The International Database on Longevity: Structure and contents

Amandine Cournil[1], Jean-Marie Robine[2],
Heiner Maier[3], Jutta Gampe[3], and James W. Vaupel[3]

[1] INSERM, Démographie et Santé, Université de Montpellier I et CRLC, Val d' Aurelle, 34298 Montpellier Cedex 5, France.
E-Mail: acournil@ish-lyon.cnrs.fr
[2] INSERM, Equipe Démographie, et Santé Centre Val d'Aurelle - Parc Euromédecine, 34094 Montpellier Cedex 5, France.
E-Mail: robine@valdorel.fnclcc.fr
[3] Max Planck Institute for Demographic Research, Konrad-Zuse-Str. 1, 18057 Rostock, Germany

Abstract. The International Database on Longevity contains exhaustive information on validated cases of supercentenarians that allows unbiased estimates of mortality after age 110. This chapter describes the structure and contents of the database, including the different categories of age-validation procedures.

1 Introduction

Individuals who attain ages of 110 years or more are special. They are still rare, but the growth in the number of these so-called 'supercentenarians' in recent decades has permitted research. The emergence and proliferation of these truly oldest-old is of great interest to aging research, as their stories shed light on the prospects for human longevity.

Investigating extremes also helps to validate theories. While there is now undisputed evidence that the exponential increase in human death rates decelerates after about age 80, the ultimate trajectory of human mortality at highest ages has yet to be established. Different patterns have been suggested, but data at the highest ages have been too scarce to permit reliable assessments. Extreme ages are, by definition, rare, but sufficiently large samples are nonetheless needed to allow for accurate estimates. Furthermore, reports of extreme ages must always be

H. Maier et al. (eds.), *Supercentenarians*, Demographic Research Monographs,
DOI 10.1007/978-3-642-11520-2_2,

considered with caution. Age exaggeration is common, and only meticulously validated information should be trusted.

This context has inspired an international collaborative effort to gather demographic data for the highest ages in the form of the "International Database on Longevity" (IDL). The aim of this database is to compile exhaustive information on validated supercentenarians (110 years and over) as reliable data that can be used to estimate mortality trajectories. These data are provided by different countries involved in the IDL project as contributors.

Two aspects are crucially important, and must be stressed. First, the data collection should be such that the resulting samples allow unbiased estimates of death rates at the highest ages. Although a complete list of all supercentenarians would be ideal, for scientific purposes it is sufficient but mandatory that the collected cases are free of age ascertainment bias. This condition has implications for the data collection process. For example, media coverage is more common for the oldest cases, while individuals who died at younger ages (i.e., shortly after their 110th birthdays) will be underrepresented in the press. Thus, screening press reports generally does not constitute an appropriate means of data collection for mortality trajectory estimation.

If the probability of identifying cases is to be age-independent, then we need to have access to what we could call a complete list of supercentenarians. As we will never be able to determine whether a list is truly complete—i.e., that it does not miss a single case— we prefer the concept of an 'exhaustive' list. Such exhaustive lists will comprise all supercentenarians who have been reported and identified in a given region or country, and for a given time period, by national or regional authorities. A failure to include the age of every single supercentenarian should not prevent researchers from producing an estimate of death rates that is free of age bias.

The second crucially important aspect is age validation (see also the chapter by Poulain in this volume). The IDL aims to provide information only on supercentenarians whose ages have been validated. Two levels of scrutiny are used in the classification of the validation procedure, and we will describe the differences between the levels of validation in more detail in Section 2.2.

The IDL was originally designed for the collection of data on supercentenarians, and currently contains information only about individuals who have reached an age of at least 110. However, an expansion of the IDL to younger ages is already being prepared. For several countries, information on so-called 'semi-supercentenarians' (aged 105+) is

available, and will be included in the IDL in the future. Naturally, the number of semi-supercentenarians is much higher, which has implications for the feasible validation procedure.

How the supercentenarians were identified, and how their ages were validated in the individual countries that contribute to the IDL, is described in the chapters that make up Part II of this volume. In this chapter, we will outline the structure of the IDL, provide a summary of the different levels of age validation, and give a brief overview of the information contained in the IDL as of October 31, 2008.

While some of the information about the individuals in the IDL, particularly the details necessary for age validation, contains personal and, hence, confidential information, the ultimate aim of the IDL is to make reliable demographic data on the oldest-old humans accessible to the research community. Therefore, a version of the database is made public on the Internet, and can be accessed at `http://www.supercentenarians.org`.

2 Database structure

The International Database on Longevity (IDL) collects information on supercentenarians as well as on the validation process of the cases. A list of variables is given in Table 1. A few additional entries are available on each individual. This information is strictly confidential as it permits identification of the person. However, such information is mandatory in the validation process. Hence these variables, given in Table 2, are accessible only to researchers directly involved in the validation process.

A subset of variables of the IDL, marked by an asterisk in Table 1, is publicly accessible via the Internet. Further details on the public version of the IDL will be given in section 2.3.

