DEMOGRAPHIC RESEARCH
A peer-reviewed, open-access journal of population sciences

DEMOGRAPHIC RESEARCH

VOLUME 32, ARTICLE 35, PAGES 983–1030
PUBLISHED 19 MAY 2015
http://www.demographic-research.org/Volumes/Vol32/35/
DOI: 10.4054/DemRes.2015.32.35

Research Article

Reconstructing trends in international migration with three questions in household surveys: Lessons from the MAFE project

Bruno Schoumaker

Cris Beauchemin

Table of Contents

Demographic Research: Volume 32, Article 35
Research Article

Reconstructing trends in international migration with three questions in household surveys: Lessons from the MAFE project

Bruno Schoumaker[1]

Cris Beauchemin[2]

Abstract

BACKGROUND
Data on migration trends are crucially lacking in developing countries. The lack of basic information on migration contrasts sharply with the increasing importance of migration in the policy agenda of both sending and receiving countries.

OBJECTIVE
The general objectives of this paper are: to show how trends in international migration can be reconstructed with three questions in a household survey; to evaluate the precision of the estimates; and to test how sensitive the estimates are to several methodological choices and assumptions.

METHODS
Migration trends are reconstructed with event history models. The reconstruction uses data collected through migration surveys conducted in cities in three countries (Senegal, Democratic Republic of Congo, and Ghana) as part of the MAFE (Migration between Africa and Europe) project. Specifically, two types of data are used: simple data on the first migration of children of household heads, collected through household surveys, and full migration histories of children collected in biographic surveys. First, we evaluate the precision of our estimates using data collected in the household questionnaire. Next, the sensitivity of our results to different methodological choices and assumptions is evaluated.

RESULTS
Estimates using household data may be affected by large confidence intervals, and migrations trends are influenced by the simplifying assumptions that are made when

[1] Université catholique de Louvain, Belgium. E-Mail: bruno.schoumaker@uclouvain.be.
[2] INED, France. E-Mail: cris.beauchemin@ined.fr.

using these data. Despite these limitations, estimates based on three simple questions provide useful information on migration levels and trends.

1. Introduction

There is a crucial lack of data for studying trends in international migration flows. This is especially true in developing countries, but to some extent also in developed nations. Census data allow estimating bilateral stocks of migrants for many countries (Parsons et al. 2007), but they give no direct information on migration flows[3]. Administrative statistics on immigration flows are mainly limited to developed countries, and they suffer from various imperfections (Poulain, Perrin, and Singleton 2006)[4]. Statistics on outmigration flows are even less common, and they are also seriously deficient (OECD 2008). As a consequence, reconstructing departure and return trends in most countries is not possible with the existing data. The lack of basic information on migration flows is in sharp contrast to the increasing importance of migration in the policy agenda of both sending and receiving countries (Jensen 2013).

Demographic surveys offer useful opportunities to collect original data on international migration (Jensen 2013; Kasnauskienė and Igoševa 2010; Bilsborrow 2007; Zaba 1987). The reconstruction of internal migration trends with survey data is relatively common (Beauchemin 2011; Piché, Gregory, and Lavoie 1984). However, measuring trends in international migration with survey data is less frequent. The Mexican Migration Project (MMP) was a pioneer in this regard (Massey 1987; Donato 1998), focusing on flows between Mexico and the USA. The MAFE (Migration between African and Europe) project[5] also collected data to estimate trends in

[3] See Abel (2013) for an interesting way of estimating migration flows from tables on stocks.

[4] Only legal migrations are recorded in migration statistics. Moreover, data published in some countries refer only to permanent migration (e.g., in the US), or they exclude asylum seekers from migration statistics (e.g., in Belgium). Definitions of migration also vary across countries (e.g., 3 months in Belgium, 12 months in France).

[5] The MAFE project is coordinated by INED (C. Beauchemin) and is additionally formed by the Université Catholique de Louvain (B. Schoumaker), Maastricht University (V. Mazzucato), the Université Cheikh Anta Diop (P. Sakho), the Université de Kinshasa (J. Mangalu), the University of Ghana (P. Quartey), the Universitat Pompeu Fabra (P. Baizan), the Consejo Superior de Investigaciones Científicas (A. González-Ferrer), the Forum Internazionale ed Europeo di Ricerche sull'Immigrazione (E. Castagnone), and the University of Sussex (R. Black). The MAFE project received funding from the European Community's Seventh Framework Programme under grant agreement 217206. The MAFE-Senegal survey was conducted with the financial support of INED, the Agence Nationale de la Recherche (France), the Région Ile de France and the FSP programme "International Migrations, territorial reorganizations and development of the countries of the South". For more details, see http://mafeproject.site.ined.fr/en/. The MAFE datasets are

international migration from three African countries (Democratic Republic of Congo, Ghana, and Senegal, see Figure 1), as well as trends in return migration. The general approach followed in these projects is to collect a few simple questions on migration in household surveys, and reconstruct trends in migration with event history models. The method consists in reconstructing trends in migration with retrospective information on the first migration (date and destination) collected on all surviving children of household heads. Such data can potentially improve the knowledge of levels, trends, and patterns of international migration. The availability of socio-demographic data (gender, education, etc.) may also allow richer descriptions of migration than with other data sources. By including such questions in existing surveys (Labour Force Surveys, Living Standard Surveys, Demographic and Health Surveys), data on international migration could be collected at a relatively low cost[6].

The general objectives of this paper are to show how migration trends can be measured with only three questions in a household survey and to then evaluate the precision of the estimates and test how sensitive those estimates are to several methodological choices and assumptions. Three general issues are assessed: (1) whether including information on deceased children matters or not; (2) whether collecting data from household surveys or from the network module in biographic surveys influences the results; and (3) to what extent collecting data on only the first migration affects the estimates through 'filtering effects' (i.e., effects due to the fact that some migrations are neglected). The MAFE household and biographic data allow us to compare the results for three different contexts in which the patterns of migration (and thus the sensitivity to methodological choices) are potentially different.

After this introduction, the paper is divided into four parts. In the next part (part 2), we review some discussions about the collection of migration data in surveys carried out in origin countries, and we present the MAFE data. Part 3 of the paper presents the methodology used to compute trends in migration (first departure) and exposes the baseline results. Next, these results are contrasted with other approaches in order to assess the proposed methodology's sensitivity to various assumptions. The conclusion discusses suggestions for future surveys on migration and for further research.

available to the international scientific community through the portal of the Reseau Quetelet (www.reseau-quetelet.cnrs.fr).

[6] As discussed later, the sample size may be a limiting factor in some cases.

Figure 1: Map of Africa and Europe showing the three African MAFE countries and the six European MAFE countries (prepared by the authors with R's mapdata package)

2. Collecting data on migration in surveys at origin

2.1 Advantages and drawbacks of household surveys

The general approach to collecting emigration data in household surveys consists in obtaining information on migrants and on a comparison group. Data is collected on people who have migrated (whether they live abroad or have returned) and people who have never migrated. A great advantage of collecting data through a household survey in origin countries is the possibility of obtaining data on migration to all destinations at a relatively low cost[7]. Moreover, given that household surveys are conducted regularly in most countries, adding a few questions on migration may be quite cost-effective (Jensen 2013).

The collection of migration data in household surveys also has well-known limitations (Bilsborrow 2007; Beauchemin 2015). By definition, data on people living abroad are collected from proxy respondents. The information collected on emigrants in this way cannot be as detailed as the information that would be provided directly by the emigrants themselves, and this information is also thought to be less reliable (Bilsborrow 2007). The questions on the migration experience of migrants living abroad are usually few and simple. Another drawback of household survey data is that, if entire households emigrate, information on the migration of its members may not be collected (depending on who the information is collected from). As summarized by Bilsborrow (2007:4), "the more people emigrate from a country as entire households, the more a survey in the place of origin will fail to cover emigrants from that country, and the less useful the survey conducted only at the place of origin becomes". Working with survey data also means that the measurements are affected by sampling errors. These will depend on the sample size, on the sample design, and on the prevalence of migration. Given that international migration is a relatively rare event, oversampling households/areas with migrants is usually recommended (Bilsborrow 2007).

Another important question is: who are the migrants on whom information is to be collected? Even though surveys on international migration commonly ask questions in the origin households about migrants, there is no standardized methodology to identify migrants. Each survey adopts its own approach to defining the migrants to be included in its household questionnaire. Some define them on the basis of social obligations and expectations, as was the case for the Push-Pull project that recorded "those who are presently residing elsewhere but whose principal commitments and obligations are to that household and who are expected to return to that household in the future or whose

[7] Compared to surveys collecting information in origin and destination countries.

family will join them in the future" (Ünalan 2005: 220)[8]. Others use residential criteria, such as the NESMUWA surveys (Network of Surveys on Migration and Urbanization in West Africa), which recorded individuals who had previously lived in the household for at least 3 months and who had been living abroad for at least 6 months at the time of the survey (Bocquier 2003). In their guidelines for surveys on international migration, Bilsborrow and his co-authors recommend defining an international migrant as follows: "a person who used to live in the country in which the interview is being conducted and was a member of the household of the person being interviewed but who left at some point during the five years preceding the interview to live abroad for at least six months" (Bilsborrow et al. 1997: 247). Some censuses that include questions on international migrants also focus on the former members of the household that moved to live abroad within the last five years preceding the census. Another, completely different approach consists in referring to family relationships, such as in the MMP (Massey 1987), which records all children of the household head, whatever their place of residence (in Mexico or abroad)[9].

Obviously, the definition of the migrant population has an impact on the analysis potential. When the intention is to reconstruct migratory trends for a period of two or three decades, it is not sufficient to collect data on only the more recent migrants (e.g., those who left within the 5 years preceding the survey). Furthermore, the point is not only to register long-term international migrants but to also collect information on individuals of the relevant comparison group (the comparable persons who could have moved but did not). When analyzing migration trends over long periods, using household members (at the time of the survey) as the reference group is thus not a valid option. Groups of people defined by permanent links (children or siblings of the respondent) are preferable. In this way, a sample of all the people who previously lived in the origin country can be constituted, and information is collected regardless of their status at the time of the survey (living abroad or not, living in the household or not, alive or not). This is in some ways similar to the data collected on mortality from birth histories or sibling survival histories.

[8] Note that this reference to the future household is conceptually problematic. Indeed, the concept of household refers to the group of people who live together in a residence, under the authority of the head, at the time of the survey. At another time, the group may be different (with members disappearing and new members arriving) or the head might change, as well as the place of residence. The reference to the future is thus not at all clear when talking about a household: Does it refer to the group, the place, or the head?

[9] See also Woodrow-Lafield (1996) for estimates of US emigration based on a multiplicity sampling method, using data collected on parents, siblings, and children. Indirect methods developed in the 1970s and 1980s also collect migration data on children or siblings (Zaba 1987), but only on the number of migrants among children or siblings. No data are collected on each child separately; as a result, no information is available on the dates and destination of migration. These methods have been used to estimate stocks of migrants rather than migration trends (Zaba 1987).

2.2 The MAFE data

The data used in this paper come from the MAFE project. The MAFE project is a multi-site project on international migration. Its objectives and questionnaires were inspired by the Mexican Migration Project (Beauchemin 2015). The objectives of the MAFE project are to measure trends and patterns of migration, causes of departures and returns, and consequences of international migration on economic and family outcomes. Both household and individual data were collected in 2008 and 2009 in cities of three sub-Saharan countries (Accra and Kumasi in Ghana, Dakar in Senegal, and Kinshasa in DR Congo), and individual data were collected in six destination countries (Belgium, France, Italy, Netherlands, Spain, UK). The same questionnaires were used in all the settings, making data comparable across countries.

2.2.1 Household surveys

Household surveys were conducted in sending countries (in 2008/2009) among representative samples of households of selected cities (1,187 in Accra/Kumasi; 1,141 in Dakar; 1,576 in Kinshasa). The samples were limited to these cities because of budget and time constraints. Dakar represents around a quarter of the total population of Senegal, Kinshasa around 12% of the population of DR Congo, and Accra and Kumasi about 17% of Ghana's population (Beauchemin 2015). Outmigration is known to be higher from these cities than from the rest of the country (Beauchemin 2015). As a result, our estimates are not representative of the countries[10].

Two-stage stratified random samples of households were selected in each city. Stratification was used in order to increase the number of migrants in the sample (Schoumaker and Mezger 2013). First, primary sampling units with a high level of out-migration were over-sampled (except in Ghana). Second, within the selected primary sampling units, households with migrants were oversampled. In the listing phase, households with return migrants and households in which at least one adult was living abroad were identified; these households were oversampled by a factor of about 5 in Senegal and Ghana and 4 in DR Congo[11].

[10] We later used the name of the countries, although the results apply to migration trends from cities.
[11] Sampling weights are used in all the analyses to take oversampling into account.

Figure 2: Questions on the first departure and the first return from the MAFE household questionnaire (Ghana)

<table>
<tr><th colspan="3">1st out-migration</th><th colspan="2">1st return</th></tr>
<tr>
<td>A12.
Has "Name" already lived outside of Ghana since he/she was born / settled the 1st time in Ghana?
10. No, never
➔ Next person
11. Yes, for less than a year
➔ Next person
03. Yes, for one year or longer</td>
<td>A13a.
In which year did you/he/she leave Ghana for the first time for at least one year?
Don't Know:
How old were you / was he/she?</td>
<td>A13b.
What was the destination country when you/he/she left Ghana for the first time?
(i.e. the first country where he/she stayed for at least a year)</td>
<td>A13c.
Did you/he/she return to Ghana for at least a year since you/ he/she first left?
1. Yes
➔ mark RETURN in the flap
2. No
➔ Next person</td>
<td>A13d.
In which year did you/he/she return for the first time?
(Indicate the date of the first return that lasted one year or longer)
Don't Know:
How old were you / was he/she?
➔ Next person</td>
</tr>
<tr><td>|__|__|</td><td>Year |__|__|__|__|
|__|__|years old</td><td></td><td>❑ Yes ❑ No</td><td>Year |__|__|__|__|
|__|__|years old</td></tr>
<tr><td>|__|__|</td><td>Year |__|__|__|__|
|__|__|years old</td><td></td><td>❑ Yes ❑ No</td><td>Year |__|__|__|__|
|__|__|years old</td></tr>
<tr><td>|__|__|</td><td>Year |__|__|__|__|
|__|__|years old</td><td></td><td>❑ Yes ❑ No</td><td>Year |__|__|__|__|
|__|__|years old</td></tr>
<tr><td>|__|__|</td><td>Year |__|__|__|__|
|__|__|years old</td><td></td><td>❑ Yes ❑ No</td><td>Year |__|__|__|__|
|__|__|years old</td></tr>
<tr><td>|__|__|</td><td>Year |__|__|__|__|
|__|__|years old</td><td></td><td>❑ Yes ❑ No</td><td>Year |__|__|__|__|
|__|__|years old</td></tr>
</table>

Data were collected on all the usual members of the household[12], as well as on a series of people related to the household but who were not household members at the time of the survey. These additional people were the following:

1. All the children of the household head living out of the household, whatever their place of residence (including those who are deceased). They may be international migrants or not. This category thus includes domestic migrants.

[12] See Table 1 for a definition of households in the MAFE surveys.

Since data was also collected on the children living in the household, information is available on all children of the household head;

2. The partners, mothers, or fathers of any household member who are living abroad[13];
3. All other persons living abroad who are relatives of the household head or his/her partner, and who have been in regular contact with the household over the past 12 months.

The household questionnaire includes an introductory module containing socio-demographic variables on each individual (usual household members and additional people) and a detailed module on migration experience[14]. In order to compute trends in migration and return, five simple questions were asked for all individuals (Figure 2):

- a screening question (A12) indicating whether or not each individual had lived for at least one year out of his/her origin country (whatever the time of departure) [15];
- two questions on the first departure to another country (A13a and A13b) related to the year and the destination country ;
- two questions on the first return (A13c and A13d), one indicating whether a return had occurred, and (if yes) the year of the first return.

In this paper we focus on the computation of trends of departure; therefore only the first three questions (A12, A13a and A13b) are used.

[13] Note that only the living-abroad partners of household members were systematically identified in MAFE-Senegal. Mothers and fathers of household members were registered only if they contributed to the domestic economy (thus entering into the third category). By contrast, living-abroad parents of household members in MAFE-Congo and MAFE-Ghana were systematically recorded.

[14] The questionnaire is available online at mafeproject.site.ined.fr/en/.

[15] Questions on migration were restricted to stays (abroad and/or at origin in case of return) of at least 12 months. This minimum length is widely used in migration studies, although a variety of thresholds are used. By focusing on stays of at least 12 months, short-term moves (such as seasonal migrations) are ignored. Estimates of international migration would thus have been higher had short-term migration been included. The impact of the length of stay on the number of migrations is expected to be greater for migration to African countries than to Western countries, as migrations of short duration are expected to be more frequent to neighbouring countries.

2.2.2 Biographic surveys

The MAFE biographic surveys are also used in this paper as a means for evaluating some of the assumptions of the household data and for testing alternative data collection methods. Biographic data were collected among individuals aged 25 and over, both in origin and destination countries: non-migrants and return migrants were interviewed in Africa (around 1,500 individuals per selected country), and migrants (at the time of the survey) were interviewed in six European destination countries (between 150 and 280 migrants per destination country: Belgium, Italy, France, Spain, the Netherlands, UK, see Figure 1). In Africa, return migrants and partners of migrants were oversampled, and they were between 2 and 4 times more likely to be selected than non-migrants (weights are used to correct for this oversampling). Life histories were collected, including full migration histories of the interviewees. Interestingly, the questionnaire includes another module that can be used to reconstruct migration trends (even if it was not designed for that purpose): the so-called 'network module'. Each respondent in the biographic survey (regardless of his/her migration status) was asked to reconstruct migration histories of a series of relatives who had lived at least one year out of the country of origin. Figure 3 illustrates the way data was collected in the network module. For each individual (children, brothers, sisters, father, mother, and other relatives or close friends the interviewee could have counted on to migrate), the list of all the changes of country (for at least one year) and the dates of the changes were collected in a grid. For children, siblings, and parents, the information covers the period from the first departure until the time of the survey: for the others the information is collected from the time of encounter with the interviewee. By selecting the appropriate information registered in the network module it is possible to mimic the data that would have been obtained in the household survey and to thus test some of the assumptions made when computing migratory trends with only the three simple questions. More generally, the network module is useful for evaluating the consistency of migration trends with two different tools, and it also allows documenting joint migration of parents and children. In this paper we mainly use the data collected in the network module in the origin countries. The data were collected from among samples of 1,062 individuals in Senegal, 1,243 in Ghana, and 1,638 in the DR Congo (Schoumaker and Mezger 2013).

Figure 3: Illustration of the network module in MAFE biographic questionnaire (Ghana)

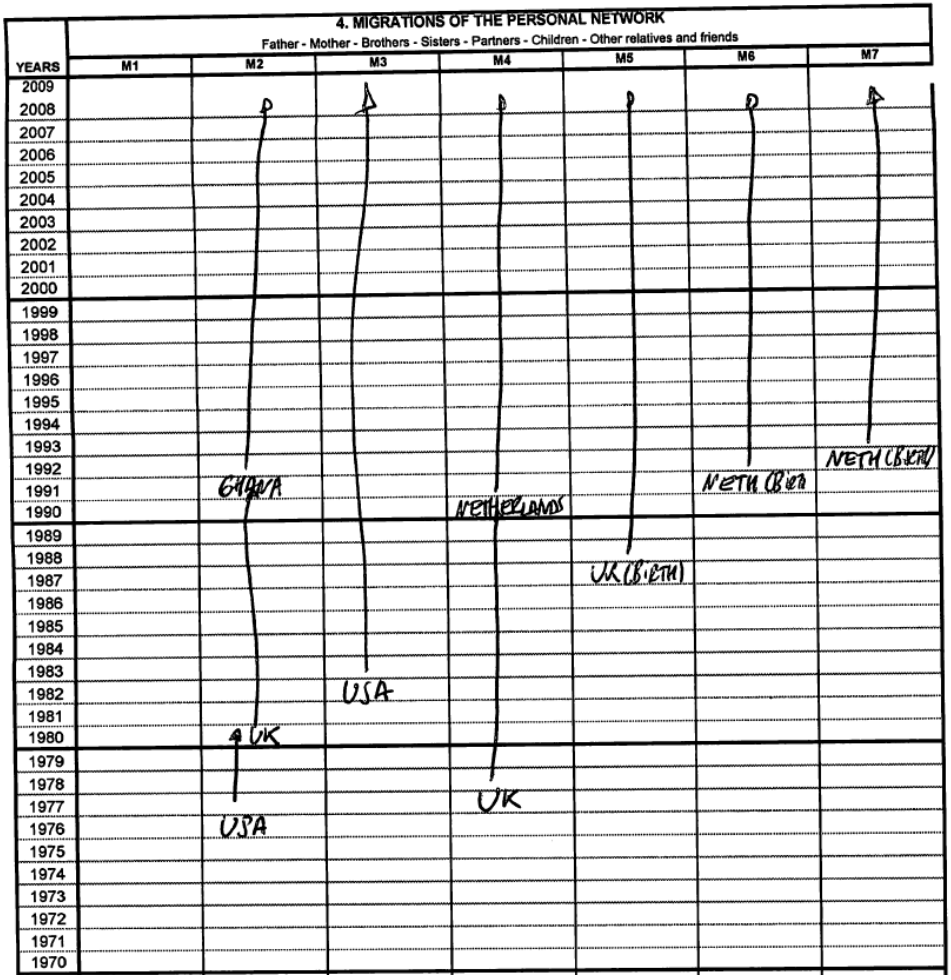
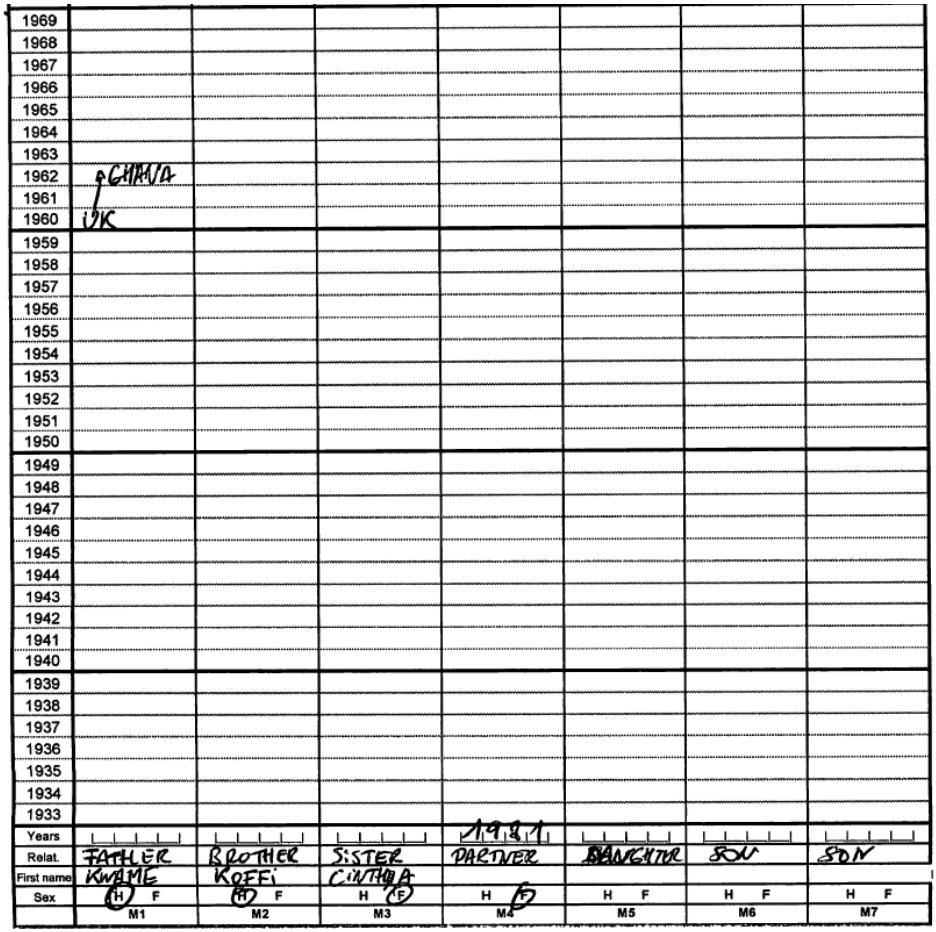


Figure 3: (Continued)



Note: Each column represents the migration history of the relatives who lived at least 1 year out of the interviewee's origin country. Histories start at the time of the first stay abroad for parents, siblings, and children, and at the time of first encounter with the interviewee for the others (partners and other relatives or close friends the interviewee could have counted on to migrate). Histories end at the time of the survey.

3. Reconstructing trends in departure with three questions

In this part, we present the methodology used to reconstruct migration trends (departure) for the three flows under study in the MAFE project (DR Congo, Ghana, and Senegal). Baseline results (i.e., results obtained with the three simple questions included in the household questionnaire: A12, A13a, and A13b) are presented, and their confidence intervals are computed. The sensitivity of the estimates to various assumptions is assessed in part 4.

3.1 Methodology

3.1.1 Choosing the appropriate population

As discussed before, an important concern when generating migration trends is the use of an appropriate population at risk of migration. Adopting a retrospective approach that looks back at several decades precludes considering household members as the reference population, because the household composition changes over time. Fixed relationships are more suited to ensuring that the same categories of individual are included in the numerator (the people who migrated) and the denominator (the people who were at risk of migrating). Table 1 summarizes the list of persons included in the MAFE household questionnaire and indicates their potential migratory status according to their place of residence at the time of the survey. This table helps to identify which individuals should be selected to compute migratory trends.

Theoretically, we need to include in the analyses all people who have been at risk of moving abroad, regardless of their current place of residence. The household heads are not eligible since, by definition, they cannot be living abroad (migrants) at the time of the survey. Actually, household heads' children are the only eligible category of people in our data, since they were registered regardless of their place of residence at the time of the survey. Information on deceased children was also collected. All other groups of people included in the survey are not eligible because we do not have the entire population at risk of migration. For instance, information on migration is available for the father of the head of the household only if the father lives in the same household or if the father lives abroad, but not if the father lives in another household within the country. As a result, migratory trends are computed in this paper by using only the information on the heads' children.

Table 1: Information collected in the MAFE household surveys on international migration by relationship to household members, migration status, and place of residence (at the time of the survey)

Relationship to the head or to another member of the surveyed household	The individuals are considered **members of the surveyed household**		The individuals are **NOT** considered **members of the surveyed household** (but information is or is not collected on them)**			
	Non migrant	Return migrant	Non migrant	Return migrant	Migrant (abroad)	Deceased
Household head	Yes	Yes	No	No	No	No
Children of the HH head	Yes	Yes	Yes	Yes	Yes	Yes
Other HH member	Yes	Yes	No	No	No	No
Spouse of a HH member	Yes	Yes	No	No	Yes	No
Father/mother of a HH member*	Yes	Yes	No	No	Yes	No
Other migrant declared by the respondent	No	No	No	No	Yes	No

Definition: In MAFE, the household is classically defined as a group of individuals who live together and partly or totally share their resources to satisfy their essential needs (housing, eating). To be considered members of a household, individuals must have been living there for at least 6 months or must intend to live there for at least 6 months.

* This information was systematically recorded only in MAFE-Congo and MAFE-Ghana. In MAFE-Senegal, parents are recorded only if they have been in regular contact with the household over the past 12 months.

** See section 2.1 for a precise list of the categories of individuals who are not considered household members but for whom, information was anyway collected in the household survey.

Working on the heads' children has a drawback: the number of older adults in the sample is limited, since they can only be reported by their parents. It thus imposes an upper age for the analyses and it limits the possibilities of going back in the past to describe migratory trends. This is illustrated in Figure 4, which presents the number of person-years (i.e., the number of years lived by all children from their birth until the time of the survey or until death) and the number of events (first departure) by age groups and by 5-year periods in the DR Congo MAFE household survey. Figure 4 (a) shows that the number of person-years of people at risk of experiencing migration quickly decreases as one goes back in time, and it is very low before the 1980s. The number of person years above age 40 is also very low before the mid-1990s. Figure 4 (b) shows the number of events by 5-year period and by age group. The number of migrations before the 1980s is also very low, and events are very few above age 40. These data come from the MAFE-Congo survey, but the same issue applies to the three countries. Analyses are thus limited to age 40 and to periods starting in 1975.

Figure 4: Number of person years and events among children of household heads, by 5-year period and by age group – Data: MAFE Congo Household survey

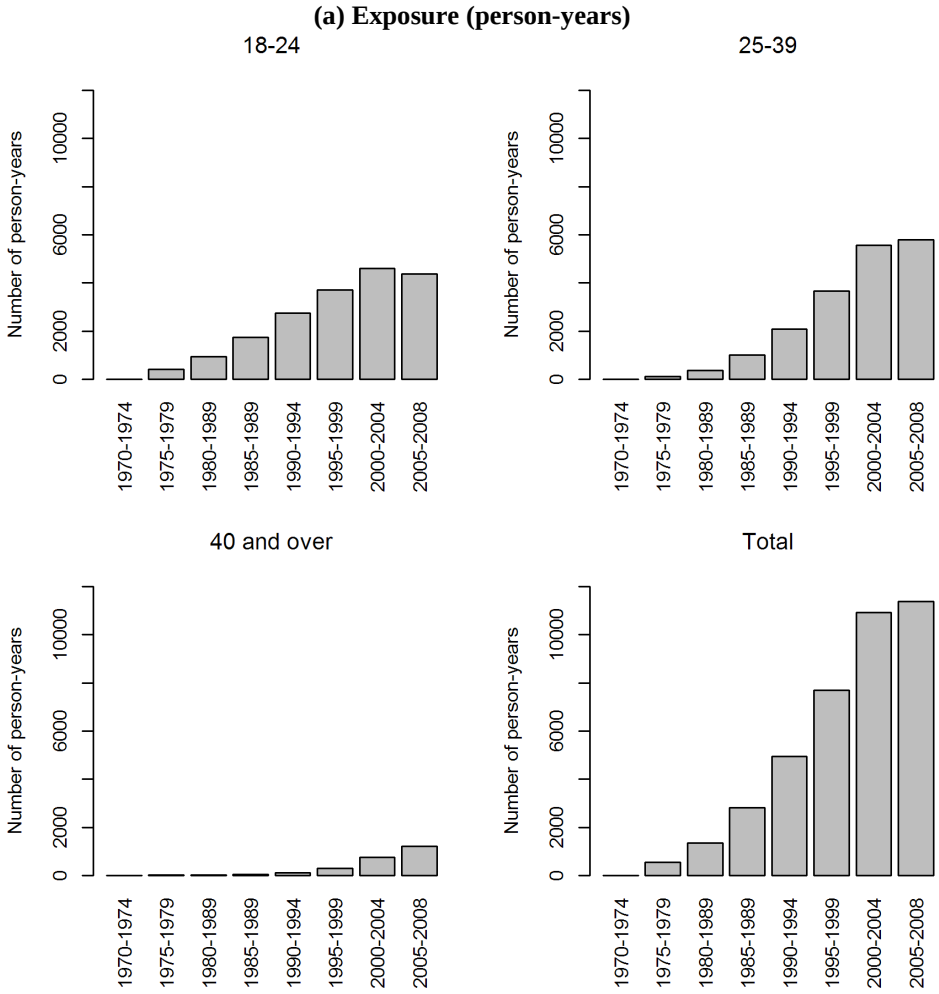
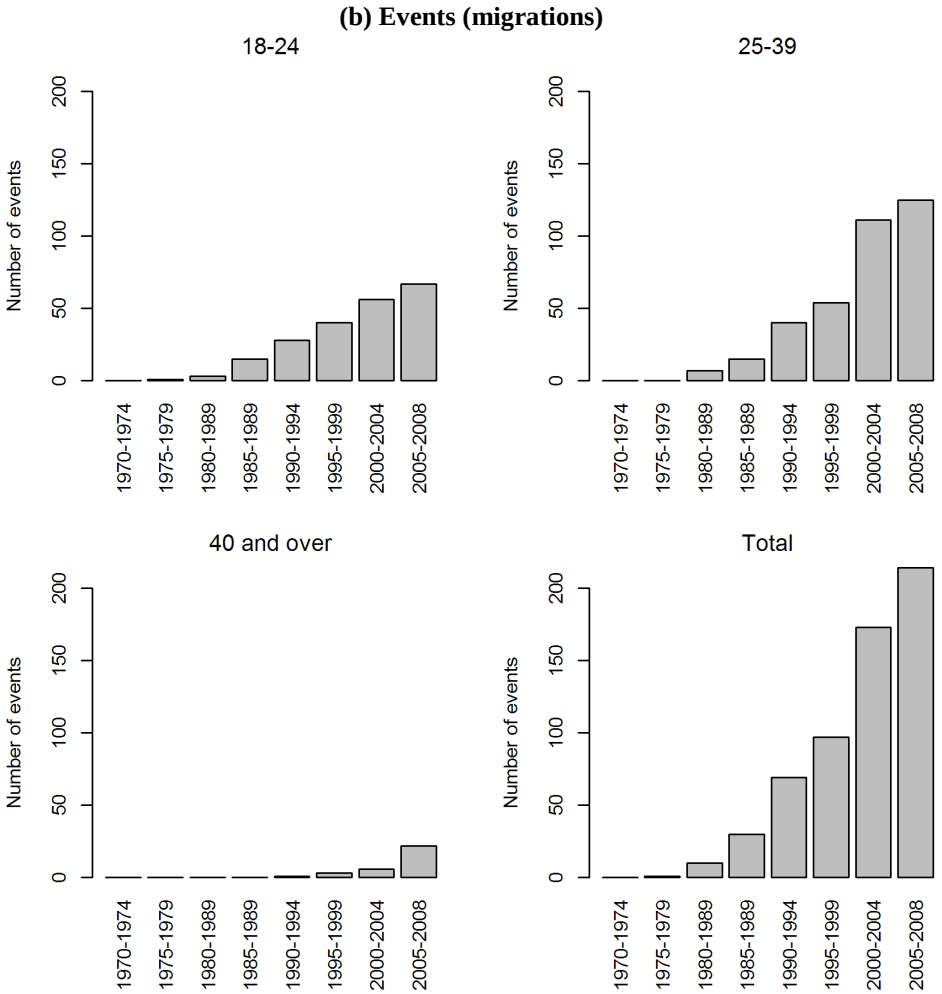


Figure 4: (Continued)



The correlation between parents' migration and children's migration is another potential issue when collecting data on the children through their parents. Stated briefly, if children migrated with their parents (or more generally, if both parents and children have migrated out of the origin country), and if the parents have not returned (and are not in the country to report their children's migration), migration of children will be underreported. However, our data indicate that this is not a serious issue for the

measurement of migration at adult ages (after age 18). The network module of the MAFE surveys conducted in Europe shows that around 90% of the parents of the migrants from the DR Congo, Ghana, and Senegal living in Europe are living in their origin country[16]. In other words, joint migration of adult children and their parents is the exception rather than the rule, and underreporting of migration because of joint migration is low.

3.2 Computing trends

We describe here how trends can be retrospectively computed using information collected in the household surveys on the heads' children.

3.2.1 Models

As in the MMP (Donato 1998), the trends in departure are computed using a discrete time event history model, with only age and period effects (called the age-period model). No other covariate is included, because the purpose of the model is to reconstruct migration trends regardless of the socio-economic characteristics of people[17]. Data are organized as a person-period dataset, in which each individual is represented as many times as the number of years between the time s/he turns 18[18] and (1) the first migration, or (2) age 40, or (3) the time of the survey if the person never migrated and is under 40. The migration variable (dependent variable) takes the value 0 for all years, except for the year of migration (last year in the person-period data file) if the individual migrated (value equal to 1). The statistical model is specified as follows:

$$\log\left(\frac{p_t}{1-p_t}\right) = \alpha + f(\text{age}) + g(\text{period})$$

[16] 93% of the fathers and 82% of the mothers of Congolese migrants living in the UK or Belgium live in the DR Congo; 90% of the fathers and 91% of the mothers of the Ghanaians living in the UK or the Netherlands live in Ghana; and 92% of the fathers and 97% of the mothers of the Senegalese in France, Italy, or Spain live in Senegal.

[17] Additional covariates could be included to reconstruct trends by socio-economic category.

[18] The starting time of observation can vary. Here, we look only at adult migration. If the individual has migrated before age 18, s/he will not be included in the risk set. As a result, the analysis of first migration will be limited to a subsample of people who have not migrated before age 18. This 'filtering' issue is addressed later in the paper (section 4.4.1).

where p_t is the conditional probability of experiencing the event (first migration) at age *t*, given that the event has not already occurred. α is the constant, *f* (age) is a function of age, and *g* (period) is a function of the time period. This model relies on the assumption that the age effect is constant over time. In this paper, age and period effects are estimated using a set of dummy variables. Age is controlled with two age groups (18–24, 25–39)[19], and different specifications of periods are tested. Based on the age effects and the period effects, a cumulative probability of first migration is computed for each period (Donato 1998). The indicator measures the probability of doing at least one international migration before age 40 in a hypothetical cohort that includes the age-specific probabilities of migration within a given period (see Appendix 1 for an illustration of regression coefficients transformed into cumulative migration probabilities).

3.3 Baseline results

The baseline results are those computed using the three simple questions on migration included in the household questionnaire. The population under study is the group of the surviving children of household heads at the time of the surveys. Migration probabilities are computed between ages 18 and 40. Results of the models are reported in Appendix Tables 2 to 7. Appendix Table 11 classifies and names all the models that were tested (including those presented in section 4); the names of these models are reported in the figures and in the results tables.

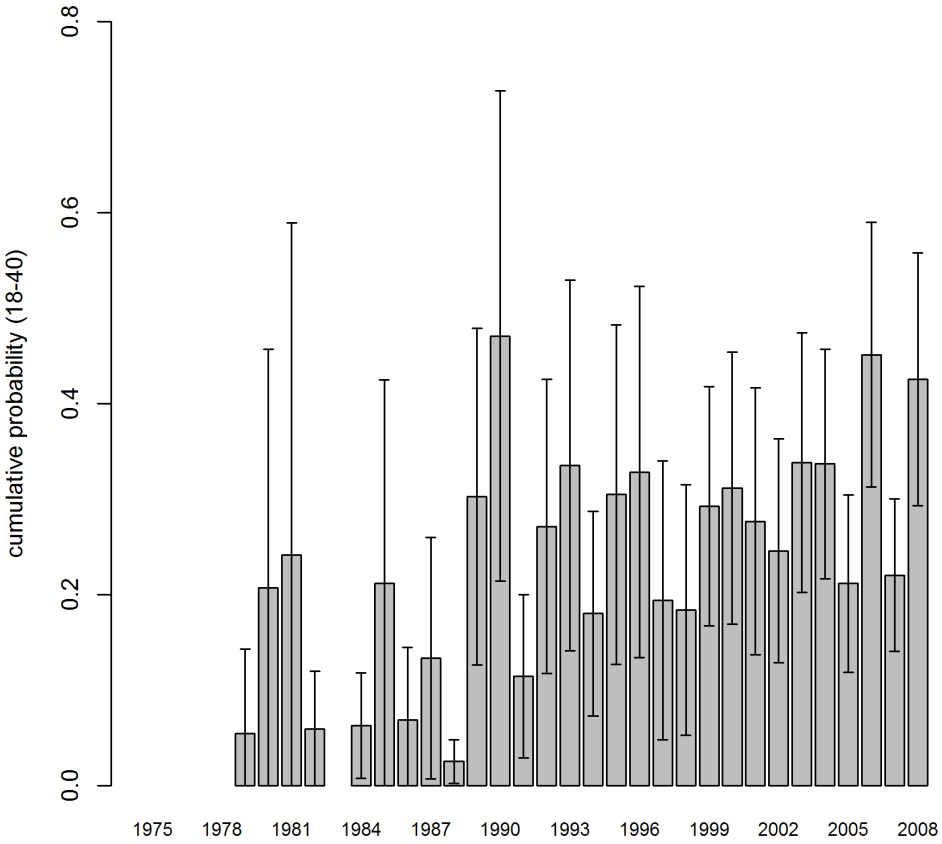
3.3.1 Temporal frame, sample size and precision

In this section we compare estimates obtained with different degrees of temporal detail. Even though the reconstruction of yearly estimates may be desirable for a detailed analysis of migration changes (including the effects of changes in policies or the impact of events like wars and crises), such estimates are hardly reliable. Apart from the fact that data collected from proxy respondents may be inaccurate, reconstructing migration by single year is affected by large sampling errors. Figure 5 shows reconstructed migration trends from the DR Congo between the mid-1970s and the year 2008. As is clear from this figure, confidence intervals are much too large for these estimates to be interpreted in a meaningful way[20]. This shows a limitation of the MAFE data – based

[19] Several functions of age were tested and the results vary very little.
[20] Samples in Senegal and Ghana are smaller, so confidence intervals will be even larger for those countries.

on relatively small samples – but it raises a more general sampling issue because, in almost all contexts, international migration is a rare event.

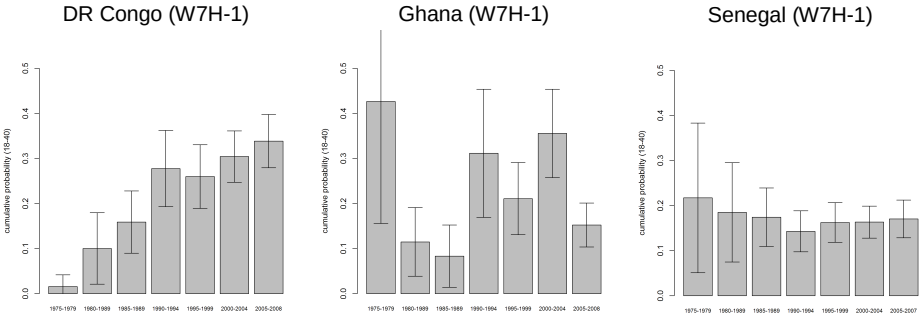
Figure 5: Cumulative probability of 1st departure (18–40) from DR Congo to all destinations by single calendar year (90% confidence intervals). Population: Surviving children of household heads. Data: MAFE DR Congo Household Survey



As expected, confidence intervals are smaller when larger periods are used (Figure 6), but the precision of estimates deteriorates as one goes back in time. The 5-year estimates (7 periods) may be useful in detecting important changes (e.g., migration from Ghana decreased between the late 1970s and the early 1980s, as expected because of the deterioration of the migrants' situation in Nigeria in the late 1970s), but sampling

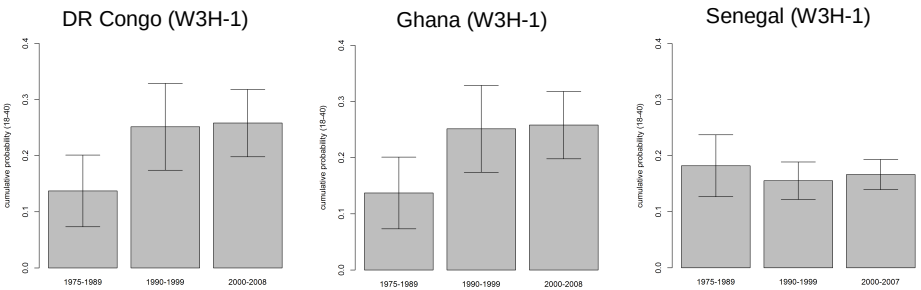
errors are too large (at least before the 1990s) to be used for depicting migration trends in a reliable way. In the rest of this paper, we therefore use three broad periods (of around 10–15 years) [21], as in Figure 7[22].

Figure 6: Cumulative probability of 1st departure (18–40) from DR Congo, Ghana, and Senegal to all destinations by 5-year period (90% confidence intervals).



Population: Surviving children of household heads. Data: MAFE Household Surveys

Figure 7: Cumulative probability of 1st departure from DR Congo, Ghana, and Senegal (age 18–40) by 3 broad periods (90% confidence intervals).



Population: Surviving children of household heads. Data: MAFE Household Surveys

[21] The selection of the 3-period models is based on the size of confidence intervals, as well as on the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). Both AIC and BIC are reported for all the models (Appendix tables). Comparisons of AIC and BIC for the 3-period models and 7-period models show that these indicators tend to be lower for the 3-period models, except in Ghana.

[22] The sensitivity of the trends to shifting the three broad periods by a few years was tested (e.g., 1988–1997 instead of 1990–1999). Estimates vary a little, but the interpretations of the trends are not affected by these changes.

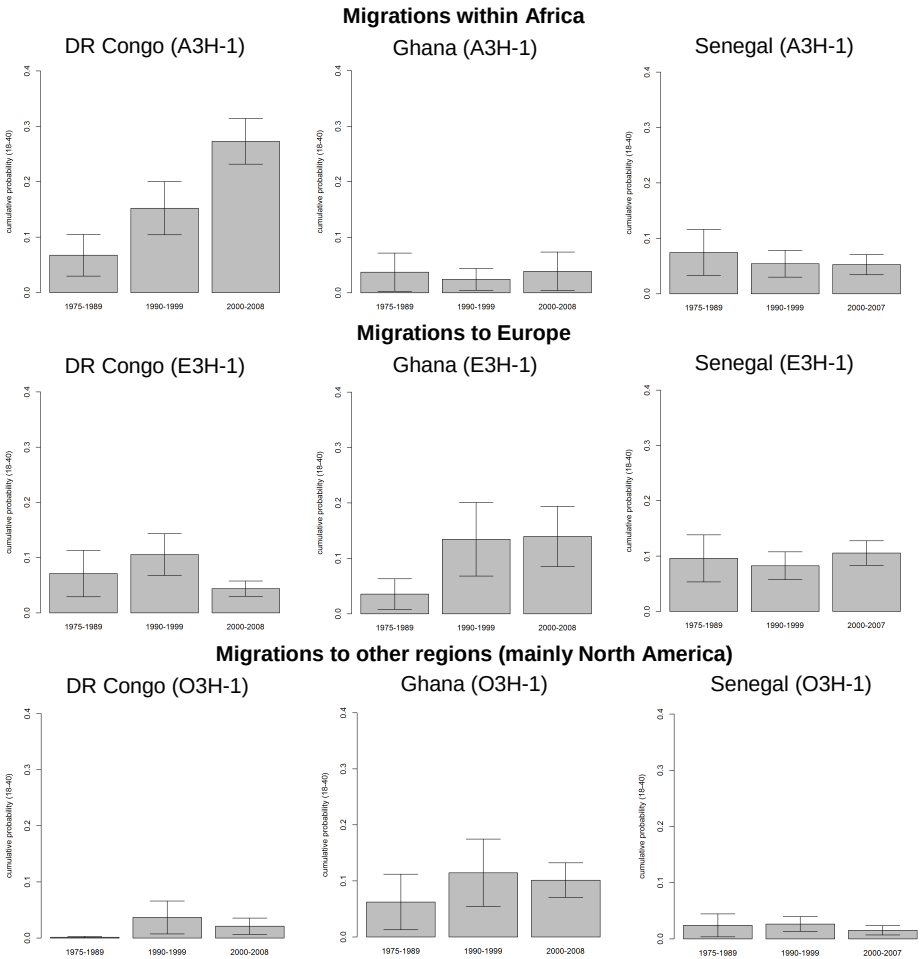
3.3.2 Migration trends by destination

Reconstructing migration trends by destination is also of major importance, both from a policy point of view and from a theoretical point of view. However, because the number of migrations to specific destinations is smaller than migrations to all destinations together, relative sampling errors are larger, especially with smaller periods (results by 5-year period are presented in Appendix 1)[23].

Despite the limitations due to sample size, clear differences are visible by destination (Figure 8). For instance, the increase in Congolese migration results almost exclusively from the increase in African destinations. By contrast, migrations to Europe slightly increased between the 1980s and the 1990s (not significantly), and then decreased significantly ($p<0.01$). Migrations to other regions have remained low. Ghanaian migrations to Europe increased significantly ($p<0.05$) between the 1980s and the 1990s, and slightly decreased in the early 2000s (not significantly). A similar trend is observed for migrations to North America (other regions), but with larger confidence intervals in the 1980s and not statistically significant changes. By contrast, migrations to African destinations remained low throughout the period. Senegalese migrations have not changed in a significant way. In short, despite the large confidence intervals, these data allow us to detect some strong differences and changes.

[23] Even though the country of destination is collected, we consider destinations as broad regions (Europe, Africa, other regions). A separate model is estimated for each destination. The dependent variable is equal to 1 if a migration to the specific destination occurs and 0 otherwise. Migration to another destination is treated as right-censoring.

Figure 8: Cumulative probability of 1st departure (18–40) from DR Congo, Ghana, and Senegal to Europe, Africa, and other regions, by 3 broad periods (90% confidence intervals).



Population: Surviving children of household heads. Data: MAFE Household Surveys.

4. Sensitivity of the estimates

The MAFE data contains additional information that allows us to evaluate the effects on the estimates of some methodological choices made during the collection or analysis stages. In the first section we use additional information from the household questionnaire to test to what extent the results change if we include (or not) deceased children in the risk set. In the second section we take the baseline results computed with the household data and compare them to the results obtained with the same methodology applied to alternative data (network module from the biographic surveys). And in the third section, we assess what we call the 'filtering effects' of age and destination of first migration (i.e., the fact that some information is lost when we collect only the information on the first departure). Finally, the last section combines several corrections to assess the robustness of the baseline results. All the results are reported in Appendix Tables 8 to 10.

4.1 Including deceased children or not?

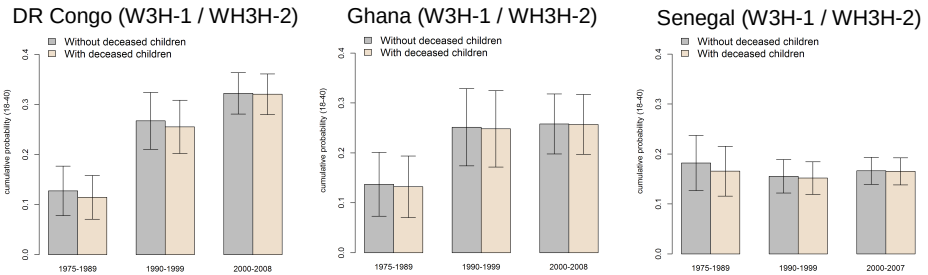
One problem with using retrospective information is that the information is usually only collected for surviving individuals. Using these data to measure trends relies on the assumption that deceased people would have behaved similarly to the surviving people, and/or that their proportion in the total population is small enough to have a minor impact on retrospective estimates. These assumptions are usually thought to be benign, because mortality at adult ages is relatively low, and differential mortality among migrants and non-migrants is not expected to be high at those ages (Massey, Goldring, and Durand 1994). However, in a context such as the DR Congo where life expectancy is low, differential mortality between migrants and non-migrants may have a greater impact on reconstructed trends.

It is possible to quantify – to some extent – the impact of this assumption with the MAFE data. Data were collected on both surviving children and deceased children of the head of household, and the age at death was collected for the deceased children. Even though data on mortality is not perfect (mortality at young ages is usually underreported), we only use information on mortality above age 18, and we are interested in finding an order of magnitude for the impact that mortality has on migration estimates.

Figure 9 shows that the cumulative probabilities of migration do vary when deceased children are taken into account, but differences are much smaller than the confidence intervals. In the three countries, computing outmigration with deceased people included in the data set (until their death) leads to lower estimates. This results

from the negative correlation between mortality and migration (people who died were less likely to migrate). Not surprisingly, differences are larger in earlier periods (around 10% lower in the DR Congo and in Senegal, around 5% in Ghana) because of the higher proportion of deceased people. Differences in the most recent period are negligible. Excluding deceased children thus tends to slightly underestimate any migration increase. However, this underestimation is not statistically significant, which suggests that collecting information on deceased children (a sensitive topic for respondents and interviewers) is perhaps not worthwhile.

Figure 9: Comparisons of cumulative probability of 1st departure (18–40), computed with the household data including or not deceased children (DR Congo, Ghana, Senegal) – all destinations (90% confidence intervals).
Population: children of household heads (surviving or deceased).
Data: MAFE Household Surveys.



Notes: Without deceased children: baseline results.
With deceased children: deceased children included in the risk set until their death.

4.2 Using network data to reconstruct trends: a comparison with household data

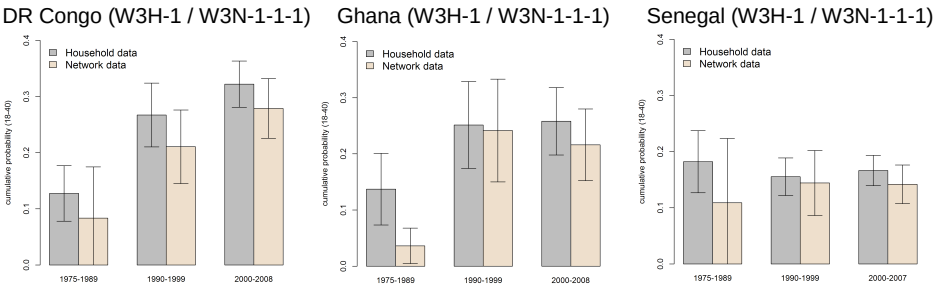
As mentioned before, network data is used to test some of the assumptions of the household survey data. Before doing that we reconstruct first migration trends using the network data and the same method as with the household survey data. We select respondents currently living in the origin countries, and full migration histories of the children of the respondents are obtained from the network module[24]. These migration histories are merged with the birth histories of the respondents, so that the dates of birth

[24] Note that the respondent is not necessarily the head of the household.

(and death) of all the children are available. We then select the first migration of each child, to obtain the same information as collected in the household survey.

Figure 10 compares trends by broad periods from the two sources. Overall, the general trends are broadly similar, but not equal. In the DR Congo the level of migration is lower with the network data for all periods. In Senegal and in Ghana the estimates are much lower with the network data in the first period, but fairly similar in the two recent periods. The reasons for these discrepancies are not entirely clear and could stem from the omission of early migrations in the network module, as well as from the fact that the risk sets are not the same. Even though the results are not dramatically different (except for the first period in Ghana), these differences are a healthy reminder that results may vary from one tool to another.

Figure 10: Comparisons of cumulative probability of 1st departure (18–40), computed with the network data and the household data (DR Congo, Ghana, Senegal) – all destinations (90% confidence intervals).



Notes: Household data: baseline results.
Network data: same approach as the baseline results, with network data of the biographic survey
Population: surviving children of household head (household data) or of respondent (network data).
Data: MAFE Household Surveys and MAFE network modules from the Biographic Surveys.

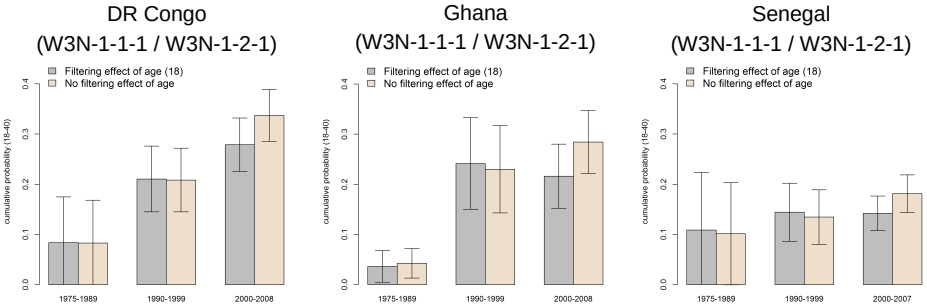
4.3 Filtering effects of first migration: age and destination

4.3.1 Filtering effect of age

In the household survey, information was collected on the first migration of all the children of the head of the household, regardless of their age at the time of their migration. This may be a sensible approach from a data collection point of view, because any additional criteria may complicate data collection and negatively impact data quality. However, one may be interested in the first migration at an adult age (e.g.,

age 18 – as in this paper) and not from birth. Because data is collected on the first migration only, any migration before age 18 will remove the individuals from the risk set for migrations after 18. The baseline results (section 3.1) were computed in this way. A migration at a lower age will thus have a filtering effect, as it prevents migration at a higher age from being recorded. The network module of the biographic survey is used to test the sensitivity of the estimates to this filtering effect. To do this, two series of estimates are computed with the network data. A first series uses the same approach as with the household data; i.e., a respondent who has migrated before age 18 is removed from the risk set. In the second series the respondent is not removed from