

University of Groningen

## Emigration Rates From Sample Surveys

Willekens, F.J.; Zinn, S.; Leuchter, M.

*Published in:*  
Demography

*DOI:*  
[10.1007/s13524-017-0622-y](https://doi.org/10.1007/s13524-017-0622-y)

**IMPORTANT NOTE:** You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

*Document Version*  
Publisher's PDF, also known as Version of record

*Publication date:*  
2017

[Link to publication in University of Groningen/UMCG research database](#)

*Citation for published version (APA):*

Willekens, F. J., Zinn, S., & Leuchter, M. (2017). Emigration Rates From Sample Surveys: An Application to Senegal. *Demography*, 54(6), 2159-2179. <https://doi.org/10.1007/s13524-017-0622-y>

### Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

### Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

# Emigration Rates From Sample Surveys: An Application to Senegal

Frans Willekens<sup>1,2</sup> · Sabine Zinn<sup>3</sup> ·  
Matthias Leuchter<sup>4</sup>

Published online: 16 October 2017

© The Author(s) 2017. This article is an open access publication

**Abstract** What is the emigration rate of a country, and how reliable is that figure? Answering these questions is not at all straightforward. Most data on international migration are census data on foreign-born population. These migrant stock data describe the immigrant population in destination countries but offer limited information on the *rate* at which people leave their country of origin. The emigration rate depends on the number leaving in a given period and the population at risk of leaving, weighted by the duration at risk. Emigration surveys provide a useful data source for estimating emigration rates, provided that the estimation method accounts for sample design. In this study, emigration rates and confidence intervals are estimated from a sample survey of households in the Dakar region in Senegal, which was part of the Migration between Africa and Europe survey. The sample was a stratified two-stage sample with oversampling of households with members abroad or return migrants. A combination of methods of survival analysis (time-to-event data) and replication variance estimation (bootstrapping) yields emigration rates and design-consistent confidence intervals that are representative for the study population.

**Keywords** Emigration rate · emigration data · sampling · bootstrap · Senegal

---

✉ Frans Willekens  
F.J.C.Willekens@rug.nl

<sup>1</sup> Netherlands Interdisciplinary Demographic Institute (NIDI), Lange Houtstraat 19, 2511CV The Hague, The Netherlands

<sup>2</sup> Population Research Centre, University of Groningen, Groningen, The Netherlands

<sup>3</sup> Leibniz Institute for Educational Trajectories, University of Bamberg, Bamberg, Wilhelmplatz 3, 96047 Bamberg, Germany

<sup>4</sup> Department of General, Thoracic, Vascular and Transplantation Surgery, University of Rostock, Schillingallee 35, 18057 Rostock, Germany

## Introduction

Approximately 3.4 % of the world population lives in a country other than their country of birth, totaling 244 million people in 2015 (United Nations Department of Economic and Social Affairs, Population Division 2016). In 2015, emigrants transferred US\$601 billion in remittances, including US\$441 billion to developing countries—nearly three times the amount of official development assistance (World Bank 2016). In 2015, 1.2 billion tourists travelled to an international destination (United Nations World Tourism Organization 2016). The number of international arrivals increases a steady 4 % every year, and some of these travelers will want to settle abroad. A worldwide Gallup survey in 2005 found that 14 % (or 630 million) of the world's adults (aged 15+) say that they would like to emigrate if they could (Esipova et al. 2011), but only 3 % of them indicated that they started making preparations to leave. Although a considerable proportion of people desire to emigrate, few do because the decision to emigrate is a complex, multistage process (Klabunde et al. *Forthcoming*; Willekens 2017).

Although the root causes of emigration are relatively well understood, little is known about the level of emigration. The emigration rates by country are unknown. Emigration is the component of population change that is the most difficult to estimate because emigration is not accurately registered, and censuses and most surveys cover the resident population and exclude emigrants. Surveys may record emigration intentions, but intentions are often not good predictors of behavior. Despite this limitation, emigration research has often focused on intentions rather than actual behavior (see, e.g., Dibeh et al. 2017; Van Dalen and Henkens 2007, 2013). Surveys among immigrants have often recorded reasons for leaving the country of origin, but such data do not help much to explain emigration because of selection bias. Stayers are excluded from immigration surveys, but their characteristics and decision processes are essential for understanding why some individuals emigrate while others stay. To study the determinants of emigration and to infer the influence of migration on other demographic events (such as marriage and childbirth), outgoing migration should be properly defined and measured and should be related to the population at risk and the duration of exposure.

Reliable estimates of properly defined emigration rates are essential for advancing demographic modeling and projection because they permit replacing net migration by inflows and outflows separately, a long-standing goal in modeling (Bilsborrow 2012; Rogers 1990). Flow models are superior for studying migration dynamics (for a recent review of flow models, see Willekens 2016). In population projections, most statistical offices and the United Nations use net migration and rely on the inertia of net migration because push and pull factors are not measured, or they are too difficult to predict (Azose et al. 2016). At times of change, when forecasts are needed most, trend extrapolation is not a reliable strategy. Instead, up-to-date estimates of emigration rates and their determinants are needed.

In this article, we present a method for estimating emigration rates and their precision from survey data. The emigration rates are occurrence-exposure rates that are obtained by dividing emigration counts and the population at risk, weighted by the duration of exposure—a common practice in formal demography and event history analysis (see, e.g., Aalen et al. 2008; Preston et al. 2001; Willekens 2014). The estimates and their precision account for the sample design. The method is applied to

data of the Migration between Africa and Europe (MAFE) survey, which was established using a multistage, stratified sampling design with oversampling of migrant households. The method also accounts for nonresponse. This method is an important tool for demographic estimation (Moultrie et al. 2013).

## Emigration Rates: Conceptual and Estimation Issues

*Emigration* is usually defined as a change of usual residence to a different country (United Nations 1998: 18, box 1). For practical applications, such as the population census, a person is a usual resident of a country if he or she has lived in that country for at least 12 months or intends to stay for at least one year. The concept involves a threshold of time spent in the country of destination. The United Nations introduced the threshold to make data internationally comparable and to distinguish long-term migration from short-term migration (visits lasting 3–12 months) and visits of less than 3 months. The migration concepts used around the world reveal many nuances (Bilsborrow et al. 1997) that complicate the combination and analysis of the data.

In this section, we briefly review sources of data on emigration and approaches used to determine the target population and to estimate emigration rates.

### Sources of Emigration Data

The direct measurement of emigration involves recording the event at border crossing or shortly before leaving the country. Countries with a population register (typically countries in Europe) require their residents to deregister before leaving. Most emigration is measured indirectly by comparing the country of current residence and the country of residence at some previous point in time, usually recorded at birth or one or five years ago through censuses and surveys. A cross-classification of the population by country of current residence and country of birth gives the native-born population and the foreign-born population (migrant stock) by country of origin. If such data are available for all countries, determining lifetime emigration for all countries is possible.

Countries with a population register can record emigration if residents deregister before they leave the country. However, underregistration of emigration is a problem, revealed when the population census or the administration finds out that individuals no longer reside in the country. In some countries, such as Poland, residents who migrate to another country do not need to deregister unless they intend to stay abroad permanently. The 2011 census revealed that 1.9 million residents of Poland (5 % of the population) were living abroad for more than three months (Wiśniowski 2017).

A few countries, such as Saudi Arabia, require foreigners to have an exit visa before they can leave. Most countries require foreigners to have a visa or a permit to stay in the country and expect them to leave voluntarily before the authorization ends. If departure is not registered, outward mobility is not measured. In addition, people may stay longer than the authorized duration. The United States and the European Union are introducing exit registration systems to identify visa overstayers (Orav and D'Alfonso 2016; U.S. Department of Homeland Security 2016).

Some countries share data on border crossing so that the record of entry into one country establishes an exit record from the other. For example, the United States and

Canada share data on foreigners and nationals crossing their border (Canada Border Services Agency 2016). Nordic countries (Sweden, Norway, Denmark, Finland, and Iceland) share data, too. A resident of one of the Nordic countries who registers in another Nordic country is automatically removed from the register of the country of origin. Romania cooperates with Italy and Spain, two main destination countries of its emigrants, to produce emigration statistics (Pisica 2016).

