Measuring Irregular Migration

www.irregularmigration.eu

USING MORTALITY DATA TO EXTRAPOLATE THE STOCK OF THE IRREGULAR MIGRANT POPULATION

AUTHORS:

JOHAN SURKYN

Funded by the European Union

Co-funded by:

UK Research and Innovation

Toronto Metropolitan University

Canada Excellence Research Chair in Migration & Integration

Publication Information:

Project Acronym:	Measuring irregular migration and related policies (MIrreM)
Project No.	101061314
WP	WP6 – Methods Innovation Lab
Type of Publication	Technical Note outside
Title	Using mortality data to extrapolate the stock of the irregular migrant population
Version:	1
Date:	30/07/2024
Responsible Partner:	Vrije Universiteit Brussels (VUB)
Contributing Partners:	University for Continuing Education Krems (UWK)
Authors	Johan Surkyn (VUB)
Reviewers:	Albert Kraler (UWK)
Dissemination Level:	Public

Revision History:

Version	Date	Author	Organisation	Description
1	08/06/2024	Johan Surkyn	VUB	Initial draft
1	11/06/2024	Albert Kraler	UWK	Reviewer feedback
1	04/07/2024	Johan Surkyn	VUB	Preparation of final draft
1	30/07/2024	Adriana Harm & Albert Kraler	UWK	Formatting, addition of abstract, keywords and other bibliographical information and publication

Summary

This guide is intended to give an overview of practical steps to be taken to apply mortality-based extrapolation as a means of providing estimates of the size, age structure and (aggregated) national background of "hidden populations", that is, populations that are present within the national borders of a country a but who are not official residents of that country because they are not captured by population register data or other data (e.g. censuses) relying on an official residence criteria, such as the "usual residence" criterion . The method is composed of two steps: In the first step, an estimate of the population not captured by official data is made, using "unlinked deaths", i.e. the number of recorded deaths that cannot be matched to a registered resident and expected mortality risks as the basis for the estimate. In a second step, irregular immigrants are identitified, counting all regular status visitors and immigrants and subtracting the aggregated number of regular immigrants from the hidden population estimated in step. In short, the two-step method starts by first making a distinction in terms of registration status (=registered member of official population or not) and only in a second step distinguishing the population in terms of legal status (regular/irregular).

Summary 3
Table of contents 4
LIST OF TABLES 5
LIST OF FIGURES 5
THE MIRREM PROJECT 6
INTRODUCTION 7
DURATION OF STAY 8
BELGIAN CONTEXT 8
BASIC PRINCIPLE OF MORTALITY EXTRAPOLATION 10
Step 1: Mortality extrapolation procedure 11
1. Collecting mortality data 11
2. Check and aggregate mortality data 11
3. The donor mortality profile 12
4. Use and application of life table variables 13
5. Testing 14
Calculation example 14
Interpreting results 18
Step 2: Distinguishing between regular and irregular status population 18
Diplomats: 18
Visa waivered tourists: 19
Visa holders: 19
Visa obliged tourists: 19
Asylum seekers during procedure 19
Posted workers 19
Final result: 20
SOME THOUGHTS ON ESTIMATING THE DISTRIBUTION OF DURATION OF STAY 21

LIST OF TABLES

Table 1: m(x) calculation for recent immigrants of non-EU origin in Belgium 15
Table 2: observed and fitted m(x) values 16
Table 3: Extrapolation based on 3-component fitted mortality rates + population pyramid 17
Table 4: Regular mobile citizens 20

LIST OF FIGURES

Figure 1: fitting model life tables to observed values using Mortpak 16

MIrreM examines estimates and statistical indicators on the irregular migrant population in Europe as well as related policies, including the regularisation of migrants in irregular situations.

MIrreM analyses policies defining migrant irregularity, stakeholders' data needs and usage, and assesses existing estimates and statistical indicators on irregular migration in the countries under study and at the EU level. Using several coordinated pilots, the project develops new and innovative methods for measuring irregular migration and explores if and how these instruments can be applied in other socio-economic or institutional contexts. Based on a broad mapping of regularisation practices in the EU as well as detailed case studies, MIrreM will develop 'regularisation scenarios' to better understand conditions under which regularisation should be considered as a policy option. Together with expert groups that will be set up on irregular migration data and regularisation, respectively, the project will synthesise findings into a Handbook on data on irregular migration and a Handbook on pathways out of irregularity. The project's research covers 20 countries, including 12 EU countries and the United Kingdom.

TO CITE:

Surkyn, J. 2024. Using mortality data to extrapolate the stock of the irregular migrant population. Technical Note. Krems: University for Continuing Education Krems (Danube University Krems). DOI: https://doi.org/10.5281/zenodo.13133775

KEYWORDS

Mortality-based extrapolation of population; Irregular migrant population; Unregistered population; Estimation method; Mortality; Demography.

FUNDING ACKNOWLEDGEMENT

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the Research Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

In addition, MIrreM benefit from funding provided by UK Research and Innovation (UKRI) under the UK government's Horizon Europe funding guarantee. The Canadian research component of this project is undertaken, in part, thanks to funding from the Canada Excellence Research Chairs Program of the Government of Canada.

