Interface Demography Working Paper

No. 2020-01

How to calculate time-to-event? Five methodological approaches using the 2011 census.

Katrien Vanthomme*, Mariska Bauwelinck, Johan Surkyn & Laura Van den Borre

* Corresponding author
E-mail: Katrien.Vanthomme@vub.be
Phone: +32 2 6148128

DOI: 10.5281/zenodo.11060048

Interface Demography, Vrije Universiteit Brussel, Pleinlaan 5, 1050 Brussels, Belgium
Website: interfacedemography.be

How to calculate time-to-event? Five methodological approaches using the 2011 census.

Katrien Vanthomme*, Mariska Bauwelinck, Johan Surkyn & Laura Van den Borre

Abstract

The availability of the 2011 census-linked data with extensive follow-up information provides new research opportunities. This working paper aims to explore the various methodological approaches that can be applied when analysing survival data and provides some insights on the difference in research outcomes applying the different approaches.

Keywords

Truncation, Censoring, Event time, Census-linked data

Introduction

The availability of the 2011 census-linked data with extensive follow-up information provides new research opportunities, but also some methodological challenges. The array of methodological options may be a bit daunting. This working paper aims to present five different methodological approaches in the preparation of survival data (i.e. the calculation of exposure time) and provides guidance for the application of these approaches in different research outcomes.

Some of the terminology used in this working paper is specific for survival analyses. So, before diving in, we provide an overview of the most important concepts based on previous work (Kleinbaum and Klein 2012; Singer and Willett 2003). The analysis of the occurrence and the timing of events builds on three methodological components:

a) The event under study: e.g. death, migration(s), etc.
b) The date of entry: the initial starting point of the study.
c) Time scale: Is time recorded in precise units (continuous time) or in thicker intervals (discrete time)?

This working paper focuses on the calculation of time-to-event, which is the period between the beginning of the study (date of entry) and the occurrence of the event. Yet, time to the event of interest cannot always be observed in survival analyses. This analytical problem is known as censoring. Data are censored when we have only partial information about the event during follow-up. There are two main reasons for censoring. Firstly, a person may not experience the event during the time period under study (i.e. follow-up period). Any possible event after the end of the follow-up period is 'lost' in the data and can no longer influence analyses. Secondly, a person can be lost to follow-up when we do not dispose of further updates of a person's status during the follow-up period. Because we have no way of knowing whether the event under study occurred and when, the time period after loss to follow-up cannot be considered for analysis.

Censoring is an important methodological feature of a dataset. Data can be left-censored; right-censored or interval-censored. Left-censoring occurs when exposure time cannot be calculated due to an unobserved date of entry for certain individuals under study. This results in a true exposure time is less than or equal to the observed exposure time. Right-censoring occurs when exposure time is uncertain because the occurrence of the event is unknown. The true exposure time is equal to or greater than the observed exposure time. Interval-censoring incorporates left- and right-censoring. A respondent may be lost during follow-up but may appear again in the dataset after some time. In the case of interval-censoring, the true exposure time lies within the known time interval of the duration of data collection.

This working paper presents five possible approaches to investigate event occurrence using the 2011 census-linked mortality and migration data. We also provide a brief overview of different results these approaches generate.

Data

The following overview of methodological approaches is based on the 2011 census-linked data, and can be applied to any similar dataset. This nationwide dataset was composed through an individual-level record linkage between the Belgian 2011 census (n≈11 million Belgian residents) and national population register data on migration and mortality for the follow-up period 01 January 2011 - 31 December 2015. During this 5-year period almost 5% of the population died and 2% emigrated. For this working paper, the event under investigation is all-cause mortality. The date of entry is defined by the data design. Due to the availability of exact dates regarding birth, death and migration, the analyses can be conducted on a continuous time scale. The combined dataset is right- and interval-censored.

Methodological approaches

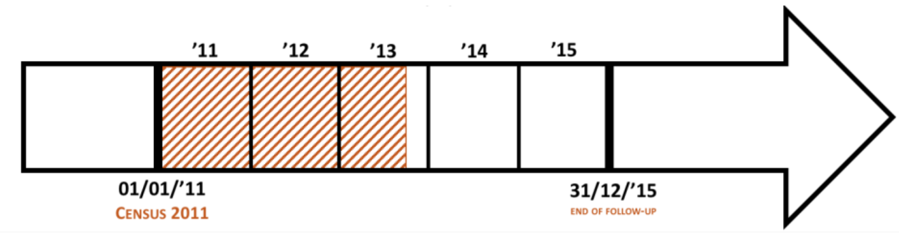
The following overview provides a description of five different approaches. Exposure variables were constructed in correspondence to these approaches. The choice of exposure variable depends strongly on the specific research question.