Sample surveys are the main source of emigration data today. The United Kingdom relies on the International Passenger Survey (IPS) to determine the number of people leaving the country and their intended duration of stay abroad (Office of National Statistics 2017). The Survey of Migration at Mexico's Northern Border (Encuesta sobre Migración en la Frontera Norte de México, EMIF-Norte) is designed to determine the magnitude and characteristics of the migration flows of Mexican workers to the United States (El Colegio de la Frontera Norte 2017). Using the EMIF data, Rendall et al. (2009) found that the male emigrants identified by EMIF are double those reported by U.S. data sources (in which they appear as immigrants). The authors attribute this to a better capture of unauthorized and circular migrants in the EMIF.

Household surveys and labor force surveys are frequently used sources of data on immigration (see, e.g., Martí and Ródenas 2007, 2012; Rendall 2003). They can also be used as sources of data on emigration, provided that they collect information on household members abroad (from proxy respondents) and return migrants.

Surveys that use respondents as proxy informants to collect information on nonresident household members are *network* or *multiplicity surveys* (Kalton 2009). Woodrow-Lafield (2010) reviewed multiplicity surveys in the United States used to determine emigration. Surveys in the Households International Migration Surveys in the Mediterranean Countries (MED-HIMS) program include modules for household members abroad and return migrants (Eurostat 2017).

The International Labour Organisation (ILO) developed a migration module for national labor force surveys (LFS) to collect information on workers abroad and return migrants. The module was introduced in several countries, including Egypt, Thailand, Ecuador, and Ukraine (Bilsborrow 2016b; International Labour Organization 2013). The data offer insight in the proportion of migrant workers and return migrants and the causes and consequences of emigration. Recently, Wiśniowski (2017) used the LFS of Poland and the United Kingdom to estimate the migration flow from Poland to the UK. The LFS of Poland includes information on household members abroad.

Household and labor force surveys have a crucial shortcoming, however. The number of emigrants (and immigrants) captured in the surveys is usually very small, which is particularly problematic for estimating flows during a given year or period. Martí and Ródenas (Martí and Ródenas 2007; 2012) concluded that LFS data can be used for compiling statistics on stocks on migrants (lifetime migrants) but not on flows.

The solution is to design a sample strategy that ensures that the sample includes a sufficient number of households with migrants and/or return migrants (Groenewold and Bilsborrow 2008). That strategy is adopted in specialized migration surveys, such as the NIDI/Eurostat push-pull migration project (Schoorl et al. 2000) and the MAFE survey (Beauchemin 2015). In both surveys, information on nonresident household members, including contact details, is collected. Emigration rates estimated from these survey data are not representative for the study population unless the sample design is accounted for. In this article, we present methods for estimating emigration rates and

quantifying their precision by taking into account the complex sample design and the selectivity resulting from nonresponse.

### Emigration Rates and Emigration Probabilities

The emigration rate relates the number of emigrations to a target population. Definitions of emigration vary considerably between countries (Jensen 2013), and a harmonized methodology for estimating emigration does not exist. As a consequence, emigration rates are not comparable. In this section, we illustrate the variety of emigration rate concepts used. Most are not consistent with the rate concept in demography and event history analysis. The concept that we propose and use in this article is fully consistent, though.

The Organisation for Economic Cooperation and Development (OECD), which maintains one of the most important databases on international migration in the world, defines the *emigration rate* of a country as the proportion of the population born in that country residing abroad. The OECD uses census data on populations by country of residence and country of birth. The denominator of the emigration rate is the population born in a country, including those living abroad (*expatriates*) and excluding the foreign-born population currently residing in that country. When native-born and foreign-born populations cannot be separated, the OECD takes the total resident population as the denominator of the emigration rate (OECD 2000). According to this definition, in 2010 and 2011, 2.4 % of the population born in Africa was living in an OECD country (OECD and United Nations Department of Economic and Social Affairs (UNDESA) 2013; see also Arslan et al. 2016). Kim and Cohen (2010) and Westmore (2015) used census data on foreign-born population by country of residence to predict the number of lifetime migrants in a given year from demographic and economic variables in that year. A weakness is that current or recent drivers of migration can hardly explain lifetime migration and migrant stock unless the drivers and their effects remain stable over a long time. In a study of emigration to the United States, Clark et al. (2004) and Hatton and Williamson (2011) used U.S. census data on migrant stock, too. Hatton and Williamson used a two-step procedure to derive the emigration rate from a source country to the United States. In the first step, they calculated the number of immigrants in the United States during the five years prior to the census who were born in the source country, divided by the population in the source country at the beginning of the five-year period. The source country is the country of birth and not the country of last residence or residence five years prior to the census. In a second step, they used a regression model to predict migration during a five-year period from key economic and sociodemographic variables measured at the beginning of the five-year period.

Abel (2013) and Abel and Sander (2014) used census data on foreign-born population to estimate recent migration flows and the proportion of residents that emigrate in a five-year period. A discussion of the method is beyond the scope of this article. The flow estimates that they obtained revealed the unexpected: contrary to the common belief, international migration flows did not increase over the past two decades. Rather, the volume of global emigration flows declined from 0.75 % of the world population moving over the five-year period from 1990 to 1995 to 0.57 % from 1995 to 2000.

Methods proposed in the literature for estimating emigration rates differ significantly. The golden rule is to approximate emigration counts and populations at risk as good as possible with the data at hand. Jensen (2013) reviewed a number of methods for estimating the number of emigrants and made a few references to the estimation of emigration rates. Van Hook et al. (2006) and Van Hook and Zhang (2011) estimated the emigration rate of the foreign-born population in the United States from Current Population Survey (CPS) data. The CPS is a monthly conducted survey with a quasi-longitudinal design in which the same household is included in the survey for four consecutive months and again for four additional consecutive months in the following year. In March of every year, the CPS includes an additional module on emigration and other reasons for attrition. Leach and Jensen (2014) proposed an occurrence-exposure rate to quantify emigration intensities in the United States. For estimation, they used the American Community Survey to compute the rate as the ratio of (net) emigrations and an approximation of the population at risk of emigrating. Schoumaker and Beauchemin (2015) used the MAFE data on migration between Africa and Europe to estimate levels of emigration. They used data on counts and exposures, but they did not estimate emigration rates. Instead, they estimated conditional probabilities of emigration at a given age, given that the event has not already occurred, using a discrete-time event history model (Allison 1982). Hanson and McIntosh (2010) measured the emigration rate as the proportion of a birth cohort that has emigrated. The authors used the Mexican population censuses of 1960, 1970, 1990, and 2000. Their target population was composed by the cohorts in the year when they are observed for the first time in the censuses. Take as an example the cohort of 8-year-old boys born in the state of Zacatecas and appearing in the Mexican census for the first time in 1960. By observing how many of them are aged 18 in the 1970 census, aged 38 in the 1990 census, and aged 48 in the 2000 census, the authors were able to construct a series of 10-year emigration rates specific to age, state of birth, and year of birth (and other personal attributes observed in the census).

### **Estimation of Emigration Rates From Household Surveys, Illustrated With MAFE**

The MAFE project involved the collection of data in three African countries (Senegal, Ghana, and Democratic Republic of Congo) and six destination countries in Europe, including Italy, France, and Spain (Beauchemin 2015). To illustrate our estimation method, we use the MAFE data of Senegal. The Senegal survey, organized in 2008, covered the Dakar region and comprised a household survey and an individual biographic survey. The survey uses a multistage sample with the population stratified based on migrant status and with oversampling of migrant households (Schoumaker and Mezger 2013). The sampling procedure implies different inclusion probabilities of households. Furthermore, unit nonresponse has to be accounted for. In sum, 87 % of the households contacted participated in the survey (Razafindratsima et al. 2011). To address nonresponse and simultaneously account for the sample design, the MAFE survey contains nonresponse-adjusted sampling weights. Hence, the sample is not self-weighted, and the nonresponse-adjusted sampling weights of households have to be used to obtain emigration rates that are representative for the study population.

