

# The Dark Side of Numbers: Updated

*William Seltzer*

The purpose of this chapter is to update and summarize the research results presented in Seltzer and Anderson.<sup>1</sup> It is based initially on a presentation to the conference, *Bevölkerungsforschung und Politik in Deutschland im 20. Jahrhundert*, held in 2003 in Berlin under the auspices of the German Association for Demography (DGD) and the Technische Universität Berlin, which itself drew on a paper by Seltzer and Anderson.<sup>2</sup> In addition, in many places the present chapter also draws directly on a subsequent paper, “On the Use of Population Data Systems to Target Vulnerable Population Subgroups for Human Rights Abuses”.<sup>3</sup>

All these works report on continuing research on how population data systems, particularly population registration systems and population censuses, have been misused in different countries to target vulnerable population subgroups for human rights abuses. On occasion, these abuses have included such internationally recognized crimes as genocide, crimes against humanity, and forced migration. The targeted groups have been defined in terms of race/ethnicity, national origin, mother tongue, and social class.

In addressing this serious topic, it should be stressed that most population data collection efforts are not associated with such targeting and misuse. Indeed, national population data systems are often the only source of reliable data needed to plan and monitor development efforts in many fields.<sup>4</sup> Fortunately, there are a number of safeguards that governments and national statistical offices can use that can reduce both the likelihood of such misuse or its harm if it does take place. Moreover, countries can take special care to avoid or minimize the use of the riskiest sorts of data collection programs.

It also should be recognized that governments may gather information for a wide variety of investigative purposes. This chapter does not address the topic generally but focuses only on the misuse of the national statistical system to target population subgroups.

The chapter opens with a short discussion of the different data systems involved followed by a presentation of a conceptual framework of data types useful for considering such targeting threats and operations. The next section presents a summary recent research on the identification of specific instances of such targeting, providing some further discussion of episodes occurring in Europe in the 1930s and 1940s and references to individual studies so that those interested can explore these cases in more detail. Finally, the chapter concludes with a section describing some of the major safeguards against such misuses and a section discussing the issues raised more broadly.

---

<sup>1</sup> Seltzer & Anderson 2001.

<sup>2</sup> Seltzer & Anderson 2003.

<sup>3</sup> Seltzer 2005.

<sup>4</sup> United Nations 2003.

## I. The Main Population Data Systems Involved

The population data systems discussed in this chapter include regular population censuses, population registration systems, and various other kinds of administrative reporting systems. These systems and their major variants are summarized in Table 1, along with information on their population coverage, the level of geographic and subject-matter detail for which they can produce useful results, and their usual periodicity. This last factor is important in determining the timeliness of the results generated.

<b>Data source</b>	<b>Population units covered</b>	<b>Geographical detail</b>	<b>Subject-matter detail</b>	<b>Temporal dimensions</b>
<i>Population census</i>				
Regular	Usually all in a country	Very detailed	Limited	Every 5 or 10 yrs
Special	Usually limited to a state, province, or city	Very detailed	Very limited	Ad hoc, with no fixed periodicity
<i>Sample survey</i>				
One time	Often 1,000 to 100,000	Very limited	Can be quite detailed	One time and ad hoc
Periodic	Usually, fewer than one-time	Very limited	Can be quite detailed	Information obtained for a sample every month, quarter, etc.
Longitudinal	Usually, fewer than periodic	Very limited	Can be quite detailed	Information obtained for the same units every month, quarter, etc
<i>Administrative record systems</i>				
Vital (or civil) registration	Usually all vital events in a country	Very detailed	Live births, deaths, etc. and related factors	Continuous
Population registration	Usually all in a country	Very detailed	Can be quite detailed	Continuous

Table 1: Sources of Population Data.

The United Nations defines a population census as “the total process of collecting, compiling, evaluating, analyzing and publishing or otherwise disseminating demographic, economic and social data pertaining, at a specified time, to all persons in a country or in a well-delimited part of a country.”<sup>5</sup> Its essential features include: individual enumeration, universality within a defined territory, simultaneity, and defined periodicity, although the last feature

<sup>5</sup> United Nations 1998.

is often not achieved in a number of developing countries. Nevertheless, most countries of the world carry out population censuses on a fairly regular basis, most frequently at 10 year intervals.

The main advantages of a regular population census are that it obtains the same set of information from all members of the population using roughly the same procedures and a common reference date. Since all members of the population are covered in a census, censuses can be used to generate far more detailed cross tabulations than can be reliably produced from most sample surveys. In addition, comparably detailed cross tabulations can be produced from a census for the country as a whole and for all units at each level of areal disaggregation defined in the census geography (for example, province, county, town, village, postal zone, census tract, block, etc.) However, regular population censuses are generally massive undertakings which means that they are normally conducted only once or twice a decade and the questionnaire or schedule used must be kept as simple as possible. As a result, decennial census data are on average 5 years out of date and must be limited in subject matter detail.

Countries also carry out a range of sample surveys. In general, as shown in Table 1, sample surveys often directly compliment censuses in terms of their main features. For example, surveys can be strong in subject-matter detail and timeliness, but are weak in geographical detail and often exclude segments of the population that censuses cover (for example, those living in institutions and other types of group quarters.)

The third major source of population data are administrative reporting systems of one kind or another. Table 1 focuses on two such systems that are sometimes confused with one another: a civil registration system and a population registration system. The former records vital events (live births, and deaths, and in many countries, fetal deaths, marriages and divorces as well). Virtually all countries have a vital registration system legally requiring the registration of at least live births and deaths, although for many developing countries the registration of live births and deaths, particularly the many that do not take place in hospitals, is very incomplete. The civil registration of these vital events is the source of a nation's vital statistics.

