Compilers' Manual on Forced Displacement Statistics

USE CASE

D

USING GOVERNMENT ADMINISTRATIVE DATA

PHILIPPINES.
2021 birth registration in Kasanyangan, Zamboanga City.
©UNHCR/Martin San Diego

EGRISS

This is Use Case D from the *Compiler's Manual on Forced Displacement Statistics*. The Use Case describes how government administrative data sources can be used in the production of official statistics on displaced populations.

The *Compiler's Manual* is aimed primarily at technical personnel in National Statistical Systems who want to include displaced populations – refugees and / or Internally Displaced Persons (IDPs) – in official statistics. Each Use Case discusses a different scenario relevant to producing official statistics on refugees and IDPs, with a focus on the elements of statistical production cycles that are specific to refugee and IDP contexts. Spotlight examples of good practice in the production of refugee and IDP statistics are interwoven throughout the Use Case.

The Compiler's Manual and its individual Use Cases are intended to be a 'living document' which will be amended and extended as the body of expertise and knowledge develops worldwide.

Note: Paragraph numbering is per the complete version of the Compilers Manual.

The Expert Group on Refugee, IDP and Statelessness Statistics

The Expert Group on Refugee, IDP and Statelessness Statistics is a UN Statistical Commission mandated, multi-stakeholder group that works with National Statistical Offices, international organizations and civil society to develop and support implementation of international standards and guidance to improve official statistics on forcibly displaced and stateless persons.

The Compilers' Manual

The Compilers' Manual offers clear operational instructions on producing official statistics on refugees, asylum seekers, IDPs and related populations. It complements the content of the International Recommendations on Refugee Statistics and the International Recommendations on Internally Displaced Persons Statistics by providing hands on guidance.

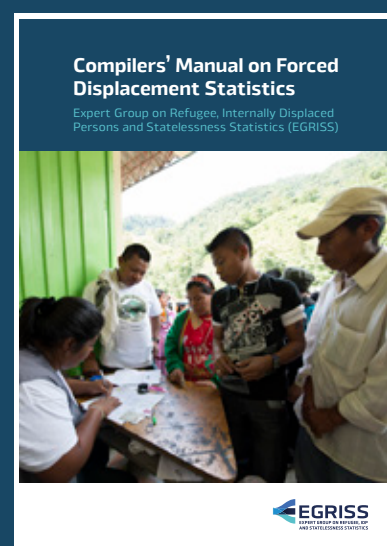## Access the complete version of the Compilers' Manual'

Compilers' Manual on Forced Displacement Statistics

Expert Group on Refugee, Internally Displaced Persons and Statelessness Statistics (EGRISS)

EGRISS
EXPERT GROUP ON REFUGEE, IDP AND STATELESSNESS STATISTICS

TABLE OF CONTENTS

Administrative data – general background

156. There are many potential sources of data on displaced people other than surveys and censuses. This section focusses on administrative data sources and covers the issues relevant to the production of statistics on both refugees and IDPs. Eurostat defines administrative data as "the set of units and data derived from an Administrative source. This record keeping can be done by institutions belonging to the government sector or by private organisations"[1].

157. Administrative data are a useful source of information for measuring both the stock of displaced populations and the flows, showing the direction of displacement trends, and potentially for developing sampling frames. If they include variables allowing the identification of the target population, or if the specific administrative source can be linked at the individual level to another data source such as another register, a survey or a census, then they have potential to be part of a rich source of regular official statistics, including socio-economic indicators.

158. The priorities when considering using administrative data as a source of information on refugees and IDPs are:

- Identify the different available administrative sources and understand how they cover displaced people.
- Establish what the access requirements are, including the legal framework for accessing any sources that are potentially useful.
- Investigate the options for linking data sources to provide greater insights.
- Investigate and understand the limitations of any sources that will be used and build links with data owners to address the limitations wherever possible.

159. Administrative data have the advantage of being readily available at no additional cost and often more timely (up to date) than survey data. This can be particularly important for refugees and IDPs, as populations can change rapidly. However, because administrative data are not collected with statistical purposes in mind it is important to spend some time understanding the data processes to be aware of any limitations, such as definitional differences or data quality issues. Any limitations relating to the use of the statistics should be conveyed at the point of dissemination.