INTRODUCTION

This guide is intended to give an overview of practical steps to be taken to apply mortality-based extrapolation as a means of providing estimates of the size, age structure and (aggregated) national background of “hidden populations”. That is, populations that are present within the national borders of a country a but who are not (official) residents of that country and are thus not captured by population register data or other data (e.g. census data) relying on an official residence criterion. The method is composed of two steps: the first step makes general estimates that **do not differentiate** between regular and irregular residents. Furthermore, the term “residents” must be taken largely, since it refers to all mobile citizens, even including e.g. tourists or transiting drivers and passengers. In other words, there is no criteria on duration of stay for a person to be considered a resident. Only this first step is based on mortality data.

The second step identifies irregular immigrants. It does so by first counting all regular status visitors and immigrants, and subtracting the aggregated number of regular immigrants from the general total number of mobile citizens estimated in step 1. Regular immigrants are counted on the basis of tourism data, visa data, asylum data and other sources. In short, the two-step method starts by first making a distinction in terms of registration status (=registered member of official population or not) and only in a second step distinguishing the population in terms of legal status (regular/irregular).[1]

[1] Some groups of short term visitors will not be identifiable through available data. For the EU+ population, the only group of short-term stayers that are detectable are tourists and business visitors staying in hotels. But those who don’t stay in hotels and are still in their first 3 month period are not identifiable and therefore wrongly considered “irregular EU+”. For the non-EU+ group, the relevant criterion for their legal status is their visa or asylum procedure status. Therefore short-term stayers from Visa-wavered countries are also not detected unless they stay in tourism infrastructure.

DURATION OF STAY

As is explained in more detail below, mortality-based extrapolation of population actually estimates person-years of presence of non-resident populations aggregated during a year, which is equivalent to an average daily headcount during that year. As an example, 100 person-years require at least 100 different people if they all stayed for the entire year, but it could just as well be 200 different people each staying for 6 months, or any other combination that adds up to 100 person-years. For the interpretation of results, which combination is behind the 100 person-years does not really matter. If we could physically do a headcount on a random day of that year, we would on average always find 100 people present regardless of the combination of numbers of people and the matching duration of stay. However, the statistical notion of a "usually resident population" includes a duration of stay criterion. In that case, the combination of numbers of (individual) incoming foreign citizens and their respective durations of stay does make an important difference. 100 persons each staying the entire year will be counted as entirely being part of the usually resident population, regardless even of their legal status. But 1200 people all staying for exactly one month will not count at all for the usually resident population.

Even if the extrapolation model itself does not differentiate between the combinations mentioned above, there are a few indications that can shed some light on the distribution of population by duration of stay, such as tourism data, visa (C-type, D-type) data etcetera. The Belgian example described below did not explicitly study duration of stay.

BELGIAN CONTEXT

All recommendations formulated here are based on experience from the Belgian testcase which, to our knowledge, was the first one performed in such detail. Therefore, this guide heavily relies on the Belgian research data infrastructure with Belgium's National Statistics Agency STATBEL as the main data provider. Belgium has a central digital population register.

That system has been in place since the early 1990's and it allows to follow individuals over time. The population register is the backbone of the statistical system in the sense that it defines the target population for the census and most other statistical collection of data.

Secondly, Belgium's system of mortality registration involves a statistical as well as an administrative element. Both these aspects come together in a statistical bulletin, which is a form where all information is registered on the demographics of the deceased (age, gender, nationality), the circumstances of death (where and when it took place, whether an autopsy was performed). It also contains the detailed causes of death. This statistical form is filled out for **every** death, regardless whether it concerns members of the official population or non-residents. STATBEL receives all those statistical forms, and links the individual forms to the population register. This leaves a number of statistical forms that cannot be linked, because no match is found between an administratively confirmed death in the population register, and a statistically recorded death. Since they cannot be linked, those unmatched cases must by default concern deaths of non-registered people (=mobile citizens). These "**unlinked deaths**" are the main basis for mortality-based extrapolation. The **"linked deaths"** on the other hand, can be used for calculating life tables (see below).

Belgium and especially Brussels hosts a large international diplomatic community linked to the European institutions and NATO headquarters. Next to that, it is also marked by a strong presence of international businessmen, researchers and lobbyists. But Belgium is also marked by a diverse historic labor migration and a relatively high number of asylum seekers. Finally, significant numbers of Dutch and French citizens have settled on the Belgian side of the borders with the Netherlands and France, whilst maintaining their official residence in those countries. This varied composition of foreign presence, both regular and irregular, is the context for the second step. Preparing for that second step it is necessary to thoroughly analyze the major groups of incoming mobile citizens, and the statistical sources that can shed light on their numbers according to legal status.

BASIC PRINCIPLE OF MORTALITY EXTRAPOLATION

The basic idea is that if mortality risks[2] are known, we also know how many people are alive to produce one death: if the mortality risk is 0.001 (=1/1000), then each death theoretically corresponds to 1.000 people alive, 10 deaths correspond to 10.000 people alive etc. In other words, inversing the mortality risk turns it into an extrapolation factor. In the above example the inversed mortality risk or extrapolation factor is 1/0.001 = 1.000.

Since mortality risks depend strongly on age and gender, extrapolations must be performed for each combination of age and gender, and then added up. So, if the mortality risks are known and applied to the "unlinked deaths", this provides an estimate (by extrapolation) of the non-registered population. Mathematically this is very simple.

However, the true mortality profile (age/sex specific mortality risks) in the non-registered population is not known. It cannot be calculated because the population at risk ("risk set") is itself the topic of the research. Furthermore, mortality profiles differ across populations and along social gradients. To get around this, we have to use another mortality profile that is hypothesized to be representative for the non-registered population: we will refer to it as the "donor mortality profile". Choice of that donor profile will have a significant impact on the extrapolation outcome and it must be done with care.

