of persons, settings and times	applicability and transferability: the extent and conditions of applying results in other contexts
Reliability	accuracy and stability of a result, usually tested by replication	clarity and dependability of a result, demonstrated by procedural clarity and safeguards such as repeated interpretations by same or other investigators
Objectivity	inter-subjective agreement, independence of data and results from individual investigators	confirmability, arguments on the basis of data, avoidance of biased statements, multiperspectivity

Source: Own compilation on the basis of different textbooks, inspired by the discussion in (Lincoln and Guba 1985) and the tabulation of Brühl and Buch (2006:22).

2.2. Procedures to ensure quality

A number of procedures are only relevant for specific methodological instruments. Many statistical instruments are only applicable to quantitative studies. A textbook for econometrics (as one specific discipline of quantitative empirical studies) lists for example quality criteria for good econometric formula: Computational costs, least squares, highest R^2, unbiasedness, efficiency, mean square error, asymptotic properties, and maximum likelihood (Kennedy 2003:11). Obviously, these criteria can be used in procedures to ensure quality in econometrics, but have no relevance for the analysis of qualitative interviews. Therefore, we would rather interpret them as indicating the relevance of specific *procedures* in econometric research, and of course within these procedures there can be specific quality criteria.

There are also procedures to ensure reliability only in specific types of qualitative studies. For example in studies dealing with qualitative interviews, there are standards for the transcriptions that transform oral material into written text (Mayring 2002), and there are coding strategies to ensure intercoder-reliability when several investigators label texts according to the same rules (Brühl and Buch 2006:28).

Here, I concentrate on those procedures to ensure quality that are of a more general character and can also be applied to our research task. In the econometrics textbook quoted above, such procedures are discussed under the heading "applied econometrics", and the formulation as

'ten commandments' indicates their normative character: "Thou shalt"[72] use common sense and theory, ask the right question, know the context, inspect the data, avoid worshipping complexity, look long and hard at the data, beware the costs of data mining, be open to compromise, do not confuse significance with substance, and report about the sensitivity (Kennedy 2003:396). These rules and procedures are applicable to all types of research.

In his text on "the quality of qualitative research", Seale categorizes strategies to achieve good research into "inner and outer dialogue", making explicit reference to the standards for internal and external validity in quantitative studies (Seale 2004). His approach prescribes a more general procedure which is useful for our purpose. With inner dialogue, he means that researchers strive for a rigorous consistent argument that connects claims and evidence, searching actively for contradicting evidence and taking alternative explanations into account that may appear *within* the framework of the study. It also includes all forms of data, method and researcher *triangulation* within the project – all efforts to ensure and improve quality by looking at several data sources, using several methods and including several researchers in the analysis. King and others (1994: 24) even recommend collecting data on as many observable implications of a theory as possible. This may also include communicative validation in the sense of giving systematic feed-back of research results to those who were subject of a study (member check).The systematic search for negative cases and alternative explanations that could disprove the claim is also one means to increase the validity of an argument. In cases in which the estimation includes statistical analysis, the usual statistical reliability criteria have to be discussed.

The outer dialogue – according to Seale 2004 - concerns external relations of the research project, including the dialogue with other researchers and also with practical or political experts. This can also help to pose non-trivial questions and to avoid "type III errors": Producing the right answer to a wrong question. "An approximate answer to the right question is worth a great deal more than a precise answer to the wrong question" (Kennedy 2003:391). It includes a discussion of the research results against the background of public discourses, the acknowledgement of the state of art and the discussion of the own results in this context. The adequacy of methods for making contributions to social questions is checked. Outer or external dialogue involves communicative validation in a more general sense of systematic discussions with scientific and practical experts, for example in conferences, workshops and individual discussion.

Documentation of research questions, concepts and procedures is a *precondition* for assessing the validity and the reliability of research. Some methodologists list transparency of documentation as independent criterion and assign a first place under the research quality criteria (e.g. Mayring 2002: 144). In order to convince readers of the validity (trustworthiness, credibility) of the results and to make results accessible to the critical assessment of other researchers, theories, concepts and procedures have to be documented in sufficient detail. Documentation includes internal documentation (archiving qualitative and quantitative data which are usually not published; detailed description of research methods and procedures) and external documentation (in the sense of a sufficiently detailed reference to these data and procedures in publications such as books and articles).

[72] Biblical English: "You should"

3. Estimates of irregular residents and entrants – problems and solutions

It is widely agreed that most numbers quoted as estimates of the size of undocumented, irregular or illegal migration stocks and flows are problematic (Tapinos 1999; Jandl 2004). However, most researchers would probably also agree that not all estimates are *equally* problematic. Before proposing a scheme for a basic evaluation of the quality of such estimates, we first look at the key problems that have to be addressed in estimating the size of irregular migrant populations. Most studies involve a discussion of particular problems and the ways that their specific study addresses these problems. They were surveyed in the context of the methods in chapter 3. The following section is based on chapter 3 and summarizes some more general problems as well as attempts to solve them.

3.1. Problems of definitions

Irregular migration is problematic to define (see above chapter 1). Therefore, it is of crucial importance that estimates of the size of irregular migrant populations clearly indicate to which definitions they subscribe. This seems to be self-evident, but still it is not always the case. Here, particularly problems with definitions of irregular migration are pointed out in order to raise sensitivity for these problems.

Definitions of irregular migration always include elements of a negative definition: Migrants who have *no* officially recognised status (or have *no* regular residence status). Therefore, changes in policies and practices towards regular migration lead to changes in irregular migration. The term remains the same, but the population concerned may change substantially.

Definitions of irregular migrants have to be particularly clear on the following issues:

- Do they refer to stocks (illegal residence) or to flows (illegal entrance and overstaying) and avoid mixing both?
- Do they specify whether a stock estimate is made for the summer or the winter (high and low season for seasonal work)?
- Does the definition exclude particular groups without regular status, for example informally tolerated persons?
- Does the definition include illegally working persons with a regular status that are liable to expulsion if their work is detected, such as working tourists and seasonal workers working in breach of their contract?
- Does the definition include persons with a seemingly regular status, but who have false papers or identities or gained their status under false pretences so that their status may be withdrawn if it comes under scrutiny?
- Does the definition correspond to the problems that are addressed? Sometimes, studies define the estimated group very narrowly, but seek to address a problem of a much larger and more varied group.

3.2. Problems of collecting data

Irregular migrants are difficult to access in surveys. The basic problem is the difficulty to identify irregular migrants – precisely because they break immigration rules and *fear*

sanctions. In addition, irregular migrants are typically a *small group* in the total population so that even efficient screening mechanisms would require very large surveys. And even if irregular migrants are included in surveys it is questionable whether they report their irregular situation correctly. Many other problems of estimating the size of the irregular population are directly or indirectly related to these problems. The most important consequence is that all data sources about irregular migrants are *biased* in one way or the other.

One way to overcome the access problem is by using *organisational data.* There are three types of occasions when immigrants can be identified in organisational data:

- When they are checked by control authorities,
- When they self-identify to control authorities in order to gain regularisation,
- When they self-identify to service and welfare organisations to seek help.