By contrast, a well-functioning population registration system strives to maintain a record for each person from birth or arrival in the country, through a person's education, work, and retirement history, to death or other permanent departure from the country. Such a system also generally strives to keep track of changes in residences and is often linked to other government registers (for example, tax, voting, social security, health). The United Nations defines a population register as a mechanism

“for the continuous recording of selected information pertaining to each member of the resident population of a country or area, making it possible to determine up-to-date information about the size and characteristics of the population at selected points in time. Because of the nature of a population register, its organization, as well as its operation, should have a legal basis. Population registers start with a base consisting of an inventory of the inhabitants of an area and their characteristics, such as date of birth, sex, marital status, place of birth, place of residence, citizenship and language. To assist in locating a record for a particular person, household or family in a population register, an identification number is provided for each entity.

The population register can contain other socio-economic data, such as occupation or education. The population register should be updated by births, deaths, marriages and divorces, which are part of the civil registration system of the country. The population register is also updated by

migration. Thus, the population register is the result of a continuous process, in which notifications of certain events, which may have been recorded originally in different administrative systems, are automatically linked to a population register on a current basis. The method and sources of updating should cover all changes so that the characteristics of individuals in the register remain current.”<sup>6</sup>

As this definition implies, to function a population registration system requires both a full inventory of the population to establish the system and means of obtaining in a timely manner information on all live births, deaths, and moves of persons included to keep the system up to date. It is generally comparatively easy technically, although not necessarily operationally, to establish a population registration system since the initial inventory can be based on a census-like operation. It should not be based on the population census itself since the latter under most statistics or census laws is carried out under confidentiality protection provisions. Thus, sharing of information between the census (a statistical system) and the population registration system (assuming it had some administrative purposes) would be a violation of statistical confidentiality.

On the other hand, the maintenance of a population registration system is an extremely difficult job logistically. Not only must all birth and deaths be reported to the vital registration system and the reports transferred to the population registration staff in a timely manner, but all moves within a country must also be recorded. Indeed, unless a country has complete birth and death registration, there is little point in trying to establish a population registration system. In addition, because population registers have been involved in some of the most serious human rights tragedies of the twentieth century (see Table 3 below), great care is needed to limit the kind of information collected and to use other safeguards against misuse.

## II. Reconceptualizing Population Data

Most of those who produce and use population data are aware of two broad classes of data, (1) the individual level data for each unit (person, family, household, or dwelling) and (2) the aggregates based on tabulating these individual records. The individual records may also be analyzed in more complex ways through multivariate analysis (for example, regression analysis). However, from the perspective of human rights concerns, the key issue is how well the data lend themselves for targeting potentially vulnerable individuals or groups. In these circumstances, a three-way classification of data types (that is, macro data, meso data, and micro data) becomes relevant.

---

<sup>6</sup> United Nations 2001.

Data type	Definition
Macro	<i>Macro</i> data refer to tabulated aggregates for national or large geographic areas
Meso	<i>Meso</i> data refer to tabulated data for sufficiently small geographic areas that the results can be used operationally to identify and target a vulnerable population subgroup. They are statistical results presented at such a fine level of geographic disaggregation, whether in tabular or graphic form, that the results may be used in conducting field operations at the local level.
Micro	<i>Micro</i> data refer to identifiable records for each individual

Table 2: Types of Population Data.

As the definitions of these concepts provided in Table 2 make clear, macro data are simply traditional census or survey tabulations for large geographic areas, while micro data refer to the information contained in the individual unit records for each member of the population covered. Initially at least, such unit records usually contain or are linked to identifying information such as name and address. The concept of meso data is a relatively new one. As Seltzer and Anderson<sup>7</sup> observed in a detailed examination of the use of population data systems to target vulnerable population subgroups,

„While the relative protection offered by the statistical aggregates of macro data and the relative vulnerability of individual records that constitute micro data have long been recognized, the special risks posed by meso data have only been explicitly examined in the past few years.”

They went on to describe the role of meso data in such targeting in these terms,

„meso data are statistical results presented at such a fine level of geographic disaggregation, whether in tabular or graphic form, that the results may be used in conducting field operations at the local level. Thus the borderline between macro data and meso data will depend in part on the size of the geographic units, the distribution of the target population among these units, and the intended operational uses. For example, census aggregates showing the number of persons in a target population for an individual small village maybe operationally useful, while similar data for a large city, would need to be further broken down by tract, ward, or even block to be operationally useful.”

Two additional points should be kept in mind when considering the concepts of macro, meso and micro data. First, traditional laws that protect the confidentiality of statistical data bar the release of individually identifiable data. In other words, they explicitly pertain to the release of identifiable micro data. Statutory protections do not generally cover the targeting of vulnerable groups through meso data, although statistical agency disclosure policies can sometimes provide considerable protection.

Second, although the present paper focuses on the risks associated with micro and meso data, it should be recognized that macro data have frequently been used in efforts to stigmatize vulnerable populations as part of an effort to mobilize public support for systematic efforts directed against such groups. (Indeed, the statistical concepts involved in producing

<sup>7</sup> Seltzer & Anderson 2003.

such tabulations have often helped shape the government's definition of the "problem" population.)