160. Statisticians should seek to build and maintain strong links with the organisations that own administrative registers, to raise awareness of the statistical uses of the data and foster collaboration: it may be possible to enhance the statistical value of administrative data through small changes to the underlying register.

---

1 Eurostat ESS Vision 2020 ADMIN (Administrative data sources).

❘ Potential sources of administrative data

161. There are many administrative data sources that can provide relevant statistics on refugees and IDPs. These include:

- **General population registers** maintained by governments, covering the total resident population of a country.
- **Other government administrative registers for the wider population** including residence permit registers, tax registers, social security registers, register of border crossings, work permits, health registers, education registers, housing registers and land ownership registers[2].
- **Specific refugee / IDP registers maintained by government administrations** or organisations responsible for refugees, asylum seekers and IDPs, including registration of displaced people in camps and/or elsewhere, asylum seekers registers (applications), registers on decisions granting international protection, resettled persons register.

162. International agencies' databases (for example, registration systems, administration of cash assistance, etc.) are also administrative registers by their statistical nature, but because they are not government-owned and come with very different access requirements they are discussed separately in Use Case E and referred to in the Compilers' Manual as "operational data".

General population registers

163. A population register is a regularly updated administrative database of uniquely identifiable individuals who are usually resident in a country. In countries that maintain a population register, the register is the basis for population statistics, particularly on stocks and flows and can support a wide range of other statistics, through data linking which is facilitated by the unique identifiers in the dataset.

164. Refugees and related populations may be included and identifiable in a population register, but it is more common that asylum seekers are not treated as part of the usually resident population and are therefore omitted. In this case, identification would have to be achieved through data matching with other sources, for example administrative records of asylum applications or residence permit registers. See further information on data linking (paragraph 174). IDPs would usually be included in a population register but may not be identifiable as IDPs without linking to another source. Access requirements vary, but often access is permitted for statistical purposes, subject to data confidentiality agreements.

---

2 The United Nations' *Recommendations on Statistics on International Migration Revision 1* (1998) provides useful guidance for utilising administrative data sources for migration statistics in general, and refugee statistics in particular. These recommendations can be adapted for IDP registers, however IDPs are often not included.

NORWAY

📍 CASE STUDY: USING THE CENTRAL POPULATION REGISTER TO PRODUCE REFUGEE STATISTICS IN NORWAY

Statistics on those with a refugee background in Norway are produced from the Central Population Register (CPR). The CPR of Norway was established in 1964 based on the 1960 Population Census. A unique 11-digit personal identification number (PIN) was introduced at the same time. The CPR includes all persons who have ever been a (legal) resident of Norway since 1960, regardless of their citizenship. It provides information on their name, address, residence status, place or country of birth, citizenship, reason for immigration, country of immigration or emigration, marital status and PIN numbers of spouse, mother and father. Persons who die or emigrate are not deleted from the register, but a code for their status is changed. All vital events and migrations and address changes are registered in the CPR.

Data on asylum seekers is transferred to the CPR from the Norwegian Directorate of Immigration (UDI). All asylum seekers must register with the Norwegian immigration authorities, and their case data and personal data are stored in a database that contains data on UN convention refugees as well as refugees who enter Norway on their own and apply for asylum.

Data are transferred regularly from the CPR to Statistics Norway, where it is linked with processed data from UDI enabling Statistics Norway to produce statistics on stocks and flows of people with a refugee background.

Government administrative registers for the wider population

165. There are many types of government administrative registers, other than population registers, that have potential to be a source of data on refugees and IDPs. The strength of these additional sources is often as a source of statistics on sub-categories of refugees and IDPs, rather than contributing to estimates of stocks and flows for the wider refugee / IDP population. For example, tax registers may provide information on refugees who pay tax; health registers may be a source of statistics on those who have engaged with the health system; and education registers may enable comparisons between children of refugees or IDPs and the wider population.

166. The extent to which refugees and IDPs are identifiable in government administrative registers will vary considerably[3] and will often rely on being able to link to other data sources, for example on asylum applications. As with population registers, access requirements vary, but often access is permitted for statistical purposes, subject to data confidentiality agreements.

167. A key limitation of government administrative registers can be the ability to identify unique individuals, rather than events. For example, registers of border crossings may include multiple records for people who have been returned to their country of origin and subsequently re-crossed the border; similarly, registers of work permits may include multiple records for individuals if people are required to renew permits. If the data source includes unique identifiers for individuals then such issues can be avoided.