## Survey

The sampling frame of MAFE was the population of the Dakar region in the 2002 census of Senegal, updated in 2008. The Dakar region consists of 2,109 census districts (enumeration areas). These form the primary sampling units. In the first sampling stage, census districts were collapsed into 10 strata of equal size. Nine strata comprised 211 districts, and one stratum consisted of 210 districts. In each of the 10 strata, 6 districts were sampled with a probability proportional to the number of households in every district. Now, all the households in the selected districts were listed, and their “migration status” was defined: that is, it was determined whether at least one migrant (currently abroad) lived there or lived abroad and returned to Senegal. On the basis of this information, a second stratification level was formed comprising two strata: households with and households without migrants. In the second sampling step, in each of the 10, first-level strata households were randomly sampled. They form the secondary sampling units. To ensure enough migrant households in the sample, the selection probabilities of households in the second-stage stratum, *households with migrants*, were at the average higher than in the opposed stratum, *households without migrants*. That way, an oversampling of migrant households was established. It should be noted that the sample is representative at the time of the survey, and not necessarily for the years for which data are collected in the retrospective survey.

In each district, 22 households were selected at random, 11 from each household type. If less than 11 households of a given type were available, the remaining households were selected from households of the other type. For instance, if only 4 households with migrants were found in a district, all of them were selected, and the 18 remaining households were selected among the households without migrants. A total of 1,320 households constituted the household sample. Of these, 1,141 households were interviewed: 458 nonmigrant households, 205 households with at least one returnee, 617 households with at least one current migrant, and 139 households with returnee(s) and current migrant(s) (Schoumaker and Mezger 2013).

The 1,141 households interviewed included 12,350 individuals in the household roster, which includes household members and people related to the household and with whom the individual maintained regular contact but who were not household members at the time of the survey: for example, grandchildren born in Europe and living in Europe. In this article, all persons included in the household roster are referred to as *household members*, for convenience. For each household member, the migration experience was recorded, with a migration event defined as a stay abroad for at least 12 months. The following persons living abroad at the time of the survey could be declared as *household migrants*: the head’s children, his/her spouse(s), and other relatives of the head or of his/her current spouse with whom the household head had regular contacts within the last 12 months. These are household members with potentially a migration experience. Of the 12,350 household members, 1,689 lived abroad for more than 12 months. Hence, 13.7 % of the household members had a migration experience. Clearly, that proportion cannot be extrapolated to the population of the Dakar region because households with migrants were oversampled. The date of birth was not recorded for 242 of the 12,350 household members; they are excluded from our analysis.

Emigration experience is quantified on the basis of a screening question (A12) in the household survey indicating whether an individual has lived for at least one year out of

Senegal (whatever the time of departure) and a question on the year of first departure (A13a). Emigration rates are estimated from data on first departures. Because MAFE is a retrospective survey, households consisting of only emigrants are not included in the sample; instead, the survey comprises households with a head who never lived abroad or who lived abroad and returned to the Dakar region. Because the total number of household heads who emigrated is not known, the number of household heads at risk of emigration cannot be determined.<sup>1</sup> The only category of household members who are registered—whatever their place of residence at survey time was—is the children of the household heads. Therefore, following Schoumaker and Beauchemin (2015), we rely on this category to estimate the emigration rate. However, excluding household heads introduces undercoverage bias such that older adults are underrepresented, making it difficult to estimate emigration rates for the past decades from the retrospectively collected MAFE data. Ideally, we should add information on household heads who emigrated, using data collected in destination countries or proxy respondents (Bilsborrow 2016a:127–129).

In the estimation, young children are excluded because they do not emigrate independently from their parents. Following Schoumaker and Beauchemin (2015), we compute emigration rates for the ages between 18 and 39 and for periods from 1975 to the survey date in 2008; that is, we consider first emigrations of persons who were aged 18–39 in 1975–2008. Likewise, the population at risk consists of persons who were aged 18–40 at least sometime in the period from 1975 to the survey date. We also exclude children who were younger than 18 in 2008 or who emigrated before reaching age 18. The MAFE data include the age at emigration in years, which is not directly observed but is computed as the difference between the year of emigration and the year of birth. Migrant status is collected for surviving and deceased household members. The migrant status of deceased children is known because the publicly available data file includes the household status of deceased children of the head of the household. A total of 153 children of heads of households had died, some at advanced ages. These children are included in our analysis. Schoumaker and Beauchemin (2015) found that including deceased children in their analysis resulted in somewhat (but not significantly) lower estimates of the emigration probabilities. This finding indicates that surviving children had a somewhat higher propensity to emigrate than children who did not survive until survey date, probably because of differences in health status.

## Method

Exposures and events (emigrations) are measured within the defined age range and time frame. Children born between 1936 and 1990 (3,392 children) enter observation sometime between 1975 and the survey date. Children born between 1936 and 1957

<sup>1</sup> Orrenius and Zavodny (2005) used similar data from the Mexican Migration Project in a study of undocumented migrations to the United States. They included heads of households, who emigrated and returned to Mexico (and are included in the survey) and excluded heads of households who emigrated permanently. In the analysis, they used two separate samples: (1) heads of households, and (2) sons of heads of households. The authors estimated the Cox proportional hazard model to determine effects of covariates and economic conditions in Mexico, but they did not consider the magnitude of emigration rates because they disregarded the baseline hazard. They considered relative effects only.

(68 children) enter observation in 1975. Those born between 1958 and 1990 enter the observation window in the year they reached age 18. An individual born in 1957 on the day and the month of the survey date reached age 18 in 1975 and contributes the maximum of 22 years, until age 40 in 2007, provided that the individual survived and did not emigrate. An individual born in 1987 who did not emigrate contributes three years to the person-years at risk of emigration: from 2005 (the year in which age 18 is reached) to the survey date. Because the date of interview is known, that date is used to determine exposure time. The date is converted into decimal year. If the individual emigrated in 2005—that is, in the year he reached age 18—the duration of exposure is set to 0.5 years to prevent zero exposure time. Accordingly, an individual who reached age 40 in 1985 and lived his entire life in Senegal contributes 10 years but at different ages than the previous individual. An individual who left Senegal at age 25 in 2004 (irrespective of destination) and returned three years later (in 2007) contributes seven years (between ages 18 and 25) before the first emigration. Here, the time spent in Senegal after the return migration is not considered because the emigration rates are based on first emigrations. Residents of Senegal aged 18–39 in 1975 (born between 1936 and 1957) enter the population at risk in 1975. Individuals born between 1958 and 1990 entered the population at risk sometime during the period 1975–2008: namely, when they reached age 18. Individuals are considered to have left the population at risk upon reaching age 40 or in 2008, whatever came first. Nine children emigrated before 1975, and they are excluded. The number of surviving children included in the estimation of emigration rates is 3,383. They were aged 18–39 sometime during the period 1975–2008.

Emigration rates are obtained by dividing event counts and the population at risk (risk set), weighted by the duration at risk. The rate is an occurrence-exposure rate. It is the same as the one estimated by an exponential model or a Poisson regression model with exposure time as offset (Aalen et al. 2008; Willekens 2014). Emigration rates are estimated for males and females separately. Age-specific emigration rates are estimated by specifying an exponential model with piecewise-constant rates. Because the recorded number of emigrations is relatively low, broad age groups are considered. The estimation method involves counting the emigrations and measuring exposure time during the age interval (and the observation window) (Blossfeld and Rohwer 2002: chapter 5; Broström 2012; Willekens 2014:102, 160). Because dates of events in MAFE are reported in calendar years and ages are reported in completed years, the piecewise-exponential model with year as the time unit gives the same result as the nonparametric Nelson-Aalen estimator (Aalen et al. 2008). Nelson-Aalen estimators of emigration rates were computed, but the results are not shown here. Another method for estimating age-specific emigration rates from survey data is the Cox model without covariates and with age as a duration variable. In the absence of covariates, the baseline hazard gives the age-specific emigration rates. To obtain the rates for population groups, such as males and females, the sample population may be stratified or sex may enter the Cox model as a stratification variable.

To estimate the emigration rate, individual exposure times must be determined. Individuals enter the observation window at different ages. The observations are left-truncated (Klein and Moeschberger 2003). They are also right-censored at survey date, at date (i.e., year) of death, or on their 40th birthday. The counting process perspective offers a natural approach to deal with left-truncated and right-censored data (Aalen

et al. 2008). The approach is quite simple if for each individual, exposure during the observation window is described by three variables: starting time, ending time, and reason for ending (emigration or censoring). The format is known as the *counting process format*.