### III. Research Results on Targeting

Table 3 presents, in highly summarized form, an updated listing of instances where efforts were made by national states to use a population census, a population registration system, or a related data system to target vulnerable population subgroups (or individuals) for adverse action. In writing about an earlier version of this table Seltzer and Anderson [2003] commented,

„We would stress that among the cases listed ... there was a wide range in severity of the consequences for the individuals and groups so targeted or identified. In some cases, targeting was part of a genocidal program. In other cases, the potential consequences were far less grave. Also some of the instances cited were fully implemented examples of targeting, while other represent intentions that were never fully implemented. Furthermore, given the range of time periods and countries covered ..., there is wide variation in the extent to which each data gathering activity listed was subject to statistical confidentiality legislation. All the cases listed ... do have two features in common: (1) they involve a population data system that was part of the national statistical system, or was created under the auspices of the national statistical authorities; and (2) in each case targeting was attempted or was an explicit or implicit goal. Our justification for using such a broad definition is simple. In view of the gravity of some of the examples, both for those targeted and for the statistical programs, agencies, and staffs involved, we consider that full exploration of the historical record is important so that we can assure that we have done all we can to avoid any new misuse by national or local governments.“

At this point Table 3 contains 17 cases. Underscoring the rapidly evolving nature of this line of research, we note that in 2001 the first time the equivalent of this table was compiled, 10 incidents were listed.<sup>8</sup> The additional cases now included relate to: (a) the Australian aborigines, (b) the population registration system in China during the Cultural Revolution, (c) the 1941 Hungarian Census; (d) Norwegian population censuses in the 19th and early 20th centuries, (e) the South African 1951 population census and that country's population registration system, (f) the 1910 U.S. population census, and (g) the recent effort made to use information collected by the U.S. National Center for Education Statistics under a pledge of statistical confidentiality to investigate and prosecute terrorism.

As is clear from Table 3, efforts to misuse population data systems to target vulnerable population subgroups, along with actual misuse have occurred in both totalitarian and democratic countries, although in democratic societies such misuses tended to occur primarily in times of national stress. Moreover, the scale of the ensuing human rights abuses was much smaller and their nature tended to be milder in democratic than totalitarian states.

Population registration systems were involved in 8 of the 17 cases listed in Table 3, regular decennial censuses in 7 cases, special censuses in 4 cases, and other or unspecified systems were involved in 2 cases. (The numbers total to more than 17 because in several of the cases listed, more than one data system was used in the targeting.)

---

<sup>8</sup> Seltzer & Anderson 2001, 487.

Although the possibility of population census-based targeting frequently receives much attention in the press and is the cause of much public fear, the record seems to be clear that population registers are an equal if not greater potential threat. Population registers were associated with such well-known gross abuses of human rights as the Jewish Holocaust in the Netherlands (over 70 percent of the resident Dutch Jews killed), Apartheid in South Africa, the Cultural Revolution in China, and the 1994 Rwandan genocide.

The targeted groups in the 17 episodes listed in Table 3 included racial and ethnic minorities (Jews, Roma, Samis, Kvens, Tutsi, and Japanese Americans), lingual minorities (German speakers in Hungary in 1945 and 1946), indigenous populations (Australian Aborigines and Native Americans), subject populations (the African and "Colored" populations in South Africa), socially defined out casts (those from a "bad" social class in Maoist China), and legal out casts (suspected draft registration violators in the United States in World War I and suspected terrorists in the United States after 9/11).

In terms of geographical scope, all regions of the world are represented in Table 3, except Latin America and Western Asia. It is not clear whether this geographic variation represents a real difference in regional experience or is an artifact of the limited research on the use of meso and micro data for targeting in these two regions.

Six of the seventeen examples listed in Table 3 relate to the targeting of Jews and Roma (sometimes referred to as Gypsies) for segregation, forced migration, and extermination by the Nazi authorities and some of their allies during World War II as part of the Holocaust (Shoah and Porajmos). Five of these examples were discussed in detail by Seltzer<sup>9</sup> and all six of them in more summary form in Seltzer and Anderson.<sup>10</sup> While the activities in each country are listed as one "incident" in Table 3, several are compound in nature. For example, while Seltzer<sup>11</sup> describes the use of a 1942 special census to identify Jews in Norway as a preparatory step to their expulsion from the country, but one not carried out by the Norwegian statistical office, Søybye<sup>12</sup> describes the persistent efforts of the Director-General of that office to take advantage of the situation to establish a population registration system.

It is not necessary to repeat here the detailed descriptions of these six cases contained in Black, Seltzer, and Søybye and the sources cited therein.<sup>13</sup> As Seltzer and Anderson observed

„although these six cases were Nazi-inspired crimes, in only two cases, Germany itself and Poland, could the misuse of the data systems be attributed solely to Nazi initiatives. In France, Henri Buhle and René Carmille, and in Norway, Gunnar Jahn, the heads of the statistical agencies, took advantage of the political climate of German occupation or influence, to expose vulnerable target populations to further risks by proposals to undertake major new data-gathering efforts to serve both statistical and administrative purposes (Rémond, 1996; Søybye, 1998).”<sup>14</sup>

With respect to the Netherlands they noted:

---

<sup>9</sup> Seltzer 1998.

<sup>10</sup> Seltzer & Anderson 2001.

<sup>11</sup> Seltzer 1998.

<sup>12</sup> Søybye 1998.

<sup>13</sup> Black 2001; Seltzer 1998; Søybye 1998.

<sup>14</sup> Seltzer & Anderson 2001, 486.