A. Classic approach

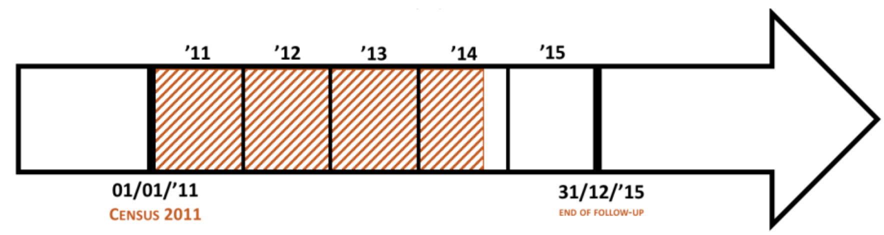
This approach has been widely accepted in international literature and has been applied in many Interface Demography studies. The approach follows all Belgian residents that are alive and living in Belgium at baseline (01/01/2011) and included in the 2011 census. Follow-up continues until one of the following events occur (whichever event comes first): death, emigration, loss to or end of follow-up period (31/12/2015). As such, we disregard any information after censoring. This could underestimate the exposure time.

Examples:

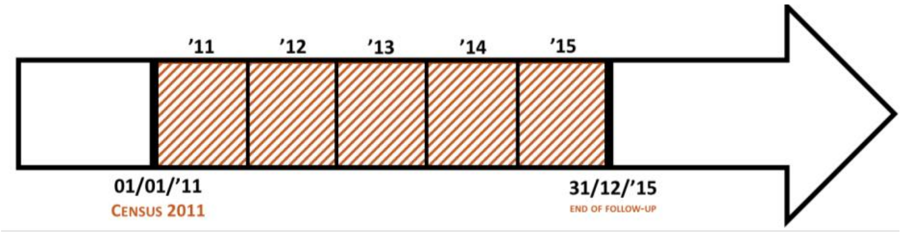
• *Person 1: alive in 2011 census, emigrates to Benidorm in 2013 (emigration)*



• *Person 2: alive in 2011 census, dies in 2014 (death)*



• *Person 3: alive in 2011 census, alive in 2016 (end of follow-up)*



B. Classic approach with return

This approach considers remigration in the calculation of exposure time. The study population includes (a) all Belgian nationals alive at baseline (2011 census) and residing in Belgium; and (b) Belgian nationals living abroad and who return to Belgium during the follow-up period.

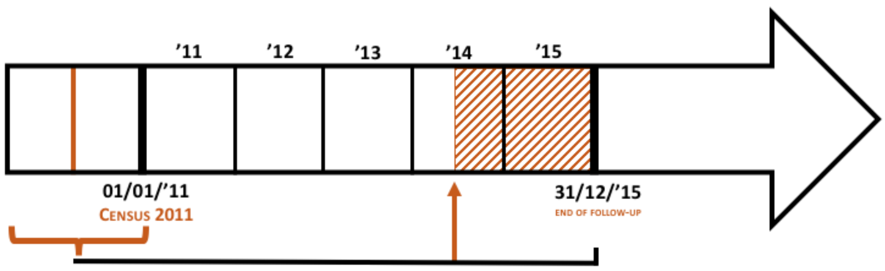
Hence, person time before emigration is considered for Belgian residents (a), as well as person time after remigration of returning Belgians (b). Person time before emigration and after remigration is considered until death, emigration, loss to or end of follow-up period. Exposure time is calculated until death, emigration, lost to or end of follow-up period, whichever event comes first.

Note that Belgian nationals living abroad and who do not remigrate are not considered.

Example:

• *Person 1,2 and 3: similar as in approach A.*

• *Person 4, who previously lived in Belgium, is abroad in 2011 (not in census). She returns to Belgium in 2014 and stays until end of follow-up.*