These data sources usually have two problems:

Firstly, administrations often count cases and not persons, so that the same person may appear repeatedly in the same data. For example, police usually accounts for apprehensions of foreigners without valid residence status, but often does not register when the same person is arrested several times. If we are interested in the size of irregular border movements, case data may be fine, but usually we are interested in data on a person basis.

Secondly, organisational data are biased with respect to the organisation's task.

- It can be assumed that the probability of being checked by control authorities varies widely between different groups of irregular migrants, and that we do not know the likelihood, although there can be some assumptions on direction of the biases from descriptions of authorities' practices. Women are less likely checked than men, old people less than young, workers in private households are less likely to be checked than those working on public sites, majority-looking persons less than those with a non-majority physical appearance.
- Selection biases of regularisation data are influenced by the criteria set for regularisation and the size and average length of stay of immigrant communities, indicating their access to help and information.
- NGOs and service organisations are also biased. Depending on the type of NGO, we can for example expect a bias towards a specific religious group, towards persons from established minorities, persons with a prior legal residence, persons with asylum background, women or persons with serious problems.

Building up trust is also a way to overcome the access problem. This is usually done in variants of the *chain-referral or snowball sampling surveys* in which the researchers start sampling from original contacts. These studies are often cost-intensive therefore limited to a city or region. They are usually biased with regard to the original contacts. More recent arrivals and more isolated persons are less likely to be included in such studies.

Advanced screening mechanisms seek to avoid the compliance problem in surveys. Randomized response and 3-cards methods allow for a statistical calculation of answers to a sensitive question such as the status question without requiring that individuals have to directly reveal it, but they do not solve the 'small size'-problem.

Indirect methods try to assess the size of undocumented populations without identifying the individual undocumented immigrants, for example by comparing registers or by drawing conclusions from economic data. However, they often have to make far-reaching assumptions

about the inclusions of irregular migrants in different registers, or the attribution of particular economic activities to irregular migrants, so that the identification problem comes in indirectly in indirect methods.

3.3. Problems of handling biased data

Biased data is not ideal, but if there is only the choice between different types of biased data, biased data is better than having no data at all. It is above all important that researchers are aware of the biases in the data they use. Awareness can be raised by logical and theoretical reasoning and by empirical investigation. Researchers may apply insights from migration and control theory to the data and thus formulate ideas about the direction of a bias. Using different data sources and expert interviews is an empirical way to become aware of biases. Particularly with regard to organisational data, investigators need to understand the work process and data reporting procedures of the organisation in order to get an idea of the data bias.

It is easier to acknowledge the existence and direction of a bias than its size. Therefore, one way of dealing with major biases is to make different assumptions about them and present different size calculations based on these assumptions. Another way is to use data triangulation in order to exclude implausible size estimates.

Some methods use statistical procedures to estimate the size of undocumented population, making statistical or discrete adjustments to the calculations in order to minimise the impact of the bias.

Capture-recapture measures (repeated capture, matching registers) are in principal capable of calculating the size of hidden populations. However, as the central assumption of a stable population is usually not fulfilled in the case of undocumented migration, the recapture sample is biased, too. Adjustments for known biases and different calculations for different groups can be made to minimise the impact of the bias.

Other studies collect information to assess the likelihood of inclusion of an individual and use this information to eliminate the bias statistically. This is the case with the centre-sampling approach in Italy (e.g. Baio a.o 2008). Link-trace and respondent-driven sampling (Salganik and Heckathorn 2004) collects network information, estimates the likelihood of being included in a survey from this, and weights the size estimate with these likelihoods. This way, they try to offset the effect of first contacts and differing likelihood to be included by snowball-sample surveys. These methods are cost-intensive and difficult to implement properly.

3.4. Problems of extrapolation and aggregation in space

Because access to irregular migrants is difficult, studies are often confined to one city or one neighbourhood in a big city, usually a place with a high expected incidence of irregular migrants. This enables the researchers to make a solid assessment of all available data sources and to include local expert knowledge in case of organisational data, and it makes building trust for snowball sampling or a thick area study (postal code method) possible in case of an attempted survey. Some local studies seek to achieve high reliability using multiple data sources, intensive and prolonged engagement in the area, method triangulation including the use of statistical methods, but the extrapolation to the national level is based on a simple multiplier method. A local multiplier is calculated for extrapolation in space (e.g. irregular migrants in relation to population or to foreign population) and applied to similar cities or the whole country.

Unfortunately, there is no reason to believe that irregular migration is distributed evenly in space. Legal migration is unevenly distributed within countries. The distribution is dependent among others on historical migration patterns and current labour market chances. Typically, persons from the same countries of origin concentrate more in some places than in others due to the role of networks in migration. Thus, if the size and characteristics of the irregular population of a certain city is estimated, a simple multiplier may not be sufficient to extrapolate the result to other cities. Therefore, even a high quality local study may lead to a low quality national estimate.

Leerkes and others (2007) analysed factors that influence the spatial distribution of irregular migrants in the Netherlands based on apprehension data. They found a proof that the relative size of regular migration, labour market opportunities, cheap housing opportunities and the share of single-person households positively influenced the presence of undocumented immigrants. Factors like this can be taken into account more or less systematically when multipliers for national estimates are chosen.

In principle, the problem can also be solved by making several local studies with different contexts and aggregating the results directly, or using them for calculating different multipliers.

Also, European countries differ considerably in size. For small countries, the aggregation problem is negligible when most of the irregular migrants are concentrated in the capital, but big countries with several differently structured cities and agricultural areas with differing types of production face manifold problems of estimating. Naturally, estimation problems in bigger countries affect the European total more than those in smaller countries.

On the European level, the problem of space is aggravated (Clarke 2000). Usually, European efforts to get an idea about the size of irregular populations rely on the review of studies that differ largely in their methods (see chapter 2). Even if they rely on the same type of data – apprehension data by control authorities – these data are not completely comparable as reporting practices differ. In addition, aggregation is complicated by the fact that there may be an overlap of data. An undocumented migrant may be caught in more than one country, for example first in the country of transit, and later, if not effectively deported in the first country, in a country where the person lives.

3.5. Problems of extrapolation in time

Measuring migration and migrant populations as such already poses some problems, because it involves mobile populations. These problems are aggravated in case of irregular migrant populations. When there are serious efforts to estimate the size of the irregular migrant populations, they are usually not undertaken every year, because they tend to be time-consuming and costly.

Therefore, the question of the depreciation of an estimate should be raised. How long can an estimate be used, and which procedures can be used to adjust older estimates without a new full estimate? This is a question that mostly concerns stock estimates. Chapter 1 outlines the links between stocks and flows. Extrapolations in time should make use of such a model and reflect on the likely changes in the stock in light of information that is available about the flows.