Place	Time period	Targeted individuals or groups	Data systems involved	Type of data <sup>a</sup>	Human rights violation or presumed state intervention	Source
Australia	19 <sup>th</sup> and early 20 <sup>th</sup> centuries	Aborigines	Population registration	Micro	Forced migration, elements of genocide	Kraly and McQuilton, 2002
China	1966 - 1976	Bad-class origin during cultural revolution	Population registration	Micro	Forced migration, instigated mob violence	Qin, 2004
France	1940 - 1944	Jews	Population registration, special censuses	Macro, micro	Forced migration, genocide	Rémond, 1996; Seltzer, 1998
Germany	1933 - 1945	Jews, Roma, and others	Numerous	Macro, micro	Forced migration, genocide	Seltzer, 1998
Hungary	1945-1946	German nationals and those reporting German mother tongue	1941 Population Census	Micro	Forced migration	Gal, 1993
Netherlands	1940 - 1944	Jews and Roma	Population registration system	Macro, meso, micro	Forced migration, genocide	Seltzer, 1998
Norway	1845 - 1930	Samis and Kvens	Population censuses	Macro, micro	Ethnic cleansing	Lie, 2002
Norway	1942 - 1944	Jews	Special census and proposed population registration	Macro, micro	Genocide	Seltzer, 1998; Søbye, 1998

a. See Table 2 for definition



Poland	1939 - 1943	Jews	Primarily special censuses	Macro, micro	Genocide	Seltzer, 1998
Romania	1941 - 1943	Jews and Roma	1941 Population Census	Macro, micro	Forced migration, genocide	Black, 2001
Rwanda	1994	Tutsi	Population registration	Macro, micro	Genocide	des Forges, 1999
South Africa	1950 - 1993	African and "Colored" populations	1951 Population Census and population register	Micro	Apartheid, voter disenfranchisement	McNeil, 2002
United States	19th century	Native Americans	Special censuses, population registers	Macro, micro	Forced migration	Seltzer, 1999
United States	1917	Suspected draft law violators	1910 Census	Micro	Investigation and prosecution of those avoiding registration	Seltzer and Anderson, 2003
United States	1941 - 1945	Japanese Americans	1940 Census	Macro, meso, micro(?)	Forced migration and internment	Seltzer and Anderson, 2000; 2003
United States	2001 - continues	Suspected terrorists	Surveys and administrative data gathered by the National Center for Education Statistics	Micro	Investigation and prosecution of domestic and international terrorists	Seltzer and Anderson, 2002
USSR	1919 - 1939	Minority populations	Various population censuses	Macro, micro	Forced migration, punishment of other serious crimes	Blum, 2000

Table 3: List of Cases Where Population Data Systems Have Been Used to Target Individuals or Population Subgroups, Where Such Efforts Were Initiated, or Where Such Targeting Has Been Seriously Contemplated.<sup>15,16</sup>

<sup>15</sup> The time periods and intended targets specified refer only to those studied in the sources cited

<sup>16</sup> Sources: See chapter references.

"the effort at establishing a comprehensive population registration system for administrative and statistical purposes was completed even before the Nazi-occupation [Methorst, 1936; Thomas, 1937]. In 1938 Methorst, who was then the director-general of the Dutch Central Bureau of Statistics and formerly also head of the Dutch office of population registration, reported on the rapid progress being made in the Netherlands in implementing a new comprehensive system of population registration that would follow each person 'from cradle to grave' and would open up 'wide perspectives for simplification of municipal administration and at the same time social research' [1938: 33: 713-714]. By early 1941 Methorst's successor as head of the population registration office, J.L. Lentz, had quickly adapted this general 'cradle to grave' system to create special registration systems covering the Jewish and the [Roma] populations of the Netherlands. These registration systems and the related identity cards played an important role in the apprehension of Dutch Jews and [Roma] prior to their eventual deportation to the death camps ... The critical role of the registration system in the over-all process has been stressed by such diverse observers as the German Generalkommissar for administration and justice in the Netherlands in September 1941 [Presser, 1969: 38] and the British historian Bob Moore [1997]."<sup>17</sup>

However, some discussion of subsequent developments related to events in France, Germany, and the Netherlands seems called for.

*France.* The dispute over the precise role of René Carmille, head of the French statistical service under the Pétain, referred to in Seltzer<sup>18</sup>, appears to continue. Led by his son, Robert Carmille, there has been a continuing effort to rehabilitate Carmille's reputation by, among other things, alleging that some of his communications to his superiors in the Vichy government were only written to deceive the government while he was deliberately using his post as head of the statistical service to sabotage the government's anti-Jewish measures. However, René Carmille's role, and the role of the French statistical service itself, appear to remain ambiguous at best, pending further independent research.

*Germany.* Our understanding of the role of the German statistical service in the Holocaust continues to benefit greatly from ongoing research by a large number of German scholars. This research, some carried out by independent scholars and some working under commission from the German statistical service, has examined operations of individual population data system and individual statisticians and demographers under the Third Reich. Initially virtually all of this extensive body of research was available only in German. However recently Aly and Roth [1984] appeared in an English translation [2004].<sup>19</sup> Unfortunately, the English translation, like the German original is marred by the use of imprecise terminology when referring to the data gathering systems described which leads to significant confusion and ambiguities in interpretation. Wietog's important study of the role of the German statistical system in the Third Reich, previously cited in Seltzer and Anderson<sup>20</sup> remains untranslated, despite a strong plea at the December 2003 Berlin conference for this work to be translated in full into English and shared internationally.