Specific refugee / IDP registers maintained by government administrations

168. Many countries use administrative systems to co-ordinate the asylum application and decision-making process or to register displaced people in camps or other locations. Where this is done by government administrations – or on their behalf by other organisations – the data can be an important source of statistics on refugee / IDP stocks and flows.

169. Administrative registers are the main data source for statistics on asylum applicants and decisions taken during the asylum application process in regions such as Europe and Northern America. In the EU and in some additional neighbouring countries, they are also the main source of statistics on the number of refugees living in the country, through the residence permits statistics (which include categories such as refugees legal status, subsidiary protection, humanitarian reasons).

170. The use of "registers of asylum seekers" is specifically recommended for estimating the stock of asylum seekers (by length of stay in the country) and the number of new asylum applications within a given year (United Nations, 1998, p. 19), and for producing statistics on the outcome of the asylum procedure.

171. By their nature, these administrative systems will have good coverage of refugees or IDPs within their scope – but (particularly for camp-based registers) this may not be the entire stock, for example it may not include refugees or IDPs outside camps. If contributing to a full picture of stocks and flows for the entire country, care should be taken to understand any potential overlap or duplication of coverage between administrative registers.

---

3   For example, this UNESCO paper discusses the challenges identifying refugees in Education Management Information Systems.

BOSNIA AND HERZEGOVINA

CASE STUDY: PRODUCING STATISTICS ON FORCIBLY DISPLACED PEOPLE IN BOSNIA AND HERZEGOVINA

Statistics on IDPs and refugees in B&H are produced by a member of the National Statistical System, the Ministry of Human Rights and Refugees. This is a state level organisation responsible for unifying the data on displaced persons which is collected by municipalities in all three administrative units of B&H, and for producing statistical reports.

Producing statistics on refugees and IDPs was initiated by the government at the beginning of the war in the 1990s. When the war ended the first comprehensive official registration of persons in B&H was carried out in late 2000 by Ministries and local authorities. The registration and estimation of the number and the status of displaced persons was first completed in 2005.

The administrative records in the Database of Displaced Persons (DDPR) are now the main source of statistics on refugees and IDPs. This key data source establishes someone's status as IDP or refugee. The database is supplemented by census data. The questionnaire from the census in 2013 asks if a person had been a refugee after 1991, or if the person has returned from 'refuge'. It asks for details of the settlement that the person was displaced from; whether the person has returned; and if the person is still legally considered to be a displaced person.

Summary of key features of administrative systems and their limitations

172. Table D.1 illustrates the types of data and analysis that different sources of administrative data may support, if they identify refugees and IDPs or they can be linked with a source that enables such identification. As well as stocks and flows, administrative data have potential to support analysis of refugees needs and integration and IDPs progress toward durable solutions, as described by the indicators referenced in Use Cases B and C.

173. It is important to note that the identification of refugees and IDPs in administrative sources (and other sources which may be considered for data linking) may not be equivalent to identification following a set of survey questions, for example. There can be important definitional differences which need to be considered, particularly if using non-traditional sources as a source of statistics for comparison or to augment survey results: the two may not be covering the same population.

Table D.1: Analysis supported by administrative sources and their key limitations

Source	Stocks and flows analysis supported	Socio-economic analysis supported	Completeness and other limitations
Population register; register of foreigners	- Key source of statistics on stocks and flows - Basic characteristics, for example age, gender - Geospatial analysis, usual residence and previous residence	n/a	- May not routinely identify refugees or IDPs. - Need to ensure statistical laws allow access - Frequency of update.
Refugee / IDP register	- Key source of statistics on stocks and flows - Basic characteristics, for example age, gender - Geospatial analysis, usual residence and previous residence		- Should be complete, but consider usual quality concerns, presence of duplicates, time lag in registering, update frequency.
Asylum application register	- Key source of statistics on stocks and flows - Basic characteristics, for example age, gender - Geospatial analysis, usual residence and previous residence	n/a	- Covers asylum seekers only, not IDPs. - Won't cover refugee-related populations. - Refugees granted asylum before establishment of register won't be identified.
Residence permit register	- Key source of statistics on stocks and flows - Basic characteristics, for example age, gender - Geospatial analysis, usual residence and previous residence	n/a	- May not accurately identify sub categories of those requiring residence permits. - Possibility of duplication
Work permit register	- Stocks of refugees / IDPs with work permits	- Legal access to labour market	- Only captures those of working age. - Only captures those within the formal working environment.
Tax register, Labour Market Information System	- Stocks of refugees / IDPs registered for tax or employment.	- Tax details - Analysis of employment status and income status, occupation, and comparisons with wider population	- Only captures subset of refugees / IDPs who pay tax