As indicated in the section on definitions, irregular migration is sensitive to policy changes. Therefore, substantial policy changes with regard to legal migration and migration control should be reflected when data is extrapolated in time. At best, these reflections allow for adjustments and extrapolations of an earlier estimate, at worst, they make last year's statement

completely useless because no reasonable assumptions about the effects of policy can be made without a renewed substantial empirical study.

Policy changes are often on a national level, for example regularisation programmes, new options for regular labour recruitment or an expansion of internal labour market control activities. On the European level, there are two fundamental changes in the reporting period: The accession of 10 new member states in 2004 and the accession of two new member states in 2007. Formerly illegally resident Third country citizens in the EU-15 member states became EU citizens with many more options for legalising their status and much less state power for expulsion, although EU accession did not automatically include the full mobility rights for the new EU-citizens. In states with substantial migration from CEE countries, estimates are devalued after the accession or have to be explicitly adjusted for the new member state nationalities.

So far, the question of extrapolation in time has been rarely addressed in the methodological discussion. In the United States, there are efforts to make a serious estimate of the size of irregular migrant populations in census years, and to make adjustments with the help of available administrative data in the years between censuses. No systematic data adjustment procedures are used in Europe. In the public and political discourse, numbers are often quoted without reference to a time frame, as indicated in the introduction.

Sometimes, extrapolation in time does not go forward from an earlier estimate but backwards from current data. This happens in case of legalisations. When 100 000 legalise and the legalisation programme requires a minimum residence duration of 5 years, we can conclude that there were at least 100 000 irregular migrants in the country in the past five years. However, this claim rests on the assumption that the residence period can be securely confirmed in the regularisation process. There can be doubts about this. Some documents may be falsified, and some real documents (such as proofs of registration in a municipal register) may not correspond to the real presence in the country.

4. Applying quality criteria to estimation of irregular migrant populations

What is a good estimate? How can the quality of estimates be classified? What conditions would estimation procedures have to fulfil so that we have considerable trust that the estimate comes close to the true unknown value of the irregular population? In other words, how would we define the high end of a quality scale? We propose the following criteria for a high quality estimate and explain them in more detail.

4.1. Documentation

Not all numbers that are quoted as "estimates" of irregular residence can be called an 'estimate' in a social scientific sense. A social scientific estimate can be defined as a number or a set of numbers gained *methodically* that aim at coming close to a true, unknown value of a parameter. This definition picks up a formulation for econometrics. Estimates require an estimation method or estimator (Kennedy 2003:4-5). These may include statistical estimation techniques but also qualitative ways of narrowing the range for the true value of irregular migrant populations.

Concerning our question, the *minimum documentation requirements for a scientific estimate* include a clear definition of the irregular population group, a specified time and space, the

description of the estimation procedure and of the underlying data sources. It also specifies if it uses non-fact based assumptions.

The press frequently quotes numbers of irregular migrant populations for whom we do not know how they were derived: There is no identifiable method and sometimes not even a source. An expert may have forwarded a number, and we do not know whether this number was estimated from intimate knowledge of data and plausibility calculations, or whether it was put forward out of the blue to deliberately dramatise or downplay a migration situation. A government body may have forwarded informally the results of a detailed calculation with administrative data to the press, without wanting to be officially connected to this calculation. We report such numbers in our study if they are widely quoted, but it is clear that they do not fulfil the minimum requirements of what is called an estimate in scientific literature. There is no way to assess the quality of the estimation. From a social scientific point of view, we cannot put much trust in an estimate of which we do not know how it was gained.

4.2. Validity

Validity requires adequate methods to be applied in an adequate way. What are adequate methods for the estimation of the irregular population? Unfortunately, no single method as described in chapter 3 will be the best recommendation to achieve these aims in all countries. Not all methods are applicable in all circumstances and countries. They depend in differing degrees on the size of the country and the group in a particular area, on the relative stability of the population during the data collection period, on the documentation and reporting practices of organisations that may deliver data, and on the control intensity. Therefore a method that produces valuable results in one context may be not applicable or yielding poor results in another context because the necessary preconditions are not fulfilled. However, some general statements on the applicability of methods can be made.

Methods involving expert assessments have a different character, depending on the *size of the space and group*. Expert's assessments are highly valuable if they speak about first and second hand experiences – cases they know personally or about which they have information in the context of their ethnic community or work environment. If researchers ask experts to make estimates about a group and a space that they are likely to oversee, they tap on an unwritten data source that is not counted and exact but nonetheless fact-based. If researchers ask experts about a group or space which they cannot oversee, they will have to make their own calculations or repeat or adjust numbers that have heard in the discourse. Therefore, such expert assessments are either displaying a discourse, or consist of estimates without any scientific documentation.

The applicability of *residual methods* presupposes that suitable databases exist, are available to social scientists in sufficient detail and with sufficient information so that their limitations can be assessed. The higher and the more comprehensive the intensity of controls directed at irregular migrants, the less likely is the *existence of such databases*. If there are databases in a country that are likely to include most irregular immigrants (such as census data in the US and municipal registers in Spain) and other data sources, that exclude irregular immigrants (such as registers of regular immigrants), these are a good starting point for an estimate. It requires a good knowledge of the all involved data sources so that the researcher is aware of over- and undercounting possibilities and errors in both databases that could influence the estimate.

Capture-recapture strategies are best for *settled irregular populations* with low risk of deportation and less useful for populations with a high percentage of returns and new arrivals in the data collection period. If one and the same individual leaves an identifiable statistical trace at different points in time, this can be used for a statistical calculation of the size of the

underlying population. This presupposes above all the existence of a relatively large micro-database that is accessible to researchers.

Surveys with irregular immigrants require the building of trust, and are thus suitable for an environment in which trust-building is possible. Therefore, estimation methods involving immigrant surveys will usually be *confined to local contexts and/ or specified nationalities* and require additional methods for extrapolation to the national context. Representative samples are not possible, and methods to adjust for sampling biases have so far only be used in Italy (centre-sampling method).

On the local level, there are more options to conduct a comprehensive and consistent study. Therefore, a generally applicable research strategy involves *a sufficient number of comprehensive local studies with sophisticated extrapolation to the national level.*

Comprehensive local studies can principally be executed in different ways. They may primarily rely on only one data source, e.g. apprehension data, expert survey, or immigrant survey and take other evidence into account in the internal and external dialogue about the outcome, or they can triangulate data from different sources from the beginning. Different data collection methods are possible and adequate in local studies, as long as they are implemented comprehensively and consistently.

The number of necessary local studies depends on the size of the country and the variation in the social and economic contexts. The state of art does not seem sufficient to indicate the best way for a sophisticated extrapolation to the national level, but it is clear that simple multiplier (e.g. applying percentage of irregular migrants to population in one city to the whole country) is likely to yield only very rough results. Regression analysis is one way to extrapolate from the local to the national level, taking the socio-economic differences into account. Differentiated socio-economic multipliers could be an alternative.