*The Netherlands.* Despite the indisputable role that the ID cards linked to the Dutch population registration system played in the Jewish and Roma Holocausts in the Netherlands and the role that meso data from the 1930 Census played in the dot maps used to target Jewish neighborhoods, senior officials of Statistics Netherlands have persisted in the view that

<sup>17</sup> Seltzer & Anderson 2001, 486, 488

<sup>18</sup> Seltzer 1988, 522f.

<sup>19</sup> Aly & Roth 1984; English translation 2004.

<sup>20</sup> Seltzer & Anderson 2001.

the statistical service, its leadership, and its outputs were uninvolved. For reasons set out in more detail previously<sup>21</sup>, their continued denials appear unjustified. It may be noted that after the end of the war, the Dutch population registration system dropped religion as a registration item and according to Begeer<sup>22</sup> introduced a series of operational safeguards, such as decentralization, designed to make its use for targeting purposes more difficult. Despite these measures, it appears that the Dutch population had such reservations about the statistical system that they effectively forced Statistics Netherlands to abandon their plans for carrying out a population census in 1981.<sup>23</sup>

I am also unaware of any informed, independent research by Dutch scholars that explicitly examined the degree to which the Dutch statistical service, its leadership, its activities and its outputs in the 1930s and early 1940s contributed to ensuing events. One issue that such research might attempt to address is whether the Dutch central statistical agency and its Director-General used the mantle of that agency's influence and prestige in the 1930s to secure the passage of the legislation needed to strengthen the county's population registration system. In this connection, it may also be noted that in the 1920s and 1930s the establishment and strengthening of national population registration systems was one of the major goals of the international eugenics movement and, according to van de Kaa<sup>24</sup>, Methorst was one of the "main protagonists" of the eugenics movement in the Netherlands in the inter-War period.

For more details about these and other individual episodes listed, see the individual sources cited in Table 3. In addition, Seltzer and Anderson provide some information about each of the listed events, except for the cases of the Chinese population registration system and the 1941 Hungarian Census.<sup>25</sup>

With respect to the former, the broad outlines of the Cultural Revolution in China are generally known. What is less widely known is the role that the population registration system played in targeting an unknown number of victims of such human rights abuses as forced migration and mob violence, sometimes leading to death, because they were identified in the register as coming from a "bad" social class. It may be noted that one of the reforms introduced, after the excesses of the Cultural Revolution was recognized by Chinese authorities, was the elimination of social class as a variable in the population register.<sup>26</sup> In the case of Hungary, individual records from the 1941 Hungarian Census were examined at the end of World War II to target those who reported German as a mother tongue for deportation to East Germany or the Soviet Union. (It should be noted that, according to Gal, those Hungarians who had actively collaborated with the Germans during World War II had already been deported or killed prior to the census-based linguistic targeting of 1945 and 1946.<sup>27</sup>)

The point of Table 3 is not to discourage the collection and use of population statistics. Rather, it is intended to remind those proposing to gather such data that they carry a heavy

---

<sup>21</sup> Seltzer 1998, 523ff; Seltzer & Anderson 2001, 486, 488.

<sup>22</sup> Begeer 1998.

<sup>23</sup> Choldin 1988, 147; van der Laan 2000.

<sup>24</sup> van de Kaa 1998, 113.

<sup>25</sup> Seltzer & Anderson 2001; 2003.

<sup>26</sup> Qin 2004.

<sup>27</sup> Gal 1993.

obligation to ensure that the systems they develop do not easily lend themselves to kinds of misuse portrayed in Table 3 and that continued attention be given to the prevention of misuse. Failure to respect these obligations, as discussed below, can lead to the public's refusal to provide complete and accurate responses. This, in turn, can deprive the government and all other data users with the statistical data they need.

## IV. Safeguards Against Misuse

Seltzer and Anderson<sup>28</sup> identified five potential safeguards against the kinds of misuse described in Table 3. These were: (1) substantive safeguards, (2) methodological and technological safeguards, (3) organizational and operational safeguards, (4) legal safeguards, and (5) ethical safeguards. It must be noted that these safeguards, used individually, rarely provide an absolute defense against misuse. However, used jointly they can often avoid, deter, delay, and minimize the adverse impact of efforts to misuse population data systems to target populations for human rights abuses. In the context of human rights abuses, delay often directly leads to mistreatment avoided or minimized, including abuse-related mortality.

*Substantive safeguards.* Briefly, substantive safeguards refer to omitting sensitive items (for example, race, ethnicity, tribal group, language, religion) from data collection systems, particularly a population census or a population registration system. Seltzer and Anderson characterized this as the "ultimate safeguard" and noted that "this safeguard, while often perceived as reducing the analytical or policy usefulness of the involved data system, has been deliberately employed in several countries that had histories of misuses associated with major abuses."<sup>29</sup>

*Methodological and technological safeguards.* Methodological and technological safeguards against operational targeting include the collection of data on sensitive topics using sample surveys based on multistage probability designs rather than complete count information from censuses or population registers or basically unclustered systematic samples based on these sources. In addition, as Seltzer and Anderson noted

„Another broad technological approach is the deliberate introduction of errors into the data set. These include systematically swapping responses for individual items between records, introducing perturbations in specific items, top (or bottom) coding of quantitative items so that unduly large (or small) responses are grouped together to protect the identity of respondents, coding categorical data in broad response categories or using only large areal units for similar purposes.“<sup>30</sup>

*Organizational and operational safeguards.* Organizational and operational safeguards involve arrangements designed to make it more difficult, or at least more time-consuming, for respondent identification information to be associated with information on sensitive data items.<sup>31</sup> The importance of organizational and operational safeguards has taken on added importance now that national statistical offices are able to store completed census and popu-

<sup>28</sup> Seltzer & Anderson 2001, 495ff.