Social security register	- Stock of refugees / IDPs in receipt of social security, as a proportion of total stock of refugees / IDPs and comparison with wider population	- Poverty and deprivation	
Border crossings register	- Flows of refugees and asylum seekers. - Contributes to estimates of total stocks.		- Duplication from individuals crossing borders multiple times. In flows likely to be more reliable than out flows. - May not identify refugee / IDP status of individuals, or accuracy of identification may be limited.
Health register	n/a	- Prevalence of health issues amongst refugee / IDP population. - Immunisation / vaccine coverage - Potential contribution to wider indicators of health status of refugees / IDPs	- Only covers those who engage with official health system. - Identification of refugee / IDP status within register is likely to be a limiting factor.
Education register	- Stocks of refugee / IDP school age children	- Educational attainment, literacy - Participation in education - Participation in preschool education and access to childcare	- Only covers those of school age, missing those who do not attend school. May not cover those who attend a non-government school.
Other sources, for example electricity subscribers	n/a	Source-specific data to link across sources to generate new insights. For example, the extent of electricity subscriptions in locations predominantly refugee / IDP might contribute to understanding of poverty and housing conditions.	

Data linking

174. There is no single source of data that can draw a complete statistical picture of refugees and IDPs. Because of the variety and the varying coverage of the possible sources, data linking (also called data integration) is a promising solution for improving the quality of refugee and IDPs statistics. Recent projects like the UNECE High Level Group for the Modernisation of Official Statistics (HLG-MOS) or the Eurostat European Statistical System 2020 vision project in the area of data integration illustrate the growing interest in data linking.

175. Data linking can address several challenges when compiling official statistics on refugees and IDPs by:

- Improving the coverage of target groups;
- Improving data availability, for example by adding variables especially socio-economic variables related to refugee and IDP populations;
- Improving timeliness, by linking survey data with regularly updated administrative data;
- Improving data quality and accuracy, for example by cross-referencing different sources in order to assess refugee and IDP status, or by deriving a more complete sampling frame.

176. Data linking is not limited to linking across administrative systems. The census, surveys of the general population or refugees / IDPs, and non-traditional data sources all have potential to add value through data linking with administrative data (or indeed with each other) for example through probabilistic data matching. This section provides some high-level key messages on data linking which can help and guide practitioners in NSOs interested in applying data linking to refugee and IDP statistics. The technical details of data linking are not specific to refugees and IDPs, so are not included here, but the use of data linking is illustrated through case studies of good practices in NSOs.

Unique identifiers

177. The key to successfully linking data sources (and the biggest challenge) is the availability of – or possibility of building – unique identifiers (UID) for statistical units, common across all the data sources involved. Examples of sources that may yield UIDs for refugees and IDPs include:

- Population register number.
- ID-Card or passport number for IDPs.
- Residence permit number for refugees.
- Tax or social security numbers.
- Operational data: UNHCR asylum seeker and refugee certificates registration number.

178. The following case studies illustrate different approaches to data linking, starting with a population register that identifies refugees / IDPs, compared with using a specific IDP register as a starting point.

NORWAY

CASE STUDY: DATA LINKING BASED ON A POPULATION REGISTER THAT IDENTIFIES REFUGEES IN NORWAY

Statistics on those with a refugee background in Norway are produced from the Central Population Register (CPR). Persons who have come to Norway as refugees or asylum seekers (as well as their family members) are included in the CPR once they have been granted permission to be residents – which can be temporary or permanent. Refugees are given a PIN on arrival, whereas asylum seekers are assigned another ID number, called the D-number, when they apply for asylum. It is only when a person is granted asylum that he or she is assigned a PIN number.