To make minimum or maximum estimates is a way to increase the reliability of the result (see below). However, minimum and maximum estimates have to be made with suitable methods in order to be considered as valid. They have to be method-based, comprehensive and consistent. It is not adequate to estimate only a subgroup of the irregularly resident population and declare this as minimum estimate because the estimate lacks comprehensiveness. The simple characterisation of apprehension or regularisation data as minimum estimates is also not adequate, as the estimation is circumvented. It means that a multiplier of 1 is assumed without assessment whether this is a reasonable approach. If a regularisation programme for example was open only for long-term residents of one nationality, regularisation figures lack the comprehensiveness required for a minimum estimate.

4.3. Reliability

Even if methods are principally adequate and well applied, how do we rate whether they deliver reliable results? What does reliability mean in a context in which we do not have any unbiased data sources, if any at all? What degree of accuracy and dependability can be expected?

A reliable method should be *replicable*. If the same type of study is conducted again, the differences in the results are ideally only influenced by a change in the size of the irregular migrant population. Even well-designed and well-conducted studies will not always be able to assure this because not only the number of irregular migrants but also the applicability of methods may be influenced by changes in policies and implementation. Replicated studies have to discuss how they attribute changes. A reliable study should at least in principle be replicable, even if adjustments and limitations may be involved.

Secondly, researchers should aim at the best possible degree of accuracy of their estimate, while raising awareness to the sensitivity of the estimate to central assumptions and limitations in data and method. This can be done by *making the quantitative implications of uncertainty visible* for the reader. They can indicate *ranges* between which the real unknown value of the irregular migrant population is likely to lie. They can make *alternative calculations* on the basis of different assumptions, as they are for example also used in population forecasting. This way, they demonstrate the robustness of their estimate to central assumptions. If they know the direction but not the size of the bias of their data, they can characterize an estimated figure as *maximum or minimum estimate*.

If we allow for ranges, there is a trade-off between accuracy and dependability of the result. The wider the range, the more likely the real value is in the range. A very wide range may lead to a result that is virtually meaningless for the political context in which the question was asked. However, even a wide range can make a meaningful contribution to policy discussions, particularly if it deviates from unsubstantiated dramatising or downplaying numbers in public discourse. The assessment whether a range is narrow enough to deliver a meaningful contribution to the knowledge in a specific country is left to the interpretation of the readers.

We can only assess the reliability of estimates if reliability considerations are documented, at least by raising sensitivity for limitations and at best by additionally making them quantitatively visible. If limitations are severe, for example because important parts of the estimates are based on assumptions and not data-based, reliability is low.

5. Proposal of a classification scheme

After having discussed the aspects that have to be taken into account in the preceding section, this section presents a classification scheme. As with all classifications into discrete classes, there is a normative and an arbitrary element in it. The norms are those of good social research as presented above. The categorisation aims at being plausible and understandable for non-expert users and suitable for differentiating the quality of existing studies.

The main classification into three quality classes follows a 'method assessment logic': The validity, reliability and documentation of estimates is assessed. The low quality class is additionally scrutinized with a 'size assessment logic'. If the size of the low quality estimates seems to be much too high, too low or seriously misleading, it gets the additional classification 'with plausibility warning' to alert readers to be particularly careful with the result.

It has to be noted that this classification scheme only aims at classifying the quality of the size estimate and does not make an attempt to classify a study as such. Studies that are of low quality with regard to a size estimate may still be of high scientific value, for example because they deliver a size estimate only 'by the way' in the introduction and aim at different issues in the main part. The following paragraphs explain the logic. Concrete rules and examples how the logic is applied in the framework of the CLANDESTINO study will be explained in a separate paper (Vogel and Kovacheva 2008).

A *high quality estimate* gives sufficient documentation on method and data so that the quality can be assessed. It uses an adequate method in an adequate way (validity) and aims at accuracy while making the limitations of the estimate quantitatively visible, and raises awareness of the limitations of the estimate (reliability). If the critical discussion of the estimates indicates some shortcomings in these aspects, we will still classify the estimate as high quality if these deficiencies are considered to be minor and not likely to influence the result in a substantial way.

A *medium quality estimate* is also scientific in terms of comprehensive documentation and awareness for the shortcomings of the estimate. Critical discussion of the estimate should come to the conclusion that there is a step towards narrowing down the range of an irregular population, but there are reasons to put not too much trust into the estimate. Medium quality estimates may be the result of a not fully convincing method choice and insufficient method application in a sophisticated study, but they can also be the best possible estimate that is only partly based on data and has to make assumptions about some subgroups of a total. Besides from documentation, all medium quality estimates should be able to claim that they are comprehensive and consistent, and that the core of the estimate relies on data, even if weak data, and not on assumptions only.

All estimates without further explanation are *low quality estimates*, but also those estimates that use clearly inadequate methods, apply methods inadequately or use inadequate data with presumably high influence on the results.

Low quality estimates may still be plausible or there could be clear indications that the estimate is misleading. In the latter case, a *plausibility warning* is added. Even a low quality estimate with plausibility warning could turn out to be justified by later estimates of a higher quality that follow a more comprehensive and consistent approach. However, the current state of knowledge justifies warning the reader to be particularly careful with such an estimate, because arguments are put forward that it seems implausible compared to reliable data, or estimates of higher quality, or compared to estimates for other geographical units.

Categorisations should always be considered as categorisations under reserve that can be revised when new studies, information or arguments are available. Table 6 summarizes the classification scheme.

Table 6: Quality classification scheme

Quality classes	Definition
High quality estimate	documentation: sufficient information on data and methods provided reliability: replicable study with limitations quantitatively indicated (e.g. ranges, alternative calculations, characterisation as minimum or maximum estimate) validity: adequacy of methods and data convincingly demonstrated and method carefully applied, comprehensive and consistent study, only minor insufficiencies
Medium quality estimate	documentation: sufficient information on data and methods provided reliability: indication of reasonable reliability indication, although not necessarily in quantitative terms validity: methods are not fully adequate and/or not fully adequately applied and/ or underlying data are not fully adequate
Low quality estimate	documentation: insufficient information, and/ or reliability assessment: lacking or failing, and/ or validity: inadequate method, inadequate method application, lacking or very weak foundation in empirical data.
Low quality estimate with plausibility warning	low quality plus a plausibility check procedure that leads to serious doubts about the size of the estimate

6. Final remarks

In the project CLANDESTINO, we collect and classify estimates for all countries under review. This way, we aim at transparency of the available data and methods and may thus indicate a way to a better assessment of the migration situation in the future.

The assessment of existing estimates will rely on country reports from Austria, the Czech Republic, Germany, Greece, France, Hungary, Italy, the Netherlands, Poland, Slovakia, Spain and the UK, thus depending on the comprehensive review of estimates and critical discussion of methods that are delivered by the country experts. Reports and tables with classified estimates are made available online. In the academic dialogue following the publication of these efforts, there may be needs for revision. The revision may concern the classification scheme itself, the rules for applying the classification scheme to estimates, or the classification of individual estimates according to the rules. The amendment and revision of tabulated information is seen as an essential part of the project method, because it is considered the only way to gain a better overview over diverse data situations in Europe.