<sup>29</sup> Seltzer & Anderson 2001, 495.

<sup>30</sup> Seltzer & Anderson 2001, 497.

<sup>31</sup> Seltzer & Anderson 2001, 497f.

lation registration forms, including name and address information, in machine readable form suitable for case-by-case matching.

*Legal safeguards.* Legal safeguards have long been perceived as the primary safeguard against the misuse of information obtained by a national statistical agency to harm respondents. The statistical and census acts of most countries bar the use of information obtained in statistical inquiries to harm respondents and their families. Moreover, as Seltzer and Anderson have observed, “in a few countries these laws and regulations even extend to barring the collection or storage of data on sensitive topics.”<sup>32</sup>

In the context of targeting vulnerable populations, however, laws relating to statistical confidentiality have two important weaknesses: first, they focus exclusively on micro data, leaving the issue of use of meso data for such targeting unregulated or at the discretion of the statistical agency; second, statistical confidentiality laws can be, and have been, set aside in times of perceived national emergency to permit the use of individual-level information gathered under a pledge of confidentiality to target population subgroups and individuals.<sup>33</sup> Accordingly, it is only prudent for a statistical agency to use a coordinated package of safeguards rather than relying solely on statistical confidentiality laws.

*Ethical safeguards.* Despite the official and scientific character of the work of national statistical agencies, these offices and their leadership and staff are subject to a number of ethical norms.<sup>34</sup> Internationally, many of these norms are embodied in the Fundamental Principles of Official Statistics adopted by the UN Statistical Commission.<sup>35</sup> Principle 6 of this document states

„Individual data collected by statistical agencies for statistical compilation, whether they refer to natural or legal persons, are to be strictly confidential and used exclusively for statistical purposes,”

which clearly precludes the use of micro data for targeting purposes. The International Statistical Institute’s “Declaration of professional ethics for statisticians” also refers to the obligation of statisticians to respect confidentiality assurances made to respondents as do the ethical statements of several national statistical associations.<sup>36</sup> In light of the limitations of legal and other safeguards, ethical standards can play and have played an important role in preventing misuse of data systems or minimizing the impact of such misuse.<sup>37</sup>

In addition to these five safeguards, Seltzer discussed a number of what were termed “prevention” and “coping” strategies for dealing with perceived ethical threats arising in government statistical work, including threats associated with the use of both meso and micro data to target vulnerable groups.<sup>38</sup>

---

<sup>32</sup> Seltzer & Anderson 2001, 498.

<sup>33</sup> Seltzer & Anderson 2001, 498; 2003.

<sup>34</sup> Seltzer 2005.

<sup>35</sup> United Nations Economic and Social Council 1994.

<sup>36</sup> International Statistical Institute 1986.

<sup>37</sup> Habermann 2005; Seltzer 2005; Seltzer & Anderson 2003.

<sup>38</sup> Seltzer 2005.

## V. Discussion and Conclusions

A key feature of the long-term health of a national statistical agency is its reputation. This reputation, in turn, is a function of its ability to serve three quite different ends: First, is its ability to provide the statistical data that users want in a timely and reliable manner. Second, is its ability to provide such data in an impartial manner. Third is its ability to maintain the confidence of data providers so that the responding public and enterprises continue to trust that their cooperation in statistical inquiries does not harm them or the appear to do so.

Closely related to the second and third factors is the importance of statistical agencies avoiding involvement in essentially administrative operations of government or lending the statistical agencies' good name to such administrative undertakings, particularly when they have a distinctly political character. For example, in the early 1950s, the South African Census and Statistics Office, with the enthusiastic support and involvement of its Director, was given responsibility for that country's newly established population registration system as a key element in the enhanced Apartheid system under the Nationalist government elected in 1948. Furthermore, the Census and Statistics Office, again with the active involvement of its Director, used the country's 1951 Population Census to establish the initial race classifications used in the population register. However, in time it became clear that the Census and Statistical Office was ill-suited to carry out the essentially administrative work involved establishing the population register, particularly the adjudication of contested racial classifications. As a result, progress in implementing the new system was slow, and by 1956 the Director was replaced and by 1959 the task and the related posts and office space was removed from the South African Census and Statistics Office.<sup>39</sup> Indeed, it took the Office several decades to recover from the experience.

Moreover, even when there is no immediate impact on a statistical agency's reputation because of its active involvement in targeting, whether on the basis of *micro* or *meso* data, there is a real possibility that even after 40-60 years such activities may cause considerable embarrassment to a national statistical agency. For example, in the 1980s the German Statistical office had to deal with strong public reactions based on its work in the late 1930s in support of the Holocaust and the US Census Bureau continues to have to defend itself for providing "proactive assistance" in targeting Japanese Americans early in World War II based on the 1940 Census.<sup>40</sup>

In these circumstances, both existing ethical norms and enlightened self-interest point in the same direction: national statistical agencies should avoid involvement in actions that might easily lend themselves to targeting vulnerable population subgroups or individuals. Table 4 presents a listing of critical and aggregating factors that, if present in an ongoing or planned data gathering effort, seem to increase the potential for targeting and related human rights abuses to take place.

---

<sup>39</sup> Seltzer and Anderson, 2003: 33ff.