In the Belgian application of mortality extrapolation, we used the mortality profile of recently immigrated registered newcomers from the same countries of origin as "donors" for the extrapolations. The idea behind this is that there exists a significant throughput of non-registered newcomers that become registered in a later stage. After all, many of these non-registered newcomers have a visa that will expire, or are already in the process of becoming regular citizens. This indicates temporal proximity between registered and non-registered groups. We can also suppose that recent immigrants are more comparable with not (yet) registered populations in terms of housing status, labor market position, integration in the Belgian health care system, etc..

[2] The demographic mortality measure actually used is the mortality rate, conventionally noted as m(x) in the life table

Step 1: Mortality extrapolation procedure

1. *Collecting mortality data*

The very first step is to collect time-stamped mortality data by (at least) age, gender and nationality. Instead of focusing purely on deaths among non-registered individuals, it is advisable to collect data on all deaths combined, and to separate the data later between a registered fraction and an unregistered fraction. Having collected data on both fractions from the same source will be beneficial in later steps, because the data will be comparable, contain largely the same covariates and are similarly organized. If available, two additional types of information (covariates) should be collected for all deaths: (1) duration of residence3, and (2) cause of death. If all that information is collected, the same source of data can be used for calculating the "donor" mortality profile as for counting the unlinked deaths.

It is also advisable to collect this data for the longest possible timeframe to maximize the sample size. Numbers of deaths tend to be small, especially in young age groups. Therefore random error is always an issue, and it is very likely that data will later have to be aggregated to reach robust estimates.

2. *Check and aggregate mortality data*

Data will almost certainly have to be age-aggregated in 5-year age groups, and also in nationality groups. But before aggregating, it is worth checking the age-at-death distribution. If mortality data for registered and unregistered citizens are both available, they can be compared to each other searching for abnormalities. In the Belgian case, deaths at age 0 (newborn deaths) counted an unrealistically high number of unlinked records. It is well-known that infant deaths are concentrated just after birth. In Belgium, babies dying in the first days or weeks after birth where often not declared at the registration office. As a result, they do not occur in the population register, and in case of dying wrongfully produce "unlinked deaths". Therefore, deaths at age 0 had to be discarded.

Another interesting check would be to investigate seasonality of deaths (linked as well as unlinked deaths). Deviating seasonal patterns in unlinked deaths can point at other factors of influence, such as tourism.

A second type of comparison between linked deaths and unlinked deaths is also particularly revealing: calculating the ratio of unlinked deaths by age, gender and country group divided by the numbers of linked deaths. These ratios or proportions are very interesting in their own right (see the case of unlinked deaths at age 0). These ratios can also be used to create a first rough population estimate. If the data providing linked and unlinked deaths are complete and compatible, and under the preliminary hypothesis that age-specific mortality levels are equal within each nationality group, a rough estimate can be calculated in a very simple and straightforward manner. Example: if unlinked female deaths in African nationalities in age group 20-25 would be 10% of linked deaths in the same group, then it is plausible that the

[3] Duration of residence will only be known for the registered fraction of deaths, which is sufficient.

unregistered female population from these nationalities would also be about 10% the size of the registered African population aged 20-25. If that registered population counts 1200 people in this age group, 120 (=10%) non-registered 20-25 year African nationals can be expected. Obviously, this simple calculation can be repeated for all age groups and both genders, and numbers can be added up to produce an initial total population estimate.

However, this initial estimate should be treated as a preliminary result only because it relies blindly on the mortality profiles of registered and non-registered populations being equal, and also being correctly and representatively recorded. Under the next bullet point in this guide, more information will be given on choosing, calculating and enhancing empirical mortality profiles so that they are suitable for making extrapolations.

A last type of check on mortality data is looking at the causes of death. Obviously, comparing cause of death between registered and non-registered populations has to be controlled for age and gender. Comparing causes of deaths is very revealing, because it may pinpoint particular health and mortality risks in the non-registered and irregular population. For instance, high age and gender-specific shares of deaths to non-natural cases (accidents, violent death, suicide) may indicate that extrapolations must take into account such higher risks. If such a higher risk factor is not taken into account, the donor mortality profile will contain mortality risks that are too low to be applied to the unlinked deaths, and therefore lead to overestimated extrapolation factors (1/mortality risk).

In the Belgian case, overall proportions of deaths to non-natural causes were not higher among unlinked deaths. The reason is probably that non-registered populations often live in urban areas, don't own cars but use public transport instead, and generally live a secluded life. By this style of living, they probably avoid the largest causes of non-natural death. On the other hand, we did observe significant age-specific differences in mortality due to non-natural causes.

3. The donor mortality profile

Mortality profiles are calculated in a demographic lifetable. Basically there are 2 options. The simplest solution is to use an existing lifetable. A very large collection of empirical lifetables can be found in the "Human Mortality Database" at the Max Planck Institute for Demography (MPID). After registration, access to the HMD-database is free. But mortality profiles found at MPID will not be specific to immigrants. Dedicated immigrant population lifetables can probably also be found online, in publications or at academic research institutes.

Instead of using empirical (observed) life tables, there is also the possibility of using model life tables. Model life tables are hypothetical tables that are either a synthesis of observed lifetables across countries and regions in the world, or calculated tables based on mathematical approximation of mortality curves (see Brass Logit System, Lee-Carter etc.). Again, they are not specific to immigrants. The United Nations provides model lifetables and also free software such as Mortpak that can be useful for applications where lifetables have to be matched, compared or selected:
https://www.un.org/development/desa/pd/data/model-life-tables.