Statistics Norway receives regular updates on vital events and other changes in the CPR and has access to many other administrative and statistical data sources that use the same PIN. This makes it possible to link information from different administrative sources to the CPR, to check for possible errors and other shortcomings, such as missing data, and to generate new data, such as country of birth (if this is not in the CPR). The contents, coverage and quality of the administrative registers have become so good that it is no longer necessary to conduct traditional population and housing censuses. Statistics Norway publishes regular analysis, including on persons with a refugee background and on refugees in the labour market.

COLOMBIA.
Embera indigenous communities of Bogota return to Alto Andágueda in 2023.
©Unit for Comprehensive Care and Reparation for Victims of Colombia (UARIV)

COLOMBIA

📍 CASE STUDY: USING AN IDP ADMINISTRATIVE REGISTER AS A STARTING POINT FOR DATA LINKING IN COLOMBIA

The official source to count the stock of IDPs in Colombia is the Single Victims' Registry managed by the Victims' Unit, a government agency responsible for providing assistance and reparation to victims of the armed conflict. The Single Victims' Registry has been counting IDPs since 1985. As at October 2019, 7.5 million IDPs were registered. IDP households who wish to be included in the Single Victims' Registry present a declaration before the Public Ministry (Ombudsman's Regional Office, Procurator's Municipal Office). According to the national legislation an IDP must declare within two years of the displacement event. Once an IDP is granted the status of victim, they are included in the Single Victims' Registry. Once registered they are entitled to access relevant assistance and reparation.

An individual assessment of vulnerability is run twice a year for IDPs registered in the Single Victims' Registry. The data used to run the assessment comes from both primary and secondary sources. Primary data are collected through a survey conducted by phone on a continuous basis. Secondary data, which is the main source of data, involves the exchange of official administrative records with other government agencies at both national and local level: for example the official education enrolment system, the social security system, the registers' office, and the housing and agricultural subsidies registers, among others. IDPs update their location directly to the Victims' Unit through services points located throughout the country.

There is a Protocol for Information Exchange between the Victims' Unit and all the government agencies which form part of the National System for the Assistance and Reparation of Victims. The process of information exchange is formalised through a Memorandum of Understanding, in compliance with the relevant legislation. The data themselves are linked through unique identifiers based on the national ID numbers allocated to all citizens.

Various statistics can be calculated out of the Single Victims' Registry: these are disaggregation by sex, age, ethnicity, type of human right violation (displacement, forced disappearance, homicide, threat, among others), geographic location or number of victims per year.

179. In practice, finding sources with UIDs can be very challenging, as most administrative systems evolve to serve a specific purpose without any thought to future data linking. For example, administrative registers may treat whole families as a single unit. In the absence of a suitable UID, it can be possible to build a synthetic UID through data matching, including imputation or probabilistic data matching. These methods can enable insights from combining survey sources with administrative data – see Case Study *Data Matching in Germany*. A key first step in data linking is to review the widest range of possible data sources, including surveys and non-traditional sources, giving consideration to the availability and quality of the sources and defining the possible outputs that could be derived through linking.

180. Estimating the unknown size of a target population when using multiple partially overlapping sources might also be attempted using the capture-recapture method. Application of this method to the refugee / IDP context is included in Annex 2.

⊙ CASE STUDY: DATA MATCHING IN GERMANY

Germany uses administrative data on residence permits from the Central Register of Foreigners (CRF) to identify refugees and foreigners in refugee like situations. The CRF does not provide information on their socioeconomic characteristics or housing and living conditions. Germany's Labour Force Survey (LFS) provides a detailed socioeconomic picture of respondents but does not confidently identify refugees and foreigners in refugee-like situations.

Using imputation-based data matching, the information from the administrative data set (donor) is used in order to impute refugee status in the LFS survey data set (recipient). This imputation is based on a set of common variables on citizenship, date of entry, age and gender. The final aim is the analysis of the unobserved joint distributions of refugee status and socioeconomic characteristics.

More details on this Case Study are included in Annex 1.

Access requirements

181. Some countries provide a legal basis for the NSO to access data and metadata for all available administrative data sources for evaluating their potential use in official statistics. If, based on the metadata, the data source is judged to be useful to build new or improve existing official statistics, then test data can be requested to further assess data quality and to produce prototype statistics. For other (non-administrative) data sources the access requirements are different. Table D.2 illustrates the likely access requirements for key data sources that may yield improved refugee / IDP statistics through data linking.