<sup>40</sup> Habermann 2005; Seltzer 1998; Seltzer & Anderson 2003.

**A. Critical factors**

1. Population studied is weak or otherwise vulnerable.
2. Data gathering or research involves variables that are on “sensitive” topics, typically topics that are or can be used to identify or stigmatize one or more vulnerable groups, or use classifications that permit the identification or stigmatization of such groups.

**B. Aggravating factors**

1. All or substantially all of population is covered, i.e., sampling is not used.
2. Longitudinal data gathering is involved, or the activity can be linked to a longitudinal system.
3. Participation is mandatory or is effectively coerced.
4. Little or no input from the subject population in planning the data gathering or research activities. (The risk potential is further enhanced if there are substantial inputs in terms of expertise, staff, or funds from foreign persons or institutions.)
5. The data gathering or research is carried out in a war, a period of civil disruption, or during or shortly after a similar emergency.
6. Little or no attention given to organizational, operational, methodological, and technological safeguards against the misuse of information obtained for non-statistical purposes.
7. Confidentiality assurances provided to respondents have limited or no legal basis.
8. Ethical reviews are not carried out, are perfunctory, or are heavily influenced by utilitarian considerations.

Table 4: Factors Contributing to Higher Risk of Population Data Collection Effort based on Potential for Respondent or Group Harm<sup>41,42</sup>

In situations where one of the critical or several of the aggravating factors are involved, national statistical agencies, their leadership, and their professional and technical staffs will need to take special care to ensure that effective substantive, methodological and technological, organizational and operational, legal, and ethical safeguards are in place. Underlying such work is a free and open discussion of the issues involved.

For such discussions to have a strong factual basis, demographers, sociologists, historians, statisticians, and other scholars will need to identify and document successful applications of these safeguards as well as examining in further detail those instances of the misuse already identified and possible additional instances where population data systems were used for targeting vulnerable groups. Such research is best carried out by those knowledgeable both in the functioning of such data systems and in historical research. In these circumstances, an interdisciplinary research team is often a sound way of proceeding. Moreover,

<sup>41</sup> Note: The presence of either or both critical factors gives rise to a presumption of risk and each additional aggravating factor present further augments such risk. On the other hand, it should be emphasized that the presence of critical and aggravating factors does not mean that actual harm has occurred.

<sup>42</sup> Source: Seltzer, William 2003. “Data collection, Ethics Issues in.” In: *Encyclopedia of Population*. Paul Demeny and Geoffrey McNicoll, eds. New York: Macmillan Reference USA, 2003, 195ff.

since the results of such research have implications for many countries, it is important that reports of studies undertaken be translated into other languages.

## References

- Aly, Götz; Roth, Karl Heinz 1984: *Die restlose Erfassung: Volkszählen, Identifizieren, Aussondern im Nationalsozialismus* [Control Without Rest: Counting, Identifying, and Separating People under National Socialism], Berlin: Rotbuch Verlag.
- Aly, Götz; Roth, Karl Heinz 2004: *The Nazi Census: Identification and Control in the Third Reich*, Philadelphia: Temple University Press.
- Begeer, W. 1998: Personal communications.
- Black, Edwin 2001: *IBM and the Holocaust: The Strategic Alliance Between Nazi Germany and America's Most Powerful Corporation*, New York: Crown.
- Blum, Alain 2000: Stalinism and Population Statistics. Paper presented at the Annual Meeting of the Population Association of America, Los Angeles, CA.
- Choldin, Harvey M. 1988: Government Statistics: The Conflict between Research and Privacy, in: *Demography* 25 (1, Feb.), 145ff.
- des Forges, A. 1999: *Leave None to Tell the Story*, New York: Human Rights Watch.
- Gal, Susan 1993: Diversity and Contestation in Linguistic Ideologies: German Speakers in Hungary, in: *Language in Society* 22, 337ff.
- Habermann, Hermann 2005: Ethics, Confidentiality, and Data Dissemination, Paper presented at the presented at the 55<sup>th</sup> Session of International Statistical Institute(Australia), April 5-12, Session #IPM 1. Paper also available at <http://unstats.un.org/unsd/>.
- International Statistical Institute 1986: Declaration of professional ethics for statisticians, in: *International Statistical Review* 227ff. Also available at <http://www.cbs.nl/isi/ethics.htm>.
- Kraly, Ellen Percy; McQuilton, John 2002: The 'Protection' of Aborigines in Colonial and Early Federation Australia: The Role of Population Data Systems. Paper prepared the First International Conference on Population Geographies, University of St. Andrews, St. Andrews, Scotland.
- Lie, Einar 2002: Numbering the Nationalities: Ethnic Minorities in Norwegian Population Censuses 1845-1930, in: *Ethnic and Racial Studies*, Vol. 25, No. 5 (September), 802ff.
- Mackensen, Rainer ed. 1998: *Bevölkerungsfragen auf Abwegen der Wissenschaften: Dokumentation des 1. Colloquiums zur Geschichte der Bevölkerungswissenschaft in Deutschland im 20. Jahrhundert*, Opladen: Leske + Budrich.
- McNeil, Melissa 2002: Term paper prepared for the course, "Uses/Abuses of Government Population, Social, and Economic Data." Fordham University, Graduate School for the Arts and Sciences, Spring semester.
- Qin, Dan 2004: Term paper prepared for the course, "Demography, Human Rights, and Ethics." Fordham University, Graduate School for the Arts and Sciences, Autumn semester.
- Rémond, René et al. 1996: *Le 'Fichier Juif': Rapport de la Commission présidée par René Rémond au Premier Ministre*, Paris: Plon éditeur.
- Seltzer, William 1998: Population Statistics, the Holocaust, and the Nuremberg Trials, in: *Population and Development Review*, 24 (3): 511ff. Paper also available at <http://www.uwm.edu/~margo/govstat/integrity.htm>.
- Seltzer, William 1999: Excluding Indians Not Taxed: Federal Censuses and Native-American in the 19th Century. Paper presented at the 1999 Joint Statistical Meetings, Baltimore, MD. A shorter version appears in 1999 Proceedings of the Government and Social Statistics Section. Alexandria, VA: American Statistical Association, 161ff. Paper also available at <http://www.uwm.edu/~margo/govstat/integrity.htm>.
- Seltzer, William 2005: "Official Statistics and Statistical Ethics: Selected Issues." Proceedings of the 55<sup>th</sup> Session of The International Statistical Institute, Sydney (Australia), April 5-12, Session # IPM 81. CD disk. Paper also available at <http://unstats.un.org/unsd/>.