The most elegant solution however is to calculate life tables using observed mortality data ("linked deaths") from the same source as the one providing information on the unlinked deaths. Life table calculation procedures and descriptions of related data needs are not included in this manual, since there is ample support and examples to be found online.

However, we do insist on mentioning that special care has to be taken in dealing with truncated records: i.e. cases where persons in the study population are not yet present on 1st of January of the observation year (=left truncation), or have left the study population before the end of the year for other causes than mortality (right truncation). Truncation is also labeled "lost-to-follow-up", because it refers to situations where an eventual death might not be recorded, and the record should therefore not be taken into consideration for calculations. Such situations are more frequent in migrant populations, because their elevated levels of mobility systematically cause more truncation. Furthermore, we know that return migration is often statistically registered much later than the actual migratory event took place. This has a distorting effect on mortality parameters. More precisely, it will lead to under-estimation of actual mortality levels.

For that reason, we excluded all truncated cases from the mortality input data in the Belgian application. Only records were taken into consideration for the risk set that pertain to people who were present both on 1st of January and 31st of December, or died within the year without lost-to-follow-up spells in-between. Even people who had a temporary absence during the year were thus excluded. People were considered absent either when they emigrated or when they statistically disappeared because of detection of loss of the official Belgian address ("administrative deletion"). This precaution has the benefit of completely deleting truncation in the data, and making life table calculation quite simple.

4. *Use and application of life table variables*

For the Belgian study so-called "abridged" lifetables were calculated, with age groups 0 (<1), 1-4, 5-9, 10-14, 15-19 ... 90-94, 95+. Note that results for the first age group (0) are not used, as we discarded deaths below age 1. This also implies that the second age group is only 4 years wide instead of the usual 5 years at higher ages.

The life table measure that is used for extrapolations is the "mortality rate", or the m(x) in the life table. Contrary to the probability of dying (or "q(x)"), the mortality rate is calculated as occurrence/exposure, or number of deaths at age x during the year divided by the total number of person-years of presence. For practical calculations this denominator is often replaced by the average population during the year, which boils down to supposing that people leaving the study during the year spent on average 6 months at risk. The average population is simply the average of the population on 1st of January and the population on 31st of December.

Since m(x) = deaths/average population, and since the extrapolation factor is 1/m(x), that extrapolation factor can be expressed as: average population/n of deaths.

Multiplying the number of unlinked deaths d(x) with the extrapolation factor is mathematically equal to dividing the number of deaths by the mortality rate: d(x)/m(x). This gives the expected average number of non-registered people during the year, or also the total number of person-years spent by the non-registered population. As explained before, both interpretations are valid, since 1.000 people spending and entire year in the study population, or 2.000 people spending half a year give the same expected number of people that would be found on an average day of the year: namely 1.000.

For each of the age groups in the abridged life table, for both genders, and for every study population these extrapolated numbers are added-up to provide totals.

5. *Testing*

Testing mortality-based extrapolation in a setting where the outcome is previously known is a perfect way of getting a feel for the technique and for discovering any flaws in the data or in the calculations. As a way of testing whether all calculations were done correctly, mortality extrapolation could be tried-out for a registered immigrant population of which numbers are already known. In the Belgian case, we performed such tests using French nationals officially residing in Belgium, i.e. the largest non-Belgian population. In the Austrian case, German nationals living in Austria could be an excellent testcase.

We calculated life tables based on mortality data observed among the French living in just one region, and used that “donor life table” to estimate the French population living in another region. Such a test immediately reveals calculation errors because the French officially living in each region are obviously recorded. Secondly, it also reveals which minimal size of population (or rather number of deaths) are required to reach stable estimates. As was already mentioned before, numbers of children are most difficult to estimate because mortality is extremely low at those young ages which has a double effect. First, just very few deaths are observed, which leads to relatively important yearly fluctuations. Second, as low mortality rates lead to very high extrapolation factors, each death weighs heavily on the extrapolation result. Such additive effects strongly influence stability of outcomes.

Calculation example

Table 1 below is an extract of calculations from the Belgian testcase. Table 1 depicts the raw data behind the “donor” life table. It calculates the extrapolation factors, first m(x) and then 1/m(x). It shows the observed number of deaths (column header “1.00”) of women of non-EU origin aggregated over the years 2011 to 2015. Only registered immigrated women having arrived in Belgium less than 10 year prior to those years are in the table. The column “Average” shows the sum of average populations over the five years. It is calculated using the midyear populations: (pop 1-1-2011 + pop 1-1-2012)/2 + (pop 1-1-2012 + pop 1-1-2013)/2 + (pop 1-1-2013 + pop 1-1-2014)/2 + (pop 1-1-2014 + pop 1-1-2015)/2 + (pop 1-1-2015 + pop 1-1-2016)/2). As mentioned earlier, midyear populations are a commonly used proxy for the average population or the person-years lived. Since both the deaths and the midyear populations are summed over 5 years, division by 5 is not necessary.

In Table 1, age groups are listed as rows, with “0” indicating less than 1 year old, “1” indicating 1 to 4 years old, “5” indicating 5 to 9, “10” indicating 10 to 14 years old, etcetera.