With this exercise, we will get to a better assessment of the size of irregular migrant populations in the EU than presently available. 'Better' is a relative statement – to a large degree we expect that our work will make the level of ignorance on this question more transparent than before, indicating that important policy decisions are taken on the basis of muddling through and guesswork. In an earlier debate, Böcker and Groenendijk (1986:30) doubted the usefulness of estimates of the total irregular migrant population, stating that policy makers would be better off with uncertainty than with a fake accuracy. In contrast to them, we argue in favour of making the best possible estimates because otherwise estimates out of the blue will be taken serious. In agreement with their statement, we argue in favour of *classified* estimates in order to substitute a fake accuracy that is currently often underlying policy decisions.

Beyond the aim of retrospectively assessing the state of knowledge, this methodological report also has a prospective aim. It indicates ways to better information in the future. Chapter 2 indicates the current relatively weak state of aggregate estimates on the EU level. Chapter 3 discusses a wide range of available methods for achieving estimates, which may be considered for future estimates. Chapter 4 lays out quality standards that future estimates would have to fulfil to avoid being classified as low quality. While high quality estimates are usually expensive and may not be possible under all circumstances, we would argue that an increasing number of medium quality estimates would already improve the information situation in the European Union considerably.

At a later stage in this project, we will discuss whether we have sufficient data to dare making a medium quality estimates for the EU level. In the current situation, EU level assessment on the size of undocumented population rests on the approximate compilation of incomparable studies and unscientific opinions so that a comparable plausibility estimate would already be an improvement.

References

Alt, J. (2003). Leben in der Schattenwelt. Problemkomplex illegale Migration. Karlsruhe, von Loeper.

Aparicio Gomes, R., Ruiz de Huidobro De Carlos, J.M. (2008), Modes of Migration Regulation and Control in Spain, in: Doomernik, J. and Jandl, M., (eds.), (2008), Modes of Migration Regulation, in: IMISCOE Research Series, forthcoming

Australian Government/ Department of Immigration and Citizenship (2007), Overstayers and People in Breach of Visa Conditions, Fact Sheet No.86, available at: http://www.immi.gov.au/media/fact-sheets/86overstayers.htm

Baio, G., Blangiardo, G.C. and Blangiardo, M. (2008) Centre sampling technique in foreign migration surveys (forthcoming)

Baldwin-Edwards, M. (2001), Semi-Reluctant Hosts: Southern Europe's Ambivalent Response to Immigration, MMO Working Paper No. 3, November 2001, Mediterranean Migration Observatory, Athens, available at: http://www.uehr.panteion.gr (4.2.2003)

BAMF (2005), Illegal aufhältige Drittstaatsangehörige in Deutschland. Staatliche Ansätze, Profil und soziale Situation. Bearb. A. Sinn, A. Kreienbrink and D. von Loeffelholz. Forschungsstudie 2005 im Rahmen des Europäischen Migrationsnetzwerks. Nürnberg: EMN Nationaler Kontaktpunkt Deutschland.

Belser, P., de Cock, M. and Ferhard M. (2005), ILO Minimum Estimate of Forced Labour in the World, ILO, Geneva, April 2005

Biffl, G. (ed.), (2001): Arbeitsmarktrelevante Effekte der Ausländerintegration in Österreich (Labour market relevant effects of the integration of foreigners in Austria), WIFO (Austrian Institute of Economic Research), Wien 2001

Bijak,J. and Izabela Korys (2006) Statistics or Reality? International Migration in Poland, CEFMR Working Paper.

Blaschke, J. (2008), Trends on Regularisation of Third Country Nationals in Irregular Situation of Stay Across the European Union. Briefing Paper, European Parliament, Civil Liberties, Justice and Home Affairs Committee, PE 393.282, Brussels

Bilger, V., Gendera, S., Hollomey, C., Jandl, M. und Stepien, A. (2006), Migration und Irreguläre Beschäftigung in Österreich: Ergebnisse einer Delphi-Studie, International Centre for Migration Policy Development (ICMPD), Vienna, 2006

Böcker, A. and C. A. Groenendijk (1996). "Vuile handen of verbrande vingers? Een antwoord of Burgers." Migrantenstudies 12(1): 27-31.

Brühl, R. and S. Buch (2006). Einheitliche Gütekriterien in der empirischen Forschung? - Objektivität, Reliabilität und Validität in der Diskussion. Europäische Wirtschaftshochschule Berlin. ESCP-EAP Working Paper.20

Burgers, J. (1996a). "Natte vingers en vuile handen. Over het schatten van het aantal illegale vreemdelingen." Migrantenstudies 12(1): 14-26.

Burgers, J. (1996b). "Natte vingers en vuile handen. Over het schatten van het aantal illegale vreemdelingen. Een reactie op Böcker en Groenendijk." Migrantenstudies 12(1): 32-33.

Cangiano, A. (2008), Foreign migrants in Southern European countries: evaluation of recent data, in: Raymer, J. and Willekens, F. (2008), International Migration in Europe: Data, Models and Estimates, John Wiley & Sons, Ltd, pp. 89-114

Chelimsky, E. (1991), Politics, Policy Making, Data, and the Homeless, in: Housing Policy Debate, Vol. 2, Issue 3, pp. 683-697

Clarke, J. (2000). The Problems of Evaluating Numbers of Illegal Migrants in the European Union. P. Bruycker. (Ed.). Regularisations of Illegal Immigrants in the European Union. Brussels, Bruylant: 13-22.

Coleman, C. and J. Moynihan (1996). Understanding Crime Data. Haunted by the dark figure. Buckingham, Phil., Open University Press.

Cyrus, N. (2008), German Country Report for CLANDESTINO, final version, August 2008

Committee on Migration, Refugees and Demography "Conference on the situation of illegal migrants in Council of Europe member states, Paris 13 December 2001)

Coste, F. (2008), French modes of migration control, in: J. Doomernik. and M. Jandl, (eds.), (2008), Modes of Migration Regulation, Amsterdam, Amsterdam University Press (IMISCOE Reports), forthcoming

Courau, H. (2008), French Country Report for CLANDESTINO, final version, November 2008

De Bruycker, P. (ed.) (2000) Les régularisations des étrangers illégaux dans l'Union européenne - Regularisations of illegal immigrants in the European Union, Bruylant, Bruxelles

De Haas, H. (2007), Irregular migration from West Africa to North Africa and the European Union, Paper prepared for International Organisation for Migration (IOM), International Migration Institute University of Oxford, Version April 2007

Delaunay, D. and Tapinos, G. (1998), "La mesure de la migration clandestine en Europe." Volume 1: rapport de synthese, Eurostat Working Papers, Series "Population et conditions sociales", 3/1998/E/No7, Luxemburg: Eurostat.