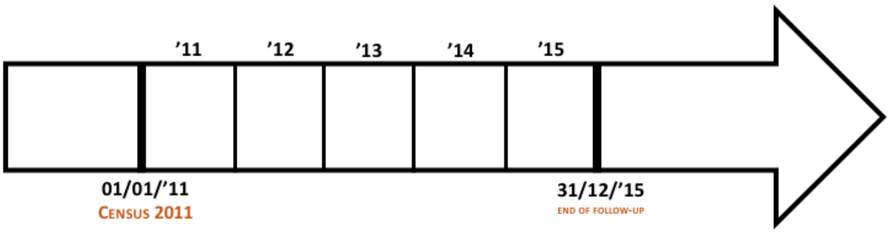


C. Non-emigrant population

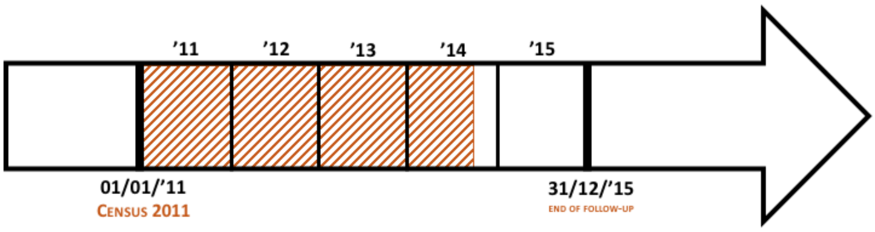
This approach excludes ever-emigrants from the study. The study population includes all living Belgian nationals residing in Belgium at baseline (2011 census) and who will not emigrate during the follow-up period. Exposure time considers death, loss to or end of follow-up period. This approach is not recommended when the outcome or important covariates vary between never-emigrants and ever-emigrants.

Examples:

• *Person 1: alive in 2011 census, emigrates to Benidorm in 2013 (excluded)*



• *Person 2: alive in 2011 census, dies in 2014 (death): similar as in approach A.*

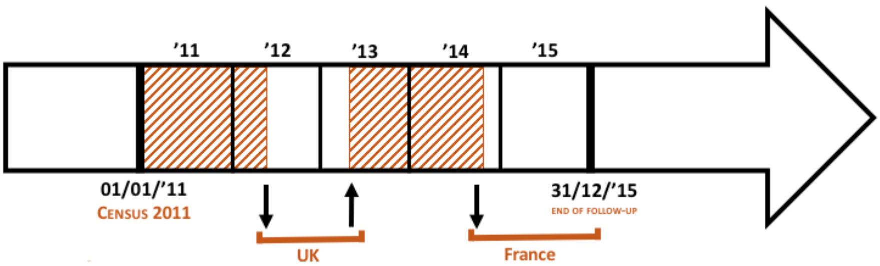


D. Interval censoring

This dynamic approach considers subsequent emigrations in the measurement of exposure time. The study population includes all Belgian residents that are alive and living in Belgium at baseline (2011 census), similar as in the classic approach. In contrast to the classic approach, the dynamic approach takes (multiple) migrations into account in the calculation of exposure time. When emigrants return during follow-up, they again contribute to the total exposure time. The method considers the period between baseline and the last event (either death, emigration, loss to or end of follow-up period), minus the time spent abroad. This method may be a useful approach for research interested in Belgian risk factors (e.g. air pollution) with short-term effects on mortality. When considering general risk factors (e.g. education) and their effects on mortality, approach E (see below) may be a better option.

Examples:

• *Person 5: alive in 2011 census, moved to the UK in 2012, returned in 2013, moved to France in 2014 and remained in France until end of follow-up*

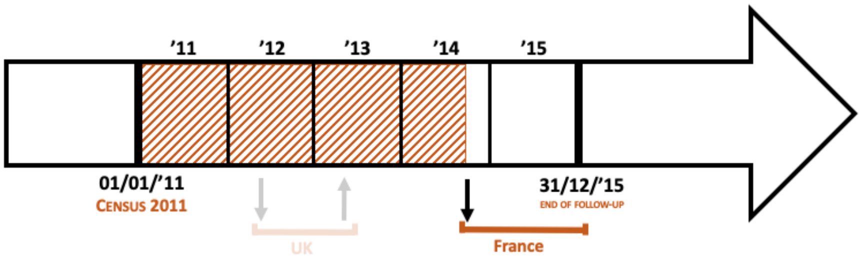


E. Classic approach until last event

We can also consider remigrations as proof-of-life and include the periods abroad as exposure time in our analyses. Different from the Classic approach (A), this approach takes the last event into account (whereas the Classic approach considers the first event). The study population includes all Belgian residents that are alive and living in Belgium at baseline (2011 census). For this population, follow-up continues until the last episode of one of the following events occur (whichever comes last): death, emigration, lost to or end of follow-up period.

Example:

- *Person 5: alive in 2011 census until his last move to France.*



How does the methodological approach affect the results?