Table 1: m(x) calculation for recent immigrants of non-EU origin in Belgium

agegr1 * diedin * verbl * SEX Crosstabulation								
Count								
SEX	verbl11			died11-15	Average			
				1.00			m(x)	1/m(x)
Female	0-9	Dur of stay	0	0	134		0,00000	-
			1	1	8084		0,00012	8084,0
			5	1	33242		0,00003	33242,0
			10	7	36473,5		0,00019	5210,5
			15	11	38899		0,00028	3536,3
			20	15	65368,5		0,00023	4357,9
			25	29	120472		0,00024	4154,2
			30	44	128266,5		0,00034	2915,1
			35	60	98677,5		0,00061	1644,6
			40	51	67295		0,00076	1319,5
			45	54	43284		0,00125	801,6
			50	73	28628		0,00255	392,2
			55	63	21102		0,00299	335,0
			60	99	16059,5		0,00616	162,2
			65	92	11331		0,00812	123,2
			70	132	8783		0,01503	66,5
			75	125	5727		0,02183	45,8
			80	118	2690,5		0,04386	22,8
			85	58	925,5		0,06267	16,0
			9.094	24	253		0,09486	10,5
			95115.00	16	93		0,17204	5,8
		Total		1073	735788.5			

The mortality rate m(x) is simply the number of deaths / average population. The inverse of the mortality rates 1/m(x) at each age-interval can be used immediately as an extrapolation factor. But in this case, only one female child died in age groups 1-4 and 5-9, which makes the mortality rates unstable. In the age group 5-9 that mortality rate is extremely low (0,00003), and the extrapolation factor is therefore extremely high (33.242). This illustrates that the random effect of any additional death that would occur in that age group has huge effects on the extrapolation factor, which is a robustness issue. Such robustness issues can be solved by replacing the observed m(x) values with comparable values from another more robust source: for example a more robust mortality curve for a larger population with the same life expectancy, borrowed from the HMD-database.

Figure 1 shows fitting model life table equivalents to the observed m(x) values, in order to deal with robustness issues. Here, the Mortpak program was used for finding the best fitting lifetable. But as can be observed, the fitted curves deviate quite strongly from the observed one at young ages. Even the 3-component fit doesn't perform well in those ages, missing the typical hump of raised mortality for young adolescents. Moreover, replacing observed m(x) values by the fitted ones only solves robustness issues from the side of the extrapolation factors. The fact that mortality rates are extremely low under age 15 also suggests that in the non-registered population (the unlinked deaths) those young age groups will also feature low and likely unstable numbers of deaths. This second source of instability can only be dealt with by aggregating deaths over several years, in large groups of national origin etcetera. Even then, extrapolations in the youngest age groups will tend to have far larger error margins than in age groups where mortality rates are higher.

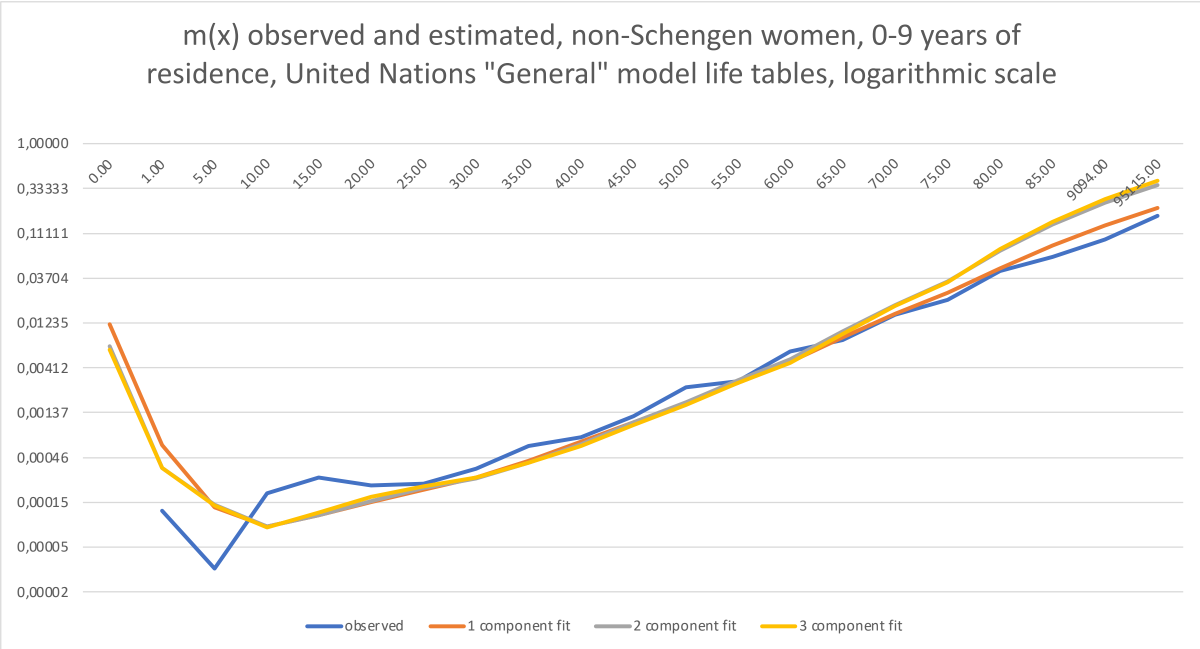


Figure 1: fitting model life tables to observed values using Mortpak[4]