2.1 Variables on the individuals

The first group of variables in the IDL contains information on the individual cases. For each person, a unique and permanent—but anonymous—identifier (IDNUMBER) is created. The gender of the person is given in the variable SEX. The variable BDATE lists the date of birth of the individual.

For deceased individuals, the entry AGEY gives the age at death in completed years, and the entry AGED shows the number of days lived

past the last birthday. The variable AGEINDAYS reports the age in days and the variable DDATE gives the date of death.

If an individual was still alive when the information was recorded, then AGEY, AGED and AGEINDAYS give the age information corresponding to the date when the individual was last known to be still alive. The corresponding date of last known survival is given by the variable ALIVEDATE. The indicator variable ALIVE records whether the information for a person in the IDL refers to a dead case, or a case that was still alive when the information was registered.

Information on the country of birth (BCOUNTRY), the country of death (DCOUNTRY), or, in the case of persons still alive, country of residence (ACOUNTRY), is also registered.

The remaining variables in Table 1 concern the process of age validation, which is discussed in the following section.

2.2 Age validation

As explained in the introduction, the database aims to compile validated cases for which the recorded age has been ascertained. The methods used for verifying ages differ among countries. Despite these differences, the IDL provides information about how meticulously each case has been validated, what documents were provided to the IDL in the course of the validation process, and the name of the researcher responsible for the validation.

The variable VALIDATION assigns the validation level of a case to one of two categories. VALIDATION=A requires that an early life document, preferably a birth record or baptism record, or an early census record, was available and checked. The respective document type is recorded in the BSOURCE variable. In the same way, the date of death of a fully validated case has to be backed up by a document, such as a death certificate or an entry in a death index. The kind of document used for verification of date of death is recorded in the variable DSOURCE. The reference person, also called informant, who is responsible for the verification, is denoted in the CONTACT field the database (see Table 5 in the Appendix).

If several types of documents can be found to substantiate either the date of birth or the date of death, only one of the available documents is mentioned. In such a case, the most reliable document is given. We are aware that the availability of several documents adds to the quality of the validation. But, for the sake of simplicity, this information is made available in the paper file that holds the various documents pertaining to the case, but not in the database.

Table 1. Variables recorded in the IDL database. Entries marked with an asterisk are contained in the online version (`www.supercentenarians.org`).

Description of the variable	Database field	Coding
Identification number	IDNUMBER*	unique numeric identifier
Age in completed years	AGEY*	Numeric data
Days since last birthday	AGED*	Numeric data
Age in days	AGEINDAYS*	Numeric data
Sex	SEX*	F: female M: male
Country of birth	BCOUNTRY*	ISO 3166 code
Country of death	DCOUNTRY*	ISO 3166 code
Country of residence	ACOUNTRY*	ISO 3166 code
Date of birth	BDATE*	dd/mm/yyyy
Date of death	DDATE*	dd/mm/yyyy
Alive or dead status	ALIVE	1: still alive 0: dead
Date when confirmed alive	ALIVEDATE*	dd/mm/yyyy
Reference person for validation information	CONTACT	3-letter code (see Table 5)
Official source (Statistics Institute)	SOURCE	Initials (see Table 4)
Validation status	VALIDATION*	A: see section 2.2 B: see section 2.2
Any particular observation concerning the validation	OBS	Text
Source of date of birth	BSOURCE	2-letter code (see Table 4)
Original document for birth date validation (photocopy, photo, scan)	BSOURCEORIG	1: yes 0: no
		0: no
Source of date of death	DSOURCE	2-letter code (see Table 4)
Location of a paper file for validation documents	PAPERFILE	3-letter code (see Table 4)

Table 2. Variables (containing confidential information) recorded in the restricted part of the IDL database

Description of the variable	Database field	Coding
Last name	LNAME	Full-text
Maiden name	MNAME	Full-text
First name	FNAME	Full-text
Birth place	BPLACE	Full-text
Region of birth	BPROVINCE	Code
State (US)		
Province (CAN)		
Region (FRA)		
Country (England & Wales)		
Place of Death	DPLACE	Full-text
Region of death	DPROVINCE	Code
State (US)		
Province (CAN)		
Region (FRA)		
Country (England & Wales)		

The entry BSOURCEORIG indicates whether information on date of birth is attested by a photocopy, photo, or scan of an original document. This information is only available for birth because it is a much more remote event compared to death, and must therefore be checked more carefully.

If a paper file containing the documents pertaining to a given case is available (with restricted access, because of the confidential information included) the location of these documents is indicated in the PAPERFILE field (see Table 4 in the Appendix).

For several countries, the validation procedure is not documented as thoroughly as category A requires, but individual cases have been thoroughly checked. Such cases are tagged as VALIDATION=B.

They can arise, for example, when a local residence register officially confirms a date of death, but without issuing any personal document on the individual. Or there may be cases in which the sequence of entries in national censuses was carefully checked by the respective national statistical office over the course of the individual's life.