Table 1 shows the results of the different approaches when calculating crude and age-standardized death rates, stratified by gender. We mapped out the population, exposure time, number of deaths, crude death rate and age-standardised death rate during the follow-up period of 2011-2015 for the total population legally residing in Belgium at the moment of the census of 2011. This table presents the results when applying a different approach for the calculation of exposure time. We observe that the included population is largest in approach B, which allows for return during follow-up, and smallest in approach C, which excludes all ever-emigrants from the study. The population included in the study is equal in approach A (classic approach), D (interval censoring) and E (classic approach until last event). However, when we compare the total exposure time, we see that these differ according to the approach used. The total exposure time is largest in the approach which allows for interval censoring (D), followed by approach E, which follows-up until the last event. The total exposure time is smallest in approach C, which excludes all ever-migrants from the study. Logically, the number of deaths is largest in scenario E, which is the classic approach taking into account the last event. In scenario B, which allows for return during follow-up, the number of deaths is also a bit larger than in the other three approaches. The crude and age-standardized death rates are different across the five approaches. The highest crude death rate can be found in approach C (excluding the migrant population). However, after age-standardization, the highest mortality rate was observed when applying approach E, which is the classic approach until the last event. The difference in the crude and age-standardized mortality rates is then the result of differences in the age structure over the population included in the approaches.

Table 1. Summary of survival data (all-cause mortality) by gender according to censoring approach – total population legally residing in Belgium at the time of the 2011 census

	Approach A	*Approach B*	*Approach C*	*Approach D*	*Approach E*
MEN					
Population	5 397 743	5 400 604	5 294 067	5 397 743	5 397 743
Total exposure	25 580 362	25 590 952	25 347 883	25 806 411	25 653 074
Mean exposure; SD	4.74; 0.90	4.74; 0.90	4.79; 0.82	4.78; 0.81	4.75; 0.87
Number of deaths	259 624	259 690	259 624	259 624	262 277
Crude death rate per 100,000	1014.9	1023.9	1024.2	1006.0	1022.4
ASMR (95% C.I.)	1432.1 (1426.6-1437.6)	1441.9 (1436.4-1447.4)	1437.1 (1431.6-1442.6)	1429.9 (1424.4-1435.4)	1442.0 (1436.5-1447.5)
WOMEN					
Population	5 596 296	5 597 777	5 498 679	5 596 296	5 596 296
Total exposure	26 757 186	26 763 219	26 540 187	26 887 747	26 795 973
Mean exposure; SD	4.78; 0.83	4.78; 0.83	4.83; 0.74	4.80; 0.77	4.79; 0.81
Number of deaths	270 825	270 847	270 823	270 825	271 951
Crude death rate per 100,000	1012.2	1015.5	1020.4	1007.2	1014.9
ASMR (95% C.I.)	981.9 (978.3-985.4)	985.3 (981.8-988.9)	984.3 (980.8-987.9)	981.1 (977.6-984.6)	985.6 (982.0-989.1)

Conclusion

This working paper gave an overview of different approaches that can be applied when calculating time-to-event in the case of e.g. survival analysis. The results of the working paper indicated that that depending on the approach used in defining the study population and the exclusion criteria in calculating the exposure time, survival outcomes may differ. Therefore, careful consideration of the approach to use is important and depends on the research question under study.

References

Kleinbaum, David G., and Mitchel Klein. 2012. *Survival Analysis: A Self-Learning Text*. 3rd ed. Statistics for Biology and Health. New York: Springer.

Judith D Singer, and John B Willett. 2003. *Applied Longitudinal Data Analysis: Modelling Change and Event Occurrence*. New York: Oxford University Press.

Appendix

Table A1 Overview of approaches with exposure variables

	Approach	**Study population at baseline**	**Censoring at**	**Time-to-event variables**
A	Classic	Belgian nationals in Belgium	First event: Emigration, Death, Loss to follow-up, End of follow-up	populatie_sc1; sterfte_sc1; DOE_sc1; DOO_sc1; exposure_sc1
B	Classic with return	Belgian nationals in Belgium and Belgian repatriates	First event: Emigration, Death, Loss to follow-up, End of follow-up	populatie_sc2; sterfte_sc2; DOE_sc2; DOO_sc2; exposure_sc2
C	Non-migrant population	Belgian nationals in Belgium who never emigrated during follow-up	Deaths, End of follow-up	populatie_sc4; sterfte_sc4; DOE_sc4; DOO_sc4; exposure_sc4
D	Interval censoring	Belgian nationals in Belgium	Multiple migrations, Deaths, Loss to follow-up, End of follow-up	populatie_sc5; sterfte_sc5; DOE_sc5; DOO_sc5; exposure_sc5
E	Classic approach until last event	Belgian nationals in Belgium	Last event: Emigration, Death, Loss to follow-up, End of follow-up	