Table 2: observed and fitted m(x) values

	observed	1 component	2 component	3 component
0.00	0,00000	0,01191852	0,00699517	0,00647883
1.00	0,00012	0,00061845	0,00035281	0,00035281
5.00	0,00003	0,00013605	0,00014405	0,00014005
10.00	0,00019	8,4018E-05	8,4018E-05	8,2017E-05
15.00	0,00028	0,00011203	0,00011203	0,00011803
20.00	0,00023	0,00015205	0,00015806	0,00017407
25.00	0,00024	0,0002061	0,00021411	0,00022812
30.00	0,00034	0,00028018	0,00027418	0,00028219
35.00	0,00061	0,00042041	0,00040238	0,00040639
40.00	0,00076	0,00066702	0,00063693	0,00060886
45.00	0,00125	0,00107264	0,00107263	0,00101235
50.00	0,00255	0,00173288	0,00178324	0,00165021
55.00	0,00299	0,00283635	0,00303291	0,00281199
60.00	0,00616	0,00474268	0,00506743	0,00465643
65.00	0,00812	0,00865917	0,00998896	0,00945524
70.00	0,01503	0,01534377	0,01916616	0,0187142
75.00	0,02183	0,0259463	0,03395968	0,03358186
80.00	0,04386	0,04752027	0,07259509	0,07546624
85.00	0,06267	0,08180337	0,1376441	0,14791191
9094.00	0,09486	0,13404542	0,23521872	0,25660344
95115.00	0,17204	0,20570356	0,3595216	0,40167234

[4] The fact that the fitted curve does not follow the shape of the observed curve very well at young ages is due to the choice of the United Nations model life table, not to the Mortpak program.

At this time, an informed choice can be made between using the observed m(x) values from Table 1, fitted model equivalents from Figure 1 or other substitute m(x) values for the final extrapolations. Here, observed as well as fitted values can both be recommended. But when nationalities are subdivided into much smaller groups, the robustness of the observed values will drop, the observed mortality curve will become erratic. In that case, fitted values will probably be required to increase robustness.

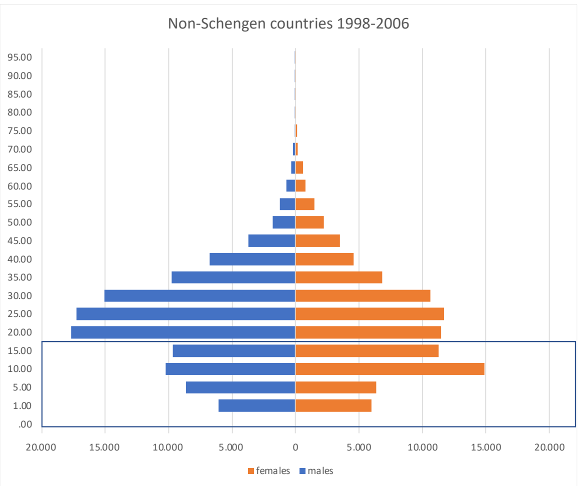
The next step in the calculations is counting the unlinked deaths by age and gender, and applying the extrapolation factors to the unlinked deaths in each age group. When these unlinked deaths at age x are noted as d(x), and the extrapolation factor = 1/ m(x), then multiplying d(x) * 1/m(x) = d(x)/m(x) or simply dividing the number of unlinked deaths by the mortality rate in each age group.

Table 3 shows calculation results, using the 3 component fitted m(x) values from Table 2, and the observed number of unlinked deaths (women, 1998-2006). In this case the extrapolated value is: (Deaths female / 3 comp fitted m(x))/9. Division by 9 controls for deaths being aggregated over 9 years. Note that about 500 unlinked deaths were observed over those 9 years of observation. The right-hand side of the graph shows the population pyramid with women in orange color, totaling at around 93.000.

As can be seen in the graph, the same type of calculation was performed for males. The male total was 109.000, making a total of both sexes combined of 203.000 non-registered mobile citizens of non-EU+ origin in the 1998-2006 period.

Table 3: Extrapolation based on 3-component fitted mortality rates + population pyramid

	Deaths female	3 comp fitted m(x)	extrapolation
.00	"377"	-	-
1.00	19	0,000352809	5983,7
5.00	8	0,000140049	6347,0
10.00	11	8,20168E-05	14902,1
15.00	12	0,000118033	11296,3
20.00	18	0,000174072	11489,5
25.00	24	0,000228125	11689,5
30.00	27	0,000282189	10631,2
35.00	25	0,000406386	6835,3
40.00	25	0,000608856	4562,3
45.00	32	0,001012346	3512,2
50.00	33	0,001650214	2221,9
55.00	37	0,00281199	1462,0
60.00	33	0,004656429	787,4
65.00	51	0,00945524	599,3
70.00	26	0,018714198	154,4
75.00	45	0,033581855	148,9
80.00	25	0,075466239	36,8
85.00	22	0,147911907	16,5
90.00	4	0,256603438	1,7
95+	5	0,401672342	1,4
Total	482		92679,5



Interpreting results

For correct interpretation of results, it must be remembered that what is actually estimated is person-years or average population. What is not estimated, is the number of individuals that have contributed to those person-years. In other words, the 93.000 person-years can theoretically have been produced by 93.000 people who were present all over the year, 186.000 people who were present 6 months each, or any other combination of people stay for complete years or shorter periods adding up to 93.000 person-years. This also means that the model does not estimate how many people exceed any threshold of duration of stay that could be set to label them as “usual residents”.

Step 2: Distinguishing between regular and irregular status population

As explained in the introduction, estimating the size of the **irregular** status population is achieved in a second step, which consists of subtracting different regular status populations from the overall results of the previous step. Therefore, data should be collected on the major groups of people who are not registered citizens, but have a regular legal status. Examples of such groups in Belgium are: diplomats, posted workers, visa waivered tourists, short-stay valid visa holders, and asylum seekers whilst in procedure. In other EU-countries, the list of major groups could be somewhat different.