- Seltzer, William; Anderson, Margo 2000: After Pearl Harbor: The Proper Role of Population Statistics in Time of War. Paper presented at the Annual Meeting of the Population Association of America, Los Angeles, CA. . Paper also available at <http://www.uwm.edu/~margo/govstat/integrity.htm>.
- Seltzer, William; Anderson, Margo 2001: The Dark Side of Numbers: The Role of Population Data Systems in Human Rights Abuses, in: *Social Research* 68 (2, Summer): 481ff. Paper also available at <http://www.uwm.edu/~margo/govstat/integrity.htm>.
- Seltzer, William; Anderson, Margo 2002: NCES and the Patriot Act: An Early Appraisal of Facts and Issues. Paper prepared for presentation at the annual Joint Statistical Meetings, New York, August 10-15, 2002. Paper also available at <http://www.uwm.edu/~margo/govstat/integrity.htm>.
- Seltzer, William; Anderson, Margo 2003: "Government Statistics and Individual Safety: Revisiting the Historical Record of Disclosure, Harm, and Risk. Paper prepared for presentation at a workshop, Access to Research Data: Assessing Risks and Opportunities, organized by the Panel on Confidential Data Access for Research Purposes, Committee on National Statistics (CNSTAT), The National Academies, Washington, DC, October 16-17, 2003. Paper also available at <http://www.uwm.edu/~margo/govstat/integrity.htm>.
- Søbye, Espen 1998: Et mørkt kapittel i statistikkens historie? [A dark chapter in the history of statistics?], in: *Samfunnsspeilet* 8, no.4 (10 September). Oslo: Statistics Norway.
- United Nations Economic and Social Council 1994: Report of the Special Session of the Statistical Commission, New York, April 11-15. E/1994/29. Also available at <http://unstats.un.org/unsd/methods/statorg/FP-English.htm>:
- United Nations 1998: Principles and Recommendations for Population and Housing Censuses, Revision 1. Series M, No.67/Rev. 1, Sales No. E.98.XVII.8. Also available at [http://unstats.un.org/unsd/publication/SeriesM/SeriesM\\_67rev1E.pdf](http://unstats.un.org/unsd/publication/SeriesM/SeriesM_67rev1E.pdf).
- United Nations 2001: Principles and Recommendations for a Vital Statistics System, Revision 2. Series M, No.19/Rev.2, Sales No. E.01.XVII.10. Also available at [http://unstats.un.org/unsd/publication/SeriesM/SeriesM\\_19rev2E.pdf](http://unstats.un.org/unsd/publication/SeriesM/SeriesM_19rev2E.pdf).
- United Nations 2003: Handbook of Statistical Organization, Third Edition: The Operation and Organization of a Statistical Agency. Series F, No. 88, Sales No. E.03.XVII.7. Also available at [http://unstats.un.org/unsd/publication/SeriesF/SeriesF\\_88E.pdf](http://unstats.un.org/unsd/publication/SeriesF/SeriesF_88E.pdf).
- van de Kaa, Dirk 1998: A Brief Note on Population Issues in The Netherlands during the 20th Century, in: Mackensen 1998, 107ff.
- van der Laan, Paul 2000: The 2001 Census in the Netherlands Integration of Registers and Surveys. Paper prepared for the conference on 'The Census of Population: 2000 and Beyond' organised by the Cathie Marsh Centre for Census and Survey Research, Faculty of Economics and Social Studies, University of Manchester, Manchester, UK, 22-23 June 2000. Reprint published. Also available at [www.ccsr.ac.uk/conference/VanderLaanpap.doc](http://www.ccsr.ac.uk/conference/VanderLaanpap.doc).
- Wietog, Jutta 2001: Volkszählungen unter dem Nationalsozialismus. Eine Dokumentation zur Bevölkerungsstatistik im Dritten Reich. [Population censuses under national socialism: A documentation of population statistics in the third reich], Berlin: Duncker and Humblot.

**Table-index**

*Table 1:* Sources of Population Data. .... 120

*Table 2:* Types of Population Data. .... 123

*Table 3:* List of Cases Where Population Data Systems Have Been Used to Target  
Individuals or Population Subgroups, Where Such Efforts Were Initiated,  
or Where Such Targeting Has Been Seriously Contemplated..... 126

*Table 4:* Factors Contributing to Higher Risk of Population Data Collection Effort  
based on Potential for Respondent or Group Harm..... 133