Table D.2: Likely access requirements for different data sources

Sources/ Access requirements	Legal environment	Availability	Sustainability	Statistical requirements	Required resources (cost, IT environment, technical skills)
Administrative registers	MOUs Statistical law	Micro/macro data	Changes in collected data: Consultation of statistical authorities by data owners. Cooperation between stakeholders	Confidentiality Statistical unit Statistical concept vs administrative	Compatibility between IT systems Data modelling
Censuses	Statistical law	Micro/macro data	Depends on statistical authorities	Low Frequency UID	Costly
General population surveys	Statistical law	Micro/macro data	Depends on statistical authorities	Frequency UID Size of the sample of the population of interest	Compatibility between IT stems Data modelling
Specific surveys[4] (Refugees, IDPs)	Refugee law Data protection policy & consent	Micro/macro data		Event driven Coverage	Data modelling
Operational data	Data protection policy & consent	Micro/macro data	Depends on data collector	Event driven Coverage	Data matching
New digital sources (social media, satellite images, mobile phones, GIS)	Memorandum of Understanding (MOUs) Commercial agreement Statistical law	Often unstructured data e.g. Tweets, Google	Problem of market entries and exits of data providers	Problem of biases in usage of digital devices and media; frame for grossing-up required to remove biases (e.g. census data) or other data selectivity models	Data scientists; Problem of competition between NSOs and businesses for skilled employees

Standardizing refugee and IDP identification in administrative systems

182. It would be a significant benefit for refugee / IDP statistics if there were some recommended standards for including refugees and IDPs in administrative systems. EGRISS will continue to work towards standards and examples of good practice for identifying refugees in administrative data systems to include in future updates of this manual.

---

4 It should be noted that the legal environment will be significantly different depending on whether the specific surveys are managed by the NSI or by other ministries, academic institutions or other Non Government Organisations.

Data quality

183. Administrative data are usually not collected with statistical purposes in mind so it is important to spend some time understanding the data processes to be aware of limitations and potential data quality issues which can affect their use. Data quality can vary significantly between administrative registers and can be as fundamental as differences in the definitions or coverage of refugees and IDPs. Removal of obsolete records is another example of a significant data quality issue common across administrative data. However, in general these data quality concerns are not unique to refugee and IDP data sources and the same processes should be followed as for any other statistical use of administrative data.

184. Statisticians should seek to build and maintain strong links with the organisations that own administrative registers, to raise awareness of the statistical uses of the data and foster collaboration: it may be possible to address data quality concerns or enhance the statistical value of administrative data through small changes to the underlying register.

185. Reviewing data quality after a mass influx of asylum seekers requires special attention. Data quality issues in administrative data occur in particular during periods of mass influx when administrative procedures are overburdened with the large numbers of incoming asylum-seekers. When the increase in asylum applications is sudden and unexpected, it is difficult to scale-up existing administrative capacities for registration and refugee status determination. As a result, data quality in administrative registers is often poorest when the demand for data on refugees is highest. Reviewing and validating the available data becomes a crucial task for statisticians in those times, since they must judge whether data quality is sufficient for the production of official statistics. When using administrative data in immediate response to a mass influx, statisticians should look out for:

- **Delayed registrations:** In response to humanitarian crises, organizing shelter and covering basic needs is a host country's first priority. Refugee status determination comes afterwards. During periods of mass influx, asylum seekers may face prolonged waiting periods for officially lodging their asylum applications. Where asylum seekers are properly registered only after having lodged their asylum application, administrative registers will cover the newly arrived only with considerable time lag.

- **Incomplete and false registrations:** Intending to assist overburdened immigration authorities, governments improvise and deviate from traditional workflows during periods of mass influx. Registration is then often no longer carried out by trained personnel in immigration offices but mobile registration units with the support of the armed forces and the police register asylum seekers in temporary accommodations or even in improvised reception facilities on the road. Considering that untrained staff registers displaced people who often lack official documents, incomplete and false registrations are hardly surprising.

- **Multiple registrations:** When asylum seekers are distributed within the host country for accommodation, small differences in spelling or translations often cause double registration. Language barriers are especially problematic when translations between different alphabets are involved e.g. between the Arabic, Cyrillic and Latin alphabet. While taking and comparing fingerprint scans would solve most of the problem, the infrastructure is not always in place. Once entered in the register, duplicate entries are difficult to identify with certainty and can thus be expected to bias results for a prolonged period.