In the Belgian example, extrapolations were not only performed for non-EU+EFTA (=“third country”) citizens, but also for EU+ citizens. These are rarely labeled “irregular” given free travel rights within the Euro-zone. In the EU+ case, irregular means non-compliant with citizen’s obligations of declaring their usual residence to local authorities and to the country’s population administration system. Contrary to non-EU+ citizens their basic right of being present within the borders of another EU+ country is not at stake.

Below, the main groups of regular visitors/residents are listed. Composition of this list and relative importance of each subgroup will vary across Europe.

Diplomats:

The EU-institutions in Brussels, NATO European headquarters, and a large diplomatic representation all contribute to a very large diplomatic community in Belgium. Two subgroups of people can freely choose not to be registered in the official population registry. One of these is referred to as “Protocol”. These people are mainly embassy and consular staff that is considered to be a protected population. This status comes with other privileges such as diplomatic immunity and administrative discretion. The second group is EU-personnel with similar rights of discretion. The EU itself reports that about 60.000 staff is working in Brussels based institutions. Until around 2017, nearly half of those (28.000) did not register, but numbers are going down. The latest figure would be around 11.000. The unregistered “Protocol” population is about 20.000 in size and is slowly increasing. Around

2016-2017, the total of Protocol + EU-staff is estimated at 45.000. Whilst EU staff can be composed almost exclusively of EU-nationals, Protocol population is much more diverse.

Visa waivered tourists:

Tourists from EU-countries + Norway, Iceland, Liechtenstein and Switzerland (EU+EFTA) can travel freely. Furthermore, other countries (mainly industrialized + Latin American) have signed bilateral agreements that relieve their citizens of visa obligations. Although there are some national differences, most EU-countries accept a very similar list of visa-waivered countries.

In Belgium, tourism statistics are kept counting numbers of nights spent in hotels and other similar infrastructure. Numbers are detailed by nationality, by motive (business/tourism) and also by number of nights per visit. This data can be converted into person-years, using an approximated 365 hotel nights = 1 person-year principle, or a variation on that theme.

Visa holders:

Since long-term visa holders are regular members of the official population, they must not be taken into account. Short-term visa holders are also regular, but non-members of official population. Numbers of short-term visa holders can be converted into person-years based on the 3 month validity of such visa: 4 visa = 1 person-year. It is likely that many visa are not used completely until their expiry date, or some not used at all. Therefore, the person-year equivalent could be raised (e.g. 5 visa = 1 person-years). A realistic estimation of Visa-effects would require more research.

Visa obliged tourists:

For visa obliged countries, separating tourism and business from other travel motives is a grey zone. In principle, tourism data on visa obliged countries of origin should not be taken into account. Otherwise, that would amount to double counting in the regular visitor group (once as a Visa-holder and once as a tourist spending registered hotelnights).

Asylum seekers during procedure

These data are publicly available and do not require any extra calculations. Latest figures for Belgium put them at around 62.000.

Posted workers

Some data on posted workers are publicly available, although not much detail is given in official statistics. However, a recent study by Leuven University was released after our estimations were published offering useful insights in the contribution of posted workers in the non-registered population. The study observed that numbers are increasing over the years. EU-nationals working in Belgium as posted workers were around 185.000 in 2016, which was a sharp increase from 112.000 in 2011. Posted workers also include non-EU nationalities mainly from continental Europe, but also from for instance Brazil. In Brussels, around 8% of the posted workers would be of non-EU+ origin (=third countries) in 2020. In the rest of Belgium that percentage is cited to be lower. From that information, around 15.000 third country posted workers can be presumed, working an average 105 days according to published statistics. In terms of person-years, that would amount to roughly 5.000 yearly equivalents. In Belgium, construction is the main sector of industry employing posted workers.

Final result:

All these groups of incoming mobile citizens were separated into EU- and non-EU origin. For some of the data cited above, published data allow direct sorting into such nationality groups (e.g. tourism statistics). For others, such as asylum seekers and issued visa, it was clear that it (quasi?) exclusively concerns third country citizens. For still others (EU-personnel) it could reasonably be assumed that it concerned EU-nationals. For the remaining groups, no detail on nationality was available ("protocol"-population).

As Table 4 illustrates, the number of regular visitors from third countries could thus be estimated with reasonable precision (in 2012-2016, we estimated their total around 83.000). For the EU-citizens this is far more difficult for several reasons. Firstly, EU-citizens can stay for up to 3 months in Belgium before being obliged to register. As mentioned before, extrapolation results are expressed in person-years, but not in terms duration of residence. We therefore do not know which fraction of EU-mobile citizens are regular in the sense of not overstaying that 3 month period. However, those staying in tourist infrastructure are considered to be in shorter stays, and therefore they are classified as regular visitors. But it is uncertain to which degree tourism data covers these populations, since a significant group of incoming visitors can be day travelers, or stay with friends or family. However, based on the data we could find (mainly hotel nights, EU-personnel data and the protocol register), and taking into account that it is almost certainly an underestimated number, around 77.500 regular EU+-citizens where estimated in 2012-2016.