The ways in which such VALIDATION=B cases arise differs between countries, and the actual validation steps are described in the country reports in Part II of this volume. The variable SOURCE in the IDL references the authority responsible for the validation (usually

national statistical offices or administrative bodies). In this case, the CONTACT field also gives the corresponding informant for the case.

Seven additional variables are available in the restricted part of the database. These are all data that identify the person, and are only accessible by researchers immediately involved in the case validation.

The last name or family name (LNAME), name at birth or maiden name for women (MNAME), and first name or given name (FNAME) are recorded in full text. For women, both marital name and maiden name are important in identifying the person at different stages of her life.

Detailed information about the place of birth and the place of death can also be found in the restricted area of the database. Although they do not directly serve as personal identifiers, this information, in combination with the date of birth, can easily be used to identify a person. Access to these variables is therefore restricted. The exact place of birth and death is recorded in BPLACE and DPLACE, respectively. Additionally, for some countries there is regional information in the BPROVINCE (and DPROVINCE) entry. These fields give the state for the United States, the province for Canada, the region for France, and the county for England and Wales.

2.3 Publicly accessible information

In keeping with the overall aim of the IDL—namely, to provide reliable demographic information on the oldest-old humans—, the information presented on the IDL website, `www.supercentenarians.org`, is intended to allow researchers to conveniently and appropriately use the IDL for their analyses. The variables included in the public version of the IDL have been discussed in Section 2.1. Besides the information on sex, ages, and dates, researchers will need additional information on data collection.

In particular, it is important to know the sampling frame of the data sets, as this allows researchers to take censoring and truncation into account in the analysis. As these sampling frames differ between countries (as do the validation procedures), the public version of the IDL is structured by country of death (or the country of residence for persons still alive). Information on the observation scheme is added in the so-called 'meta-data'. These indicate either the time period or the birth cohorts covered. The data also specify whether only deaths were sampled, or whether individuals alive were included as well. A brief summary on the validation process is also provided.

3 IDL overview

The IDL is an ongoing project with continuous updates of the information represented in the database. The following summary describes the contents of the publicly available information as of December 31st, 2008. Besides the total number of cases the number of females (f) and males (m), the age range of the individuals and the range of their birth years is given.

Table 3. Summary of the contents of the IDL as of December 31, 2008.

Country	Cases	females, males	Age range	Birth years
Belgium	5	5f	110–112	1882–1890
Denmark	2	2f	111	1884–1889
England & Wales	66	64f, 2m	110–115	1856–1895
Finland	6	5f, 1m	110–112	1878–1897
France	49	46f, 3m	110–122	1875–1893
Germany	17	14f, 3m	110–112	1883–1894
Italy	37	31f, 6m	110–113	1863–1893
Japan	78	66f, 12m	110–114	1884–1893
Quebec	10	8f, 2m	110–112	1852–1892
Norway	8	7f, 1m	110–112	1876–1893
Spain	28	20f, 8m	110–114	1878–1895
Sweden	12	11f, 1m	110–112	1874–1898
Switzerland	4	4f	110	1883–1890
USA	341	309f, 32m	110–119	1867–1889

Acknowledgements

The authors wish to thank all contributors to the IDL. Without their efforts and commitment, this ambitious project would not have been possible. Aurore Clavel and Sophie Le Roy entered the incoming data into the database. They also designed the first version of the IDL web pages. We also acknowledge the assistance of the IT staff of the MPIDR, especially Peter Wilhelm, who implemented the current website of the IDL.

Appendix

Coding of variable (see Table 1)

Table 4. Coding of entries in the IDL (see Table 1)

Birth source (BSOURCE)	
BA	Birth act
BC	Birth certificate
CS	Census list

Death source (DSOURCE)	
DA	Death act
DC	Death certificate
DI	National Death Index
TR	Transcription

Source (SOURCE)	
INE	National Statistics Institute of Spain
INSEE	Institut National de la Statistique et des Études Économiques (France)
ISQ	Institut de Statistique du Québec
ISTAT	Italian Bureau of Statistics
MHW	Ministry of Health and Welfare (Japan)
OGP	Office of the German President
ONS	Office for National Statistics (England & Wales)
SSA	Social Security Administration (USA)

Paper file (PAPERFILE)	
MPL	Université de Montpellier
DUK	Duke University

Table 5. Contact Code (CONTACT)

BDJ	Bertrand Desjardins
BEJ	Bernard Jeune
BEK	Bert Kestenbaum
DCH	Dany Chambre
FOT	Foti Tillo
GJK	Gert Jan Kuiper
HBL	Hélène Blanché
HCM	Heiner Maier
JMC	John McCormack
JMR	Jean-Marie Robine
MLP	Michel Poulain
RGG	Rosa Gómez-Redondo
ROY	Robert Young
RTH	Roger Thatcher
SAI	Yasuhiko Saito
SCO	Stéphane Cotter