After emigration rates have been estimated, the empirical survival function (i.e., the probability of staying in Senegal at a given age between 18 and 40) and the cumulative incidence of emigration (i.e., the probability of emigrating between age 18 and any older age) can easily be derived. Schoumaker and Beauchemin (2015) estimated the cumulative incidence of emigration for the age interval 18–40 as the probability that a (synthetic) person in the Dakar region at age 18 leaves Senegal before age 40. In our upcoming Results section, we compare our estimates with those obtained by Schoumaker and Beauchemin.

If the sampling design is neglected during estimation, deriving variance estimates is straightforward. See, for example, Aalen (1978) for the derivation of the variance of the Nelson-Aalen estimator and Hoem et al. (1976) for the derivation of the (asymptotic) variances corresponding to occurrence/exposure rates.

The situation differs if the sampling design is accounted for—which, in this analysis, should definitely be the case for two reasons. First, the sample of the MAFE data has been established using a stratified two-stage sampling design. Second, households with migrants have been oversampled. Both features imply (possibly very) different inclusion probabilities of sampling units and also clustering of the data. Furthermore, 13.6 % of all sampled households did not participate in the survey. This fraction is assumed to be a nonrandom subgroup of the initial sample (Schoumaker and Mezger 2013). Thus, any computation of emigration rates under negligence of the sampling design and the nonresponse might produce considerable bias when extrapolated to the population level.<sup>2</sup>

Using nonresponse-adjusted sampling weights is a simple way to counteract this problem. The household survey of the MAFE data contains such weights for each sampling unit. Here, in accordance with common practice, a sampling weight is defined as being the inverse product of the inclusion probability of a sampled household and its response probability. More details on the derivation of sampling weights in the MAFE data are given in Schoumaker and Mezger (2013). A general description of how to weight multistage sampling designs is given in, for example, Valliant et al. (2013), Kish (1995), and Heeringa et al. (2010).

In the present case, the use of sampling weights allows estimating emigration rates for the whole population in the Dakar region. We apply weights to emigrations and to exposure time. If a subject in the sample emigrated, the contribution to the event count is 1 (corresponding to one event) multiplied by the sample weight attached to that individual. Because of the oversampling of migrant households, a migration observed in a nonmigrant household receives a higher weight than a migration of a member of a migrant-household. The weighted contribution of a subject to exposure time is the actual duration of exposure in years multiplied by the appropriate sample weight. One year of observation of a member of a nonmigrant household receives a higher weight than one year of observation of a member of a migrant household.

<sup>2</sup> Several simulation studies have underpinned the need of correct variance estimation under complex sampling designs and in the presence of nonresponse. Examples are Canty and Davison (1999), D'Arrigo (2011), and Wagner and Eckmair (2006).

Schoumaker and Beauchemin (2015) used survey weights in their estimation of emigration probabilities for the Dakar region. Thus, they obtained design-based point estimates. However, without accounting for the two-stage sampling design of the MAFE data (and thus for their cluster structure), every observation in the person-period data set is treated as independent. This assumption is unrealistic and very likely increases the risk of underestimating the variability of the estimates.

The derivation of appropriate variance estimates necessitates special procedures accounting for the complexity of the survey data considered. Popular methods to do so are Taylor series linearization and replication methods (Lee and Forthofer 2006; Wolter 2007). Taylor series linearization is well suited to statistics that have a theoretical derivation of a variance formula, such as the coefficients of generalized linear regression models. In essence, replication methods conduct variance estimation by selecting from the overall sample a set of dependent subsamples. The sampling variance of the overall estimate—derived by computing parameter estimates from each subsample and calculating the variability between the subsample estimates—reflects the initial sampling process. A prerequisite of replication methods is that subsamples have to be formed such that each subsample has the same structure as the parent sample. Jackknife repeated replication, balanced repeated replication, and bootstrapping are the common replication methods that apply to stratified multistage sampling designs. A wide range of studies have noted pros and cons for all these methods (see, e.g., Heeringa et al. 2010; Lee and Forthofer 2006; Rust and Rao 1996; Wolter 2007).

Here, for sake of convenience, we use bootstrapping to derive confidence intervals. Because bootstrapping is easy and intuitive to use, it is applied in a wide range of research. The distinct steps of a bootstrapping estimation procedure applying to the stratified multistage sampling design of the MAFE data can be summarized as follows (see Wolter 2007: chapter 5.4).

First, we construct bootstrap samples ( $N = 500$ ).<sup>3</sup> A sample is restricted to children of household heads. We sample (with replacement) from each of the 10 strata of the household sample  $n_h^*$  districts, where  $n_h^* = n_h - 1$ , and  $n_h$  denotes the number of districts in stratum  $h$  ( $h = 1, \dots, 10$ ). In the MAFE data,  $n_h = 6$ ; hence, we randomly select five districts with replacement. Second, we apply the ultimate cluster principle<sup>4</sup> (Kalton 1979; Lee and Forthofer 2006) by taking all households of the  $n_h^*$  districts selected into the bootstrap sample. Thus, for each bootstrap sample  $S_b = 1, \dots, N$ ,  $S_b$  consists of  $z_{hi}$  households with  $i = 1, \dots, n_h^*$  and  $h = 1, \dots, 10$ . Third, for each bootstrap sample, bootstrap estimates  $\hat{r}_{t,a}^b$  of the emigration rates corresponding to time  $t$  and to age  $a$  are computed. The observation (emigration and exposure) on each respondent included in the subsample is weighted by the weight of his or her household in the original MAFE sample. The household weights were estimated by Schoumaker and Mezger (2013) and are included in the data. Fourth, for each estimate  $\hat{r}_{t,a}$  of an emigration rate, an accordant

<sup>3</sup> Here,  $N = 500$  constitutes a compromise between having enough sample points to derive a meaningful empirical distribution and runtime reduction.

<sup>4</sup> It is common practice to treat after the primary sampling unit all second and subsequent stages as being one. This practice is predicated on the fact that the sampling variance can be approximated adequately from the variation between the totals of the primary sampling units when the first-stage sampling fraction is small (which is usually the case).

bootstrap distribution  $\hat{r}_{t,a}^1, \dots, \hat{r}_{t,a}^N$  is determined. Given the data at hand, these distributions are not symmetric but skewed because the sample size of the data used is too small to ensure (asymptotic) normality of the estimator. Hence, we use the basic or pivotal bootstrap (Davison and Hinkley 1997:194) to derive confidence intervals. In detail, the bounds of the confidence intervals are derived from the 2.5 % and the 97.5 % percentiles  $q_{0.025}$  and  $q_{0.975}$  of the empirical bootstrap distributions:

$$2\hat{r}_{t,a} - q_{0.025}, 2\hat{r}_{t,a} - q_{0.975}.$$

## Results

The overall emigration rate of the population in the Dakar region is reported first. The rate is compared with the emigration rate from the Dakar region estimated by Lessault and Flahaux (2013), who used data from other sources. The trend in emigration rates is considered next. Finally, age-specific rates are presented. Schoumaker and Beauchemin (2015) used age-specific one-year emigration probabilities to compute the cumulative probability that an 18-year-old resident of the Dakar region emigrates before age 40. We compute cumulative probabilities from emigration rates (see, e.g., Klein and Moeschberger (2003:23) and compare the results with those that Schoumaker and Beauchemin (2015) obtained.

### *Emigration Rates by Sex*

As explained in the Survey section, the population at risk of emigration consists of children of household heads in the Dakar region. A total of 3,383 children of household heads were aged 18–39 between 1975 and the survey date. Of them, 89 left Senegal before the age of 18; they are excluded. Hence, 3,294 persons contribute to exposure time. The large majority contribute exposure but no event; only 314 emigrated during the observation window. The 3,294 persons contributed a total 33,543 person-years of exposure between 1975 and the survey date in 2008. The average emigration rate is therefore  $314 / 33,543 = 0.0094$ , or 9.4 per thousand person-years. The 95 % confidence interval of the emigration rate is (0.0082, 0.1200), disregarding the sampling design and the significant proportion of nonresponse in the household sample. On average, a subject included in the sample was exposed for 9.92 years between 1975 and 2008 and between ages 18 and 40.

Obviously, this emigration rate is overestimated because households with at least one migrant were oversampled. If each emigration is weighted by the weight of the emigrant's household, the weighted number of emigrations drops to 231.4. If years of exposure are weighted by the household weight, the duration of exposure reduces to 30,039 person-years. Hence, the emigration rate that results is  $231.4 / 30,039 = 0.0077$ , or 7.7 per thousand person-years. That rate is the unbiased estimate of the emigration rate of the population in the Dakar region. The 95 % confidence interval of the emigration rate, obtained by bootstrapping ( $N = 500$ ) with the sampling design accounted for, is (0.0064, 0.0101).