Table 4: Regular mobile citizens

Description	Third countries	EU + EFTA
Protocol + EU-institutions *	10.000	35.000
Visa wavered tourist or business traffic **	8.900	42.500
C-type visa holders ***	14.000	-
Asylum seekers during application procedure	50.000	-
Posted workers	(5.000)	(????)
TOTAL REGULAR (excl. posted workers)	**82.900**	**77.500**

* 45.000 total is verified but distribution EU+ /non-EU+ is assumed
** Estimates are based on hotel nights and could therefore be under-estimated
*** Based on 3 months of stay per C-type visa

Discounting those (83.000 non-EU+ and 77.500 EU+) regular population totals from the previously estimated totals of all non-EU+ and all EU+ mobile citizens (i.e. 195.000 non-EU+ and 294.000 EU+ in 2012-2016) leaves a final estimate of 112.000 irregular third country citizens in 2012-2016, and 217.000 irregular EU+ citizens. The first estimate did not take into account posted workers as those data were not yet available at the time of publication. The second estimate does not properly take into account the 3 months of legal stay of Schengen country citizens and probably underestimates tourist traffic since it is not completely covered by hotel data. It is therefore likely to be overestimated.

SOME THOUGHTS ON ESTIMATING THE DISTRIBUTION OF DURATION OF STAY

Producing statistics on the usually resident population requires knowledge on duration of stay. As mentioned before, mortality-based extrapolation does not inherently offer that information. External sources for estimating duration of stay of irregular and unregistered populations are scarce. Surveys could be a very valuable source of information on the subject. But in Belgium, to our knowledge, no survey data have been collected containing this information.

For third country citizens residing in Belgium, the most promising path is probably the “declaration of arrival”, which is a form that third country foreigners entering the country are obliged to fill out at the municipality of their (temporary) residence, and that must be renewed every 3 months. But since the municipality checks status regularity, it does not capture overstay, and it is limited to those having a regular status on arrival or applying for such a status. It does, therefore, include asylum applicants. Also, the means of proving regularity of a temporary stay are somewhat liberal which is beneficial for coverage. Except for the passport stamp by border police received on entry, a few other types of proof of regularity are also accepted, such as proof of full official residency in a third country, or a valid traveling ticket bearing the applicants name. Similarly, third country relatives of Belgian citizens can obtain the declaration. However, documents replacing the passport stamp will be checked by a special department of border control authorities before the applicant is granted a positive decision for filing a declaration of arrival. It should be mentioned that hotel residents, hospitalized patients and imprisoned foreigners must not file a declaration of arrival. This is not problematic as those groups are largely covered by tourism data already mentioned in this guide. Similarly, foreign patients in hospitals are also statistically reported.

Declarations of arrival will probably cover the large majority of third country arrivals because of the legal obligation and because anyone aspiring future official residency, future visa renewal or visa prolongation will be highly motivated to register their stay at the municipality. Failing to do so compromises chances of getting a positive decision later regarding other applications.

Declarations of arrival do not cover EU+ visiting citizens. For them, registered hotel stays are likely to only cover short term visitors in a tourism or work-related stay (< 3 months). That coverage is also incomplete given that short-term house rental on the residential market is not covered, nor are guest stays with family or relatives or daytrips. It seems therefore that unregistered stays of EU+ citizens can only partially be analyzed in terms of duration of stay. But there is an alternative municipal registration document that applies to those who want to apply for a durable status following their legal obligation to do so after 3 months of uninterrupted stay.

To sum up, for all those in a regular status, and for all those applying for such a regular status, several types of time-stamped documents exist that in theory would allow a statistical follow-up in terms of durations of stay. The list of documents is much longer than just the ones mentioned here, and it also includes people temporarily holding a valid proof of arrival or residence, whilst not being eligible for a permanent stay, including irregularly entering people. The situation is too complex to be presented here in detail, but very useful information can be found at (page in Dutch):
https://www.vreemdelingenrecht.be/verblijfsdocumenten/papieren-verblijfsdocumenten-en

The page describes 11 time-stamped documents in terms of their validity, target population and additional legal aspects. The main bottleneck is the availability of statistical data on that stream of documents, but this needs to be enquired more deeply. It is not unlikely that some municipalities would allow insight into their migration related data on request.

Finally, after surveys and all types of cards, registrations and permits a third source of information could be derived from studying migration behavior of recent immigrants. We know from earlier work that registered EU+ residents in Belgium are highly internationally mobile compared to third country residents. In the 2011-2017 period, five year return rates for EU+ nationalities were around or above 50%, which was about twice as high as for most low-income countries. On the other hand, non-EU+ western countries (U.S., Japan, Canada,...) were even higher at 65% return rates. But whilst interesting and relevant, information on re-migration behavior in registered regular residents is unlikely to be transferable to those in either an unregistered or an irregular status. It does seem fair however to assume that the non-tourist fraction of unregistered mobile citizens from EU+ countries is considerably more mobile than the third country fraction.

ABOUT THE AUTHOR

Johan Surkyn is a sociologist and demographer at Interface Demography, Vrije Universiteit Brussel (VUB).

COPYRIGHT CLAUSE

THE MIRREM CONSORTIUM

University for Continuing Education Krems (Coordinator)

European University Institute

University of Osnabrück

University of Maastricht

University of Turku

Complutense University Madrid

Hellenic Foundation for European and Foreign Policy (ELIAMEP)

University of Milan

University of Potsdam

Platform for International Cooperation on Undocumented Migration (PICUM)

International Centre for Migration Policy Development (ICMPD)

Migration Policy Institute Europe (MPI-E)

University of Warsaw

Vrije Universiteit Brussel

Instituto Universitário de Lisboa

Associated Partners:

Toronto Metropolitan University

University of Leicester

University of Oxford