Deutscher Bundestag (2007), Antwort der Bundesregierung auf die Kleine Anfrage der Abgeordneten Ulla Jelpke, Petra Pau, Sevim Dagdelen, weiterer Abgeordneter und der Fraktion DIE LINKE. – Drucksache 16/6832 – Drucksache 16/7089, 11 November 2007

Divinský, B. (2008), Slovak Country Report for CLANDESTINO, final version, September 2008

Doomernik, J. and Jandl, M., (eds.), (2008), Modes of Migration Regulation, Amsterdam, Amsterdam University Press (IMISCOE Reports), forthcoming

Drbohlav, D. and Lachmanová, L. (2008), Czech Country Report for CLANDESTINO, final version, August 2008

Engbersen, G. R. Staring, J. van der Leun, J. de Boom, P. van der Heijden en M. Cruijff (2002). Illegale vreemdelingen in Nederland. Omvang, overkomst, verblijf en uitzetting. Erasmus Universiteit, RISBO, 151 pp.

European Commission (2007): Commission Staff Working document.. Accompanying document to the Proposal for a Directive of the European Parliament and of the Council Providing for Sanctions against Employers of Illegally Staying Third-Country Nationals, Commission of the European Communities, Brussels, 16.5.2007, SEC(2007) 603

Fasani, F. (2008), Italian Country Report for CLANDESTINO, final version, August 2008

Frank and Snijders (1994), Estimating the size of hidden populations using snowball sampling, in: Journal of Official Statistics, 10, pp. 53-67

Futo, P. (2008), Hungarian Country Report for CLANDESTINO, final version, September 2008

Futo, P. and Jandl, M. (eds.) (various years), 200X Year Book on Illegal Migration, Human Smuggling and Trafficking in Central and Eastern Europe, International Centre for Migration Policy Development, Vienna

GéDAP (2004), Migration and Asylum in Europe 2002, Report by the Groupe d'étude de Démographie Appliquée (GéDAP) and the Berlin Institute for Comparative Social Research (BIVS) on behalf of the European Commission, 2004

General Accounting Office (1995), Illegal Immigration: INS Overstay Estimation Methods need Improvement, Washington, D.C., GAO

Gfs.bern (2005), Sans Papier in der Schweiz: Arbeitsmarkt, nicht Asylpolitik ist entscheidend. Schlussbericht im Auftrag des Bundesamtes für Migration, Bern 2005

GHK (2007) Impact Assessment on a Community instrument laying down sanctions for employers of third-country nationals with no or limited rights to work that are exceeded. European Commission Impact assessment

Global Migration Perspectives, No 53, October 2005, Crisis in countries of origin and illegal immigration into Europe via Italy

González Enriquez, C. (2008), Spanish Country Report for CLANDESTINO, final version, September 2008

Heckathorn, D. (1997), Respondent-Driven Sampling: A New Approach to the Study of Hidden Populations, in: Social Research, vol. 44, No.2, May 1997, pp. 174-199

Heckathorn, D. (2002), Respondent-Driven Sampling II: Deriving Valid Population Estimates from Chain-Referral Samples of Hidden Populations, in: Social Problems, vol. 49, No.1, pp.11-34

Heckmann, F. and Wunderlich, T., (2000), Transatlantic Workshop on Human Smuggling. A Conference Report. In: Georgetown Immigration Law Journal, Vol. 15, 2000, pp. 167-182

Hilderink H., N. van der Gaag, L.van Wissen, R. Jennisen, A. Roman, J. Salt, J.Clarke, C. Pinkerton (2003) "Analysis Forecasting of International Migration by Major Groups – Part III" EUROSTAT Working Papers and Studies 3/2002/E/no 17, European Communities, Luxemburg

Hoefer, M., Rytina, N. and Campbell, C. (2007), Estimates of the Unauthorized Immigrant Population Residing in the United States: January 2006, Office of Immigration Statistics, Policy Directorate, U.S. Department of Homeland Security, available at: http://www.dhs.gov/xlibrary/assets/statistics/publications/ill_pe_2006.pdf

Home Office, (2005), Accession Monitoring Report May 2004 – June 2005, August 2005.

ICMPD (1994), The Key to Europe – A Comparative Analysis of Entry and Asylum Policies in Western Countries. Report Prepared by the International Centre for Migration Policy Development for the Swedish Parliamentary Immigrant and Refugee Commission, Stockholm, Swedish Government Official Reports, 1994:135

ICMPD Yearbooks on Illegal Migration, Human Smuggling and Trafficking in Central and Eastern Europe, various years; identical with Futo and Jandl (ed.), various years.

Iglicka, K. and Gmaj, K. (2008, Polish Country Report for CLANDESTINO, final version, August 2008

International Organization for Migration. 1995. "Trafficking and Prostitution: The Growing Exploitation of Women from Central and Eastern Europe," Geneva.

International Organization for Migration (2000) Migrant Trafficking and Human Smuggling in Europe, Geneva.

Jandl, M. (2004). "The Estimation of Illegal Migration in Europe." Studi Emigrazione/ Migration Studies 41(153): 141-155.

Jandl, M. (2007), Proposal to UDI for an alternative estimation of the unauthorized population in Norway, ca. mid-2007, unpublished memo

Jandl, M. and Kraler, A. (2006a), Links Between Legal and Illegal Migration, in: Poulain, M., Perrin, N. and Singleton, A. (2006), THESIM – Towards Harmonized European Statistics on International Migration, UCL Presses, Louvain-La-Neuve 2006, pp. 337-371

Jandl, M. and Kraler, A. (2006b), Statistics on Refusals, Apprehensions and Removals: An Analysis of the CIREFI data, in: Poulain, M., Perrin, N. and Singleton, A., (2006), THESIM – Towards Harmonized European Statistics on International Migration, UCL Presses, Louvain-La-Neuve 2006, pp. 271-285

Jandl, M., (2003), "Estimates on the Numbers of Illegal and Smuggled Immigrants in Europe", Presentation given at the Workshop on Human Smuggling at the 8th International Metropolis Conference, 17. September 2003, available at: http://www.icmpd.org/default.asp?nav=research&folderid=408&id=296

Jandl, M., (2004), The Estimation of Illegal Migration in Europe, Studi Emigrazione/ Migration Studies, vol. XLI, No. 153, March 2004, pp. 141-155

Jandl, M., Hollomey, C. and Stepien, A. (2007), Migration and Irregular Work in Austria: Results of a Delphi-Study, International Centre for Migration Policy Development (ICMPD), Vienna, 2007

Juhasz, J. (1999) Human Smuggling and the Trafficking of Migrants in Hungary, A Country Study for the International Organisation for Migration, Geneva, IOM.

Kennedy, P. (2003). A Guide to Econometrics, Blackwell.

Klinchenko, T. et.al (1999) Country Studies on Migrant Smuggling and Trafficking: the Case of Ukraine, Geneva, IOM.

Koser, K. (2005), Irregular Migration, State Security and Human Security. A paper prepared for the Policy Analysis and Research Programme of the Global Commission on International Migration. Global Commission on International Migration, available at: http://www.gcim.org/attachements/TP5.pdf

Kraler, A., Reichel, D. and Hollomey, C. (2008), Austrian Country Report for CLANDESTINO, final version, August 2008

Laczko F., Amanda Klekowski von Koppenfels and Jana Barthel (2002) Trafficking in Women from Central and Eastern Europe: A Review of Statistical Data, in New Challenges for Migration Policy in Central and Eastern Europe, eds, F. Laczko, I. Stacher and A. Klekowski von Koppenfels, Asser Press, The Hague.

Lamnek, S. (2005). Qualitative Sozialforschung. Lehrbuch. Weinheim, Beltz.

Larson, E. M. and Droitcour, J.A. (2003), Estimating the Illegal Alien Population in the United States: Some Methodological Considerations, United States General Accounting Office, October 2003

Lederer, H. W. (2004), Indikatoren der Migration. Zur Messung des Umfangs und der Arten von Migration in Deutschland unter besonderer Berücksichtigung des Ehegatten- und Familiennachzugs sowie der illegalen Migration, efms, Bamberg 2004

Leerkes, A., Engbersen G. and van San, M. (2007), Shadow Places: Patterns of Spatial Concentration and Incorporation of Irregular Immigrants in the Netherlands, in: Urban Studies, Vol. 44, No. 8, 1–26, July 2007

Levinson, A., (2005) The Regularisation of Unauthorized Migrants: Literature Survey and Country Case Studies, Centre on Migration, Policy and Society (COMPAS), University of Oxford, Oxford

Lewin, C. (2005). Chapter 25: Elementary quantitative methods. B. Somekh and C. Lewin. (Ed.). Research methods in the social sciences. London, Sage: 215-225.

Lianos, T.H., Sarris, A. And Katseli, L. (1996), Illegal Immigration and Local Labour Markets: The case of Northern Greece, in: International Migration, vol. XXXIV, No.3, pp. 449-484

Lincoln, Y. S. and E. G. Guba (1985). Naturalistic Inquiry. Beverly Hills, Ca., Sage.

Linstone, H. and Turoff, M. (eds.), (1975), The Delphi Method- Techniques and Applications. Addison- Wesley Publishing Company. Reading, Massachusetts

Malynovska, O. (2004), International migration in contemporary Ukraine: trends and policy, in: Global Migration Perspectives, No.14, October 2004, available at: http://www.gcim.org/attachements/GMP%20No%2014.pdf

Maroukis, T. (2008), Greek Country Report for CLANDESTINO, final version, August 2008

Masser, I. and Foley, P. (1987), Delphi Revisited: Expert Opinion in Urban Analyses, in: Urban Studies 24, pp. 217-225

Massey, D. and Singer, A. (1995) "New Estimates of undocumented Mexican Migration and the Probability of Apprehension". Demography 32(2), pp. 203-213.

Mayring, P. (2002). Einführung in die qualitative Sozialforschung. Weinheim, Basel, Beltz.

Morita, Kiriro and Sassen, Saskia (1994), The New Illegal Immigration in Japan, 1980-1992, in: International Migration Review, No.1, vol.28, p.153-163

Natale, M. (1990), L'immigrazione straniera in Italia: consistenza, caratteristiche, prospettive, in: Polis, Vol. IV, No 1, Avril, pp. 5-40

Natale, M. (1998), Determining irregular foreigners in the Italian population, in: Delaunay, D. and Tapinos, G. (1998b), "La mesure de la migration clandestine en Europe." Volume 2: rapport des experts, Eurostat Working Papers, Series "Population et conditions sociales", 3/1998/E/No7

OECD (2004), SOPEMI Annual Report 2003, Trends in International Migration, OECD Publications, Paris

OECD (2006), International Migration Outlook: SOPEMI 2006 Edition,

OECD (2007), International Migration Outlook. Annual Report 2007, Organisation for Economic Co-operation and Development, Paris.

Papademetriou, D.G. (2005a) The Global Struggle with Illegal Migration: no end in sight, September. Migration Policy Institute www.migrationinformation.org

Papademetriou, D.G. (2005b) The "Regularisation" Option in Managing Illegal Migration More Effectively: A Comparative Perspective, Migration Policy Institute, September 2005 No.4

Passel, J. (2002), New Estimates of the Undocumented Population in the United States, May 22, 2002, available at: http://www.migrationinformation.com/USfocus/display.cfm?ID=19 (11.3.2003)

Passel, J. (2006), The Size and Characteristics of the Unauthorized Migrant Population in the U.S.: Estimates based on the March 2005 Current Population Survey, Pew Hispanic Centre, available at: http://pewhispanic.org/files/reports/61.pdf

Pastore, F. (2008), Italian Modes of Migration Regulation, in: Doomernik, J. and Jandl, M., (eds.), (2008), Modes of Migration Regulation, in: IMISCOE Research Series, forthcoming

Petersen, C. (1896), The Yearly Immigration of Young Plaice into the Limfjord from the German Sea, in: Report of the Danish Biological Station to the Ministry of Fisheries, Bd.6, S.1-48)

Piguet, Etienne and Losa, Stefano (2001): Travailleurs de l'ombre ? Demande de main-d'œuvre du domaine de l'asile et ampleur de l'emploi d'étrangers non déclarés en Suisse, Rapport de recherché, Swiss Forum for Migration Studies, Etude mandatée par l'Office fédéral des réfugiés, 19/2001

Pinkerton, C.G., McLaughlan and Salt, J. (2004), Sizing the illegally resident population in the UK, Home Office Online Report 58/04, available at: http://www.homeoffice.gov.uk/rds/pdfs04/rdsolr5804.pdf

Robinson, J.G. (1980), Estimating the approximate size of the illegal alien population in the United States by the comparative trend analysis of age-specific death rates, in: Demography, vol. 17, No 2, March, pp. 159-176

Salganik, M. J. and D. D. Heckathorn (2004). "Sampling and Estimation in Hidden Populations Using Respondent-Driven Sampling." Sociological Methodology 34: 193-239.

Salt J. and J. Hogarth (2000) Trafficking and human Smuggling in Europe: a review of the evidence, Geneva, IOM.

Salt, J. (1999) Current Trends in International Migration in Europe, Paris: Council of Europe.

Salt, J. A. Singleton and J. Hogarth (2001) Europe's International Migrants. Data Sources, Patterns and Trends, London HMSO.

Schneider, F. (2002) Die Entwicklung der Schattenwirtschaft (Pfusch) in Österreich und in den Bundesländern Niederösterreich, Oberösterreich, Steiermark und Tirol in den Jahren 1990 bis 2002 (The development of the informal sector in Austria and in the federal provinces of Lower Austria, Upper Austria, Styria and the Tyrol from 1990 to 2002), Linz

Schneider, F. and Enste, D., (1999), Shadow Economies around the World: Size, Causes, and Consequences, IMF Working Paper WP/00/26, 57p., available at: http://www.imf.org/external/pubs/ft/wp/2000/wp0026.pdf

Schneider, F. and Enste, D., (2000), Schattenwirtschaft und Schwarzarbeit: Umfang, Ursachen, Wirkungen und wirtschaftspolitische Empfehlungen (Shadow economy and black labour: extent, reasons, impact and economic policy implications), Forum Wirtschaft und Soziales (FWS), R. Oldenbourg Verlag, München-Wien, 2000

Schneider, F. and Klinglmair, R., (2004), Shadow Economies around the World: What do we know, Working Paper 0403, Department of Economics, University of Linz, Linz 2004

Schneider, F., (2005), Noch rückläufige Schattenwirtschaft in Deutschland – Fluch oder Segen?, Universität Linz, Linz 2005

Schneider, F., (2006), Nur noch leicht sinkende Schattenwirtschaft in Deutschland im Gegensatz zu anderen OECD-Ländern im Jahr 2006: Fluch oder Segen?, Universität Linz, Linz 2006

Schoorl J., Heering L., Esveldt I., Groenewold G., Erf R. et al. (2000), Push and Pull Factors of International Migration: A Comparative Report, Luxembourg: Eurostat, European Communities

Sciortino, G. (2003), Regularization of Foreigners in Italy, unpublished manuscript, Communication to Michael Jandl on 3.4.2003

Sciortino, G. (2004), 'Between phantoms and necessary evils. Some critical points in the study of irregular migrations to Western Europe', IMIS-Beiträge 24: 17–43.

Seale, C. (1999) The quality of qualitative research. London: Sage

Seale, C. (2004). Quality in qualitative research. C. Seale, G. Gobo, J. F. Gubrum and D. Silverman. (Ed.). Qualitative Research Practice. London, Sage: 409-419.

Sheldon, G. (2002), Forschungsdesign zur Bestimmung des Ausmasses und der Struktur der Schwarzarbeit in der Schweiz, Bericht zu einem Forschungsauftrag des Bundesamtes für Ausländerfragen, des Bundesamtes für Flüchtlinge und des Staatssekretariats für Wirtschaft, Forschungstelle für Arbeitsmarkt- und Industrieökonomik, Universität Basel, Basel 2002

Sikkel, D., van der Heijden, P. and van Gils, G. (2006), Summary, in: Methoden voor omvangschattingen van verborgen populaties, met name illegalen, WODC rapport 243, Meppel: Boom Juridische Uitgevers

Stobbe, H. (2004), Undokumentierte Migration in Deutschland und den Vereinigten Staaten. Interne Migrationskontrollen und die Handlungsspielräume von Sans Papiers, Göttingen 2004

Tapinos, G. (1999). Clandestine Immigration: Economic and Political Issues. Organization for Economic Co-operation and Development. (Ed.). Trends in International Migration. Sopemi 1999. Paris, OECD: 229-251.

Teddlie, C. and A. Tashakkori (2003). Major issues and controversies in the use of mixed methods in the social and behaviora sciences. A. Tashakkori and C. Teddlie. (Ed.). Handbook of Mixed Methods in Social and Behavioral Research. London, Sage: 3-50.

The "Regularisation" Option in Managing Illegal Migration More Effectively: A Comparative Perspective, September 2005 No.4

Thierry, X. and Rogers, G.I. (2004), Recent Immigration Trends in France and Elements for a Comparison with the United Kingdom, in: Population, Vol. 59, No.5 (Sep.-Oct. 2004), pp.635-672

UN Statistical Commission, UN Economic Commission for Europe and Eurostat, (2003), Statistical data on illegal immigration in the European Union: A discussion paper on policy needs and data availability, submitted by the European Commission, paper prepared by Ann Singleton for the Joint ECE-Eurostat Work Session on Migration Statistics, Geneva, 28-30 April 2003

United Nations Office on Drugs and Crime, Methodology, in: World Drug Report 2007, Vienna, p.257-274

Van der Heijden, P. (2006), Estimating the size of the illegal population in the Netherlands, Graduate School Lecture, University of Utrecht, 9 October 2006

Van der Heijden, P., Bustami, R., Cruijff, M. , Engbersen, G. and van Houwelingen, H. (2003), Point and interval estimation of the truncated Poisson regression model, in: Statistical Modelling, 3, pp. 305-322.

Van der Heijden, P., Cruijff, M. and van Houwelingen, H. (2003), Estimating the size of a criminal population from police registrations using the truncated Poisson regression model, in: Statistica Neerlandica, 57, pp. 289-304.

Van der Heijden, P., van Gils, G., Cruijff, M. and Hessen, D. (2006), Summary of 'An estimate of the number of illegal aliens residing in the Netherlands in 2005', October 2006

Van der Leun, J. and Ilies, M. (2008), Dutch Country Report for CLANDESTINO, final version, August 2008

Van der Leun, J. van der, G. Engbersen en P. van der Heijden (1998): Illegaliteit en criminaliteit: schattingen, aanhoudingen en uitzettingen. Onderzoeksrapport, Faculteit der Sociale Wetenschappen/Sociologie, Erasmus Universiteit, Rotterdam.

Van der Leun, J., G. Engbersen en Peter van der Heijden (1998). Illegaliteit en criminaliteit: schattingen, aanhoudingen en uitzettingen. Rotterdam: Erasmusuniversiteit

Van Meeteren, M., van San, M. and Engbersen G. (2007), Irreguliere immigranten in Belgie. Inbedding, uitsluiting en criminaliteit, Erasmus Universiteit Rotterdam, July 2007

Vogel D. (1999), Illegaler Aufenthalt in Deutschland – methodische Überlegungen zur Datennutzung und Datenerhebung. Zeitschrift für Bevölkerungswissenschaft, Jg. 24, 2/1999, pp. 165-185

Vogel, D. (2002): Ausländer ohne Aufenthaltsstatus in Deutschland. Methoden zur Schätzung ihrer Zahl. In: Minthe, Eric (ed.): Illegale Migration und Schleusungskriminalität. Wiesbaden, Eigenverlag der Kriminologischen Zentralstelle e.V., 2002, pp. 65-78.

Vogel, D. and V. Kovacheva (2008). Classification report – Quality assessment of estimates of stocks of irregular migration. Hamburg Institute of International Economics (HWWI). irregular-migration.hwwi.net

Wanner, Ph. (2002), Compter les clandestins. Méthodes d'estimation de la population sans autorisation de séjour à partir des données sur la population, Swiss Forum for Migration Studies, SFM Discussion Paper 13/2002, Neuchatel, 26.4.2002

Weil, P. (2005), La République et sa diversité. Immigration, intégration, discriminations, La République des Idées-Seuil, 2005

Widgren, J. (1994) Multilateral Co-operation to Combat Trafficking in Migrants and the Role of International Organisations, Paper Submitted to the 11th IOM Seminar on Migration, October, Geneva.

Widgren, J. (2002), Communication to J. Kempton: Question re Illegal Migrants, 18.9.2002

Woodbridge, J. (2005), Sizing the unauthorised (illegal) migrant population in the United Kingdom in 2001, London: Home Office.

Zhang, L. (2008), Report on: Developing Methods for Determining the Number of Unauthorized Foreigners in Norway, Statistics Norway, Document 2008/11, Oslo, June 2008