Males have a considerably higher emigration rate than females: 0.0089 for males (with a 95 % confidence interval of (0.0075, 0.0120)) and 0.0065 for females (with a

95 % confidence interval of (0.0038, 0.0091)). The difference is, however, not statistically significant at the 95 % level. The difference is statistically significant near the 90 % level; the 90 % level confidence intervals are (0.0083, 0.0118) for males and (0.0043, 0.0088) for females. If sampling weights are disregarded, the emigration rates are considerably overestimated: 0.0114 for males (with a 95 % confidence interval of (0.0107, 0.0147)) and 0.0073 for females (with a 95 % confidence interval of (0.0055, 0.0099)).

The weighted emigration rate is consistent with the rate that Lessault and Flahaux (2013) estimated for the population of Senegal aged 18 using the 2002 census and the Study of Migration and Urban Development in Senegal (EMUS) of 1993 (Direction de la Prévision et de la Statistique 1995). There, the emigration rate was estimated as the ratio of the number of persons who left Senegal during the five years prior to the observation (census or EMUS survey) and the population at time of observation, divided by 5. The estimates for the Dakar region that Lessault and Flahaux (2013) obtained differed considerably between the data sources. The 2002 census estimate was 0.0095, and the 1993 EMUS estimate was 0.0065. The authors did not provide standard errors or confidence intervals.

### *Emigration Rates by Period*

The emigration rate varied over time, as shown in Table 1. It was highest during the period 1975–1980, declined until the early 1990s, and increased slightly thereafter. The high female emigration rate before 1985 is remarkable. The emigration rates have large

**Table 1** Emigration rates and probabilities that 18-year-old will emigrate before age 40, by period and sex

Period	Number of Migrations (weighted)	Exposure (weighted)	Emigration Rate (and 95 % interval <sup>a</sup> )	Probability That 18-Year-Old Will Emigrate Before Age 40 (%)
<b>Males</b>				
1975–1984	6.4	654.6	0.0098 (0.0079, 0.0191)	19.4
1985–1994	28.2	3,019.1	0.0093 (0.0084, 0.0118)	18.6
1995–2008	98.3	11,274.8	0.0087 (0.0064, 0.0114)	17.4
1975–2008	132.9	14,948.5	0.0089 (0.0075, 0.0120)	17.8
<b>Females</b>				
1975–1984	8.6	635.7	0.0135 (0.0010, 0.0256)	25.7
1985–1994	19.4	2,884.6	0.0067 (0.0021, 0.0118)	13.7
1995–2008	70.6	11,570.7	0.0061 (0.0034, 0.0084)	12.6
1975–2008	98.5	15,090.9	0.0065 (0.0038, 0.0091)	13.4
<b>Males and Females Combined</b>				
1975–1984	15.0	1,290.3	0.0116 (0.0070, 0.0208)	23.6
1985–1994	47.5	5,903.7	0.0081 (0.0062, 0.0125)	16.2
1995–2008	168.8	22,845.4	0.0074 (0.0055, 0.0096)	15.0
1975–2008	231.4	30,039.4	0.0077 (0.0064, 0.0101)	15.6

<sup>a</sup> The confidence interval is estimated using the bootstrap method.

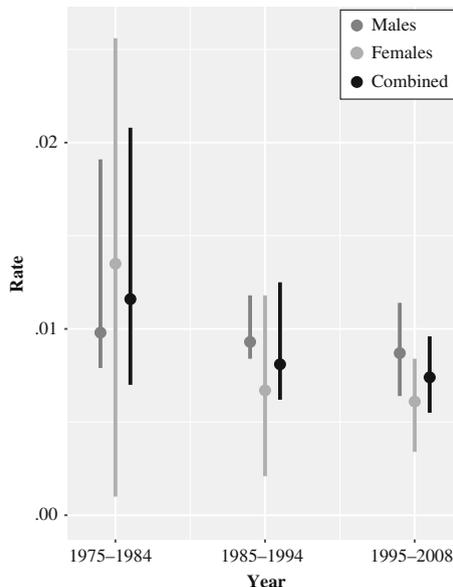
95 % confidence intervals, particularly for females in the period 1975–1984, indicating that the sample-induced variability in numbers of female emigrants (and exposure) is larger than that in numbers of male emigrants. The differences are not statistically significant. Hence, the observed differences in emigration rates between the periods considered can be produced by chance. The emigration rates and their 95 % confidence intervals are also shown in Fig. 1.

### *Emigration Rate by Age*

Table 2 shows emigration rates by age. At all ages, males have a higher emigration rate than females, but the difference is smallest in the age group 18–24. Here, females have their highest emigration rate. The highest emigration rate for males occurs for the age group 25–29. The numbers of emigrations by age are too small to produce age-specific emigration rates that are significantly different. The differences between emigration rates are not statistically significant at the 95 % level; nor are the differences significant at the 90 % level, except for the difference between age groups 25–29 and 30–39 for males and females combined. The rates are also shown in Fig. 2.

### *Emigration Probabilities*

The cumulative probability that an 18-year-old living in the Dakar region will leave Senegal at least once before age 40 can be computed from the emigration rates. If we assume that the average emigration rate of 0.0077 does not vary by age, then the probability that an 18-year-old living in the Dakar region will leave Senegal at least



**Fig. 1** Rates of emigration from Dakar region between ages 18 and 40 for females, males, and sexes combined, by period, and 95 % confidence bound

**Table 2** Emigration rates of Dakar region by age and sex for the period 1975–2008

Age	Number of Migrations (weighted)	Exposure (weighted)	Emigration Rate (and 95 % interval <sup>a</sup> )
<b>Males</b>			
18–24	70.8	8,157.9	0.0087 (0.0059, 0.0121)
25–29	37.3	3,520.0	0.0106 (0.0101, 0.0166)
30–39	24.8	3,270.7	0.0076 (0.0032, 0.0116)
18–39	132.9	14,948.5	0.0089 (0.0075, 0.0120)
<b>Females</b>			
18–24	67.6	8,184.8	0.0083 (0.0032, 0.0125)
25–29	19.8	3,603.1	0.0055 (0.0037, 0.0086)
30–39	11.1	3,303.0	0.0034 (0.0007, 0.0052)
18–39	98.5	15,090.9	0.0065 (0.0038, 0.0091)
<b>Males and Females Combined</b>			
18–24	138.4	16,342.7	0.0085 (0.0052, 0.0120)
25–29	57.1	7,123.0	0.0080 (0.0075, 0.0120)
30–39	36.0	76,573.7	0.0055 (0.0031, 0.0077)
18–39	231.4	30,039.4	0.0077 (0.0064, 0.0101)

<sup>a</sup> The confidence interval is estimated using the bootstrap method.

once before age 40 is 15.6 %.<sup>5</sup> The emigration probabilities by period and sex are shown in Table 1. The cumulative probabilities are computed assuming that the emigration rates vary by period and sex but not by age. If we consider three age intervals (i.e., 18–24, 25–29, and 30–39), then the probability that an 18-year-old emigrates before age 40 is 14.3 %.<sup>6</sup> Table 3 shows the cumulative probabilities and their 95 % and 90 % confidence intervals. The difference between males and females is statistically significant at the 90 % level but not at the 95 % level.

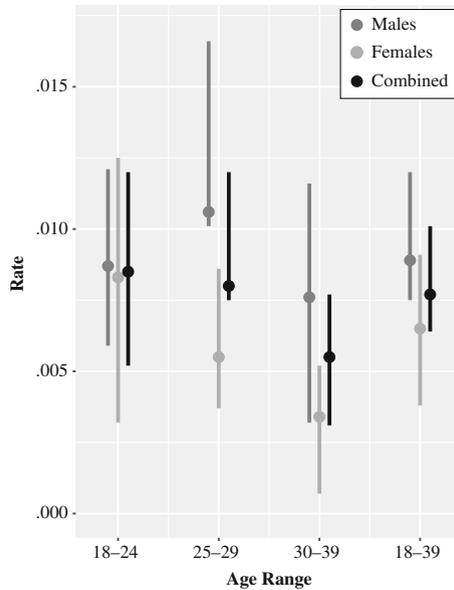
The probability that a 25-year-old male living in the Dakar region will emigrate within the next five years is 5.2 % if the individual experiences the average migration rate observed between 1975 and 2008 for the age group 25–29. It is 2.7 % for females.

## Discussion and Implications for Demography and Population Studies

Our study is the first that uses techniques of event history analysis and data on household members abroad as well as return migrants to estimate emigration rates defined as occurrence-exposure rates. Occurrence-exposure rates relate event counts to exposure times. Although the occurrence-exposure rate is central to demographic analysis and modeling, it is little used in international migration research because necessary data are lacking. Emigration surveys offer unique

<sup>5</sup>  $1 - \exp[-0.0077 \times (40 - 18)]$ .

<sup>6</sup>  $p = 1 - \exp[-(7 \times 0.0085 + 5 \times 0.0080 + 10 \times 0.0055)] = 0.1432$ .



**Fig. 2** Rates of emigration from Dakar region between 1975 and 2008 for females, males, and sexes combined, by period, and 95 % confidence bound

opportunities for estimating occurrence-exposure rates of emigration. To obtain representative emigration rates and appropriate variance estimates, the estimation must account for the sample design and nonresponse bias. Although sampling weights correct for the oversampling of particular groups (such as migrant households) and for nonresponse, their sole use is not sufficient to conduct viable variance estimation in complex multistage sampling. In this article, we combine methods for time-to-event data (concretely, survival analysis and event history analysis) for the estimation of the emigration rates and a replication method for variance estimation under a complex survey design in the presence of nonresponse. The central idea of the replication method is to draw replicate samples from the original sample by mimicking the sampling steps conducted to establish the survey sample. The variability of the weighted estimates among the replicate samples is then used as a replication-based estimator of variance. In this study, we use bootstrapping to produce replicate samples. The variability in the empirical distributions of the bootstrapped estimates determines the confidence intervals of emigration rates and the cumulative emigration probabilities. We apply the methods to data from the MAFE sample survey. We consider time trends in emigration rates

**Table 3** Probability that an 18-year-old resident of Dakar region will emigrate before age 40, with 95 % and 90 % confidence intervals (CI) given in parentheses

	Males	Females	Total
95 % CI	17.3 (14.6, 22.9)	11.2 (7.2, 15.3)	14.3 (12.4, 18.0)
90 % CI	17.3 (15.2, 22.6)	11.2 (7.9, 14.6)	14.3 (12.8, 17.4)

and age-specific emigration rates. The emigration rate models and the bootstrapping are implemented in R.<sup>7</sup>

The method that we propose does not include all sources of uncertainty. It excludes the uncertainty in the sampling frame, which in the MAFE study consisted of the 2008 population of the Dakar region, updated from the 2002 census. It also excludes the uncertainty and possible bias caused by emigration of whole households with nobody remaining in Senegal to report the characteristics of the household members.

The emigration rate of the Dakar region is less than 1 % (0.0077), with a 95 % confidence interval of (0.0068, 0.0088). The emigration rate differed significantly between males and females, and the rates varied over time. Emigration was highest in the late 1970s and the early 1980s. Emigration rates are overestimated considerably if sampling weights are not taken into account. Variance estimation that disregards the stratified two-stage sampling design of the MAFE survey and the nonresponse yields confidence intervals that are too narrow. This underlines the necessity of regarding the sampling design in order to avoid faulty research conclusions because of not having included all sources of uncertainty.

In the literature, no uniform definition of an emigrant and method for computing an emigration rate exist. The definition of an emigrant is often conditioned by data availability. As a consequence, published emigration rates are not comparable. Ideally, emigration rates relate numbers of emigrations in a given period by a group of people with given characteristics to the total duration of exposure during the same period by people with the same characteristics. The accurate measurement or approximation of exposure time is an essential component of the estimation of emigration rates. Migration surveys with information on date of migration or age at migration, collected either retrospectively or prospectively, provide unique opportunities for estimating emigration rates, provided that a sufficient number of emigrations are recorded. Emigration rates can be related to individual characteristics and contextual variables to identify the determinants of emigration (Baizán and González-Ferrer 2016; Dibeh et al. 2017) using established methods of event history analysis (see, e.g., Aalen et al. 2008; Beauchemin and Schoumaker 2016; Willekens 2014). The method of variance estimation, presented in this article, can be used to obtain standard errors if the sample size is large enough. Large sample sizes usually ensure symmetry in the distributions of the bootstrapped estimates required for deriving standard errors.

Our method provides a general approach to estimate occurrence-exposure rates and other demographic indicators, and their confidence intervals, from survey data with complex sample design. Complex survey designs are relatively common in demography and the social sciences. The Demographic and Health Surveys (DHS) have a complex design, too; the jackknife repeated replication procedure is routinely applied to estimate sampling errors for selected mortality and fertility rates (see, e.g., Adali and Türkyilmaz 2012). The Health and Retirement Survey (HRS) also uses a multistage stratified sampling strategy with oversampling of certain demographic groups (Sonnegga and Weir 2014). Cai et al. (2010) used the bootstrap method to estimate the sample design-adjusted variance of multistate life table indicators (e.g., health expectancy) from complex panel surveys that include stratification and multistage clustering.

<sup>7</sup> The R code and the data are available online from GitHub (<https://github.com/MLeuchter/MAFE/>).

Reliable emigration estimates contribute to evidence-based policies in countries of emigration and countries of immigration. Throughout history, governments have used emigration as instruments in population and labor market policies. More recently, governments have shown interest in the roles that diaspora can play in economic development and sociocultural and political change. Governments in immigration countries express an interest in the root causes of emigration, their variation between migrant categories, and their impact on emigration propensities. That insight is indispensable for the design of effective immigration policies with few unintended consequences (Czaika and de Haas 2013; Garip 2017). International migration is high on the political agenda. Scientists are challenged to produce usable knowledge by innovating the measurement, explanation, and prediction of emigration.

**Acknowledgments** We thank Cris Beauchemin and Anna Klabunde for their comments on an earlier draft. We are also grateful to two anonymous referees for their comments and suggestions. Cris Beauchemin coordinated the MAFE project; we acknowledge with thanks his permission to use the MAFE data and to upload the data to the GitHub repository. This article was written while the authors were affiliated with the Max Planck Institute for Demographic Research in Rostock, Germany.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Aalen, O. (1978). Nonparametric inference for a family of counting processes. *Annals of Statistics*, 6, 701–726.
- Aalen, O. O., Borgan, Ø., & Gjessing, H. K. (2008). *Survival and event history analysis: A process point of view*. New York, NY: Springer.
- Abel, G. (2013). Estimating global migration flow tables using place of birth data. *Demographic Research*, 28(article 18), 505–546. <https://doi.org/10.4054/DemRes.2013.28.18>
- Abel, G., & Sander, N. (2014). Quantifying global international migration flows. *Science*, 343, 1520–1522.
- Adali, T., & Türkyilmaz, A. S. (2012). Re-calculating the sampling variance of selected demographic indicators from the TDHS surveys: The effect of stratification. *Turkish Journal of Population Studies*, 34, 17–30.
- Allison, P. D. (1982). Discrete-time methods for the analysis of event histories. *Sociological Methodology*, 13, 61–98.
- Arslan, C., Dumont, J. C., Kone, Z. L., Özden, Ç., Parsons, C. R., & Xenogiani, T. (2016). *International migration to the OECD in the 21st century* (KNOMAD Working Paper No. 16). Washington, DC: World Bank. Retrieved from <https://www.knomad.org/publication/international-migration-oecd-21st-century>
- Azose, J. J., Ševčíková, H., & Raftery, A. E. (2016). Probabilistic population projections with migration uncertainty. *Proceedings of the National Academy of Sciences*, 113, 6460–6465.
- Baizán, P., & González-Ferrer, A. (2016). What drives Senegalese migration to Europe? The role of economic restructuring, labor demand, and the multiplier effect of networks. *Demographic Research*, 35(article 13), 339–380. <https://doi.org/10.4054/DemRes.2016.35.13>
- Beauchemin, C. (2015). Migration between Africa and Europe (MAFE): Advantages and limitations of a multi-site survey design. *Population-E*, 70, 13–37.
- Beauchemin, C., & Schoumaker, B. (2016). Micro methods: Longitudinal surveys and analysis. In M. J. White (Ed.), *International handbook of migration and population distribution* (pp. 175–204). Dordrecht, Netherlands: Springer.
- Bilsborrow, R. E. (2012, December). *Key issues of survey and sample design for surveys of international migration*. Paper presented at the French Institute for Demographic Studies (INED) Conference on Comparative and Multi-Sited Approaches to International Migration, Paris, France.

- Bilsborrow, R. E. (2016a). Concepts, definitions and data collection approaches. In M. J. White (Ed.), *International handbook of migration and population distribution* (pp. 109–156). Dordrecht, Netherlands: Springer.
- Bilsborrow, R. E. (2016b, December). *The global need for better data on international migration and the special potential of household surveys*. Paper presented at the Conference of the International Organisation of Migration's Global Migration Data Analysis Centre (GMDAC) on Improving Data on International Migration: Towards Agenda 2030 and the Global Compact on Migration, Berlin, Germany.
- Bilsborrow, R. E., Hugo, G., Oberai, A. S., & Zlotnik, H. (1997). *International migration statistics: Guidelines for improving data collection systems*. Geneva, Switzerland: International Labour Office.
- Blossfeld, H.-P., & Rohwer, G. (2002). *Techniques of event history modeling: New approaches to causal analysis*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Broström, G. (2012). *Event history analysis with R*. Oxford, UK: CRC Press.
- Cai, L., Hayward, M. D., Saito, Y., Lubitz, J., Haredom, A., & Crimmins, E. (2010). Estimation of multi-state life table functions and their variability from complex survey data using the SPACE program. *Demographic Research*, 22(article 6), 129–158. <https://doi.org/10.4054/DemRes.2010.22.6>
- Canada Border Services Agency. (2016). *Entry/exit initiative*. Ottawa: Canada Border Crossing Agency. Retrieved from <http://www.cbsa-asfc.gc.ca/btb-pdf/ebsiip-asfipi-eng.html>
- Canty, A. J., & Davison, A. C. (1999). Resampling-based variance estimation for labour force surveys. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 48, 379–391.
- Clark, X., Hatton, T. J., & Williamson, J. G. (2004). What explains emigration out of Latin America? *World Development*, 32, 1871–1890.
- Czaika, M., & de Haas, H. (2013). The effectiveness of immigration policies. *Population and Development Review*, 39, 487–508.
- D'Arrigo, J. (2011). *Understanding and dealing with unit nonresponse during and post survey data collection* (Doctoral dissertation). School of Social Sciences, University of Southampton, Southampton, UK.
- Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap methods and their application*. Cambridge, UK: Cambridge University Press.
- Dibeh, G., Fakhri, A., & Marrouch, W. (2017). *Decision to emigrate amongst the youth in Lebanon* (IZA Discussion Paper No. 10493). Bonn, Germany: Institute of Labor Economics.
- Direction de la Prévision et de la Statistique. (1995). *Enquête sur les migrations et l'urbanisation au Sénégal (EMUS) 1993–Rapport national* [Survey on Migration and Urbanization in Senegal (EMUS) 1993–National report] (Technical report of the Direction de la Prévision et de la Statistique, Ministère de l'Économie et des Finances). Dakar, Sénégal: Agence Nationale de la Statistique et de la Démographie (ANSD).
- El Colegio de la Frontera Norte. (2017). *EMIF Surveys of migration at Mexico's northern and southern borders*. Retrieved from <http://www.colef.mx/emif/eng/#>
- Esipova, N., Ray, J., & Pugliese, A. (2011). *Gallup world poll: The many faces of migration*. (IOM Migration Research Series No. 43). Geneva, Switzerland: International Organization for Migration (IOM), in cooperation with GALLUP. Retrieved from <http://publications.iom.int/bookstore/free/MRS43.pdf>
- Eurostat. (2017). *Households International Migration Surveys in the Mediterranean countries (MED-HIMS)*. Luxembourg City, Luxembourg: Eurostat. Retrieved from <http://ec.europa.eu/eurostat/web/european-neighbourhood-policy/enp-south/med-hims>
- Garip, F. (2017). *On the move: Changing mechanisms of Mexico-U.S. migration*. Princeton, NJ: Princeton University Press.
- Groenewold, G., & Bilsborrow, R. (2008). Design of samples for international migration surveys: Methodological considerations and lessons learned from a multi-country study in Africa and Europe. In C. Bonofazi, M. Okólkowski, J. Schoorl, & P. Simon (Eds.), *International migration in Europe: New trends and new methods of analysis* (pp. 292–312). Amsterdam, Netherlands: Amsterdam University Press.
- Hanson, G. H., & McIntosh, C. (2010). The great Mexican emigration. *Review of Economics and Statistics*, 92, 798–810.
- Hatton, T. J., & Williamson, J. G. (2011). Are third world emigration forces abating? *World Development*, 39, 20–32.
- Heeringa, S. G., West, B. T., & Berglund, P. A. (2010). *Applied survey data analysis*. Boca Raton, FL: CRC Press.
- Hoem, J. M., Keiding, N., Kulokari, H., Natvig, B., Barndorff-Nielsen, O., & Hilden, J. (1976). The statistical theory of demographic rates: A review of current developments [with discussion and reply]. *Scandinavian Journal of Statistics*, 3, 169–185.
- International Labour Organization, State Statistics Service, & Ptoukha Institute of Demography and Social Studies of the National Academy of Sciences of Ukraine. (2013). *Report on the methodology, organization and results of a modular sample survey on labour migration in Ukraine*. Geneva, Switzerland:

- International Labour Organization. Retrieved from [http://www.ilo.org/budapest/what-we-do/publications/WCMS\\_244693/lang-es/index.htm](http://www.ilo.org/budapest/what-we-do/publications/WCMS_244693/lang-es/index.htm)
- Jensen, E. B. (2013). *A review of methods for estimating emigration* (Working Paper No. 101). Washington, DC: U.S. Bureau of the Census.
- Kalton, G. (1979). Ultimate cluster sampling. *Journal of the Royal Statistical Society: Series A*, *142*, 210–222.
- Kalton, G. (2009). Methods for oversampling rare subpopulations in social surveys. *Survey Methodology*, *35*, 125–141.
- Kim, K., & Cohen, J. E. (2010). Determinants of international migration flows to and from industrialized countries: A panel data approach beyond gravity. *International Migration Review*, *44*, 899–932.
- Kish, L. (1995). *Survey sampling* (Wiley Classics Library ed.). New York, NY: John Wiley.
- Klabunde, A., Zinn, S., Willekens, F., & Leuchter, M. (Forthcoming). Multistate modeling extended by behavioral rules—An example of migration. *Population Studies*.
- Klein, J. P., & Moeschberger, M. L. (2003). *Survival analysis: Techniques for censored and truncated data* (2nd ed.). New York, NY: Springer.
- Leach, M. A., & Jensen, E. B. (2014, May). *Revisiting the residual method: Using the American Community Survey to estimate foreign-born emigration*. Paper presented at the annual meeting of the Population Association of America, Boston, MA.
- Lee, E. S., & Forthofer, R. N. (2006). *Analyzing complex survey data* (2nd ed.). London, UK: Sage.
- Lessault, D., & Flahaux, M.-L. (2013). Regards statistiques sur l'histoire de l'émigration internationale au Sénégal [Statistical views on the history of international migration from Senegal]. *Revue Européenne des Migrations Internationales*, *29*(4), 59–88.
- Marti, M., & Ródenas, C. (2007). Migration estimation based on the Labour Force Survey: An EU-15 perspective. *International Migration Review*, *41*, 101–126.
- Marti, M., & Ródenas, C. (2012). Measuring international migration through sample surveys: Some lessons from the Spanish case. *Population-E*, *67*, 435–464.
- Moultrie, T., Dorrington, R., Hill, A., Hill, K., Timæus, I., & Zaba, B. (2013). *Tools for demographic estimation*. Paris, France: International Union for the Scientific Study of Population. Retrieved from <http://demographicestimation.iussp.org>
- Office of National Statistics (ONS). (2017). *International Passenger Survey (IPS) methodology*. London, UK: ONS. Retrieved from <http://webarchive.nationalarchives.gov.uk/20160105160709/http://www.ons.gov.uk/ons/guide-method/method-quality/specific/travel-and-transport-methodology/international-passenger-survey-methodology/index.html>
- Orav, A., & D'Alfonso, A. (2016). *Smart borders: EU entry/exit system* (Briefing of the EU Legislation in Progress). Brussels, Belgium: European Parliament, European Parliamentary Research Service. Retrieved from [http://www.europarl.europa.eu/RegData/etudes/BRIE/2016/586614/EPRS\\_BRI\(2016\)586614\\_EN.pdf](http://www.europarl.europa.eu/RegData/etudes/BRIE/2016/586614/EPRS_BRI(2016)586614_EN.pdf)
- Organisation for Economic Cooperation and Development (OECD). (2000). *Emigration rates by country of origin, sex and educational attainment level* (DIOC-E Release No. 3). Paris, France: OECD. Retrieved from <http://www.oecd.org/migration/46561284.pdf>
- Organisation for Economic Cooperation and Development (OECD), & United Nations Department of Economic and Social Affairs (UNDESA). (2013). *World migration in figures: A joint contribution by UN-DESA and the OECD to the United Nations high-level dialogue on migration and development* (Report). New York, NY: UNDESA. Retrieved from <http://www.oecd.org/els/mig/World-Migration-in-Figures.pdf>
- Orrenius, P. M., & Zavodny, M. (2005). Self-selection among undocumented immigrants from Mexico. *Journal of Development Economics*, *78*, 215–240.
- Pisica, S. (2016, November). *Intra-EU data exchange: A method for improving migration statistics*. Paper presented at the Eurostat Conference on Social Statistics, Luxembourg City, Luxembourg. Retrieved from <http://socialstats2016.eu/node/173>
- Preston, S. H., Heuveline, P., & Guillot, M. (2001). *Demography: Measuring and modeling population processes*. Oxford, UK: Blackwell.
- Razafindratsima, N., Legleye, S., & Beauchemin, C. (2011). Biais de non-réponse dans l'enquête migrations entre l'Afrique et l'Europe (MAFE-Sénégal) [Non-response bias in the survey on migration between Africa and Europe (MAFE- Sénégal)]. In M.-E. Tremblay, P. Lavallée, & M. E. H. Tirari (Eds.), *Pratiques et méthodes de sondage* [Practices and methods of surveys] (pp. 95–99). Paris, France: Dunod.
- Rendall, M. S. (2003). Estimation of annual international migration from the Labour Force Surveys of the United Kingdom and the continental European Union. *Statistical Journal of the United Nations Economic Commission for Europe*, *20*, 219–234.

- Rendall, M. S., Aguila, E., Basurto-Dávila, R., & Hancock, M. S. (2009, April). *Migration between Mexico and the U.S. estimated from a border survey*. Paper presented at the annual meeting of the Population Association of America, Detroit, MI.
- Rogers, A. (1990). Requiem for the net migrant. *Geographical Analysis*, 22, 283–300.
- Rust, K. F., & Rao, J. N. (1996). Variance estimation for complex surveys using replication techniques. *Statistical Methods in Medical Research*, 5, 283–310.
- Schoorl, J. J., Heering, L., Esveldt, I., Groenewold, G., van der Erf, R. F., Bosch, A. M., . . . de Bruijn, B. J. (2000). *Push and pull factors of international migration: A comparative report* (Report 3/2000/E/no.14 of the European Commission/Eurostat). Luxembourg City, Luxembourg: Office for Official Publications of the European Communities. Retrieved from <https://www.nidi.nl/shared/content/output/2000/eurostat-2000-theme1-pushpull.pdf>
- Schoumaker, B., & Beauchemin, C. (2015). Reconstructing trends in international migration with three questions in household surveys: Lessons from the MAFE project. *Demographic Research*, 32(article 35), 983–1030. <https://doi.org/10.4054/DemRes.2015.32.35>
- Schoumaker, B., & Mezger, C. (2013). *Sampling and computation weights in the MAFE surveys* (MAFE Methodological Note 6). Paris, France: Institut National d'Etudes Démographiques. Retrieved from [http://mafeproject.site.ined.fr/en/methodo/methodological\\_notes/](http://mafeproject.site.ined.fr/en/methodo/methodological_notes/)
- Sonnega, A., & Weir, D. R. (2014). The Health and Retirement Study: A public data resource for research on aging. *Open Health Data*, 2(1), e7. <https://doi.org/10.5334/ohd.am>
- United Nations. (1998). *Recommendations on statistics of international migration: Revision 1* (Statistical Papers, No. 58, Sales No. E98.XVII.14). New York, NY: United Nations.
- United Nations. (2016). *International migration report 2015* (ST/ESA/SER.A/384). New York, NY: United Nations Department of Economic and Social Affairs, Population Division. Retrieved from <http://www.un.org/en/development/desa/population/migration/publications/migrationreport/docs/MigrationReport2015.pdf>
- United Nations World Tourism Organization (UNWTO). (2016). *UNWTO Annual Report 2015*. Madrid, Spain: UNWTO. Retrieved from [http://cf.cdn.unwto.org/sites/all/files/pdf/annual\\_report\\_2015\\_lr.pdf](http://cf.cdn.unwto.org/sites/all/files/pdf/annual_report_2015_lr.pdf)
- U.S. Department of Homeland Security. (2016). *Comprehensive biometric entry/exit plan: Fiscal year 2016 report to Congress*. Washington, DC: Department of Homeland Security. Retrieved from <https://www.dhs.gov/sites/default/files/publications/Customs%20and%20Border%20Protection%20-%20Comprehensive%20Biometric%20Entry%20and%20Exit%20Plan.pdf>
- Valliant, R., Dever, J. A., & Kreuter, F. (2013). *Practical tools for designing and weighting survey samples*. New York, NY: Springer.
- Van Dalen, H. P., & Henkens, K. (2007). Longing for the good life: Understanding emigration from high-income countries. *Population and Development Review*, 33, 37–66.
- Van Dalen, H. P., & Henkens, K. (2013). Explaining emigration intentions and behaviour in the Netherlands, 2005–10. *Population Studies*, 67, 225–241.
- Van Hook, J., & Zhang, W. (2011). Who stays? Who goes? Selective emigration among the foreign-born. *Population Research and Policy Review*, 30, 1–24.
- Van Hook, J., Zhang, W., Bean, F. D., & Passel, J. S. (2006). Foreign-born emigration: A new approach and estimates based on matched CPS files. *Demography*, 43, 361–382.
- Wagner, H., & Eckmair, D. (2006). Simulation studies for complex sampling designs. *Simulation*, 35, 419–435.
- Westmore, B. (2015). International migration: The relationship with economic and policy factors in the home and destination country. *OECD Journal: Economic Studies*, 2015, 101–122. [https://doi.org/10.1787/eco\\_studies-2015-5jrp104jpz7j](https://doi.org/10.1787/eco_studies-2015-5jrp104jpz7j)
- Willekens, F. (2014). *Multistate analysis of life histories with R*. Heidelberg, Germany: Springer.
- Willekens, F. (2016). Migration flows: Measurement, analysis and modeling. In M. J. White (Ed.), *International handbook of migration and population distribution* (pp. 225–241). Dordrecht, Netherlands: Springer.
- Willekens, F. (2017). The decision to emigrate: A simulation model based on the theory of planned behaviour. In A. Grow & J. van Bavel (Eds.), *Agent-based modelling in population studies: Concepts, methods and applications* (pp. 257–299). Dordrecht, Netherlands: Springer.
- Wiśniowski, A. (2017). Combining Labour Force Survey data to estimate migration flows: The case of migration from Poland to the UK. *Journal of the Royal Statistical Society: Series A*, 180, 185–202.
- Wolter, K. M. (2007). *Introduction to variance estimation* (2nd ed.). New York, NY: Springer.
- Woodrow-Laffield, K. A. (2010, April). *Reconsidering multiplicity surveys on emigration from the United States*. Paper presented at the annual meeting of the Population Association of America, Dallas, TX.
- World Bank. (2016). *Migration and remittances: Factbook 2016* (3rd ed.). Washington DC: World Bank. Retrieved from <http://siteresources.worldbank.org/INTPROSPECTS/Resources/334934-1199807908806/4549025-1450455807487/Factbookpart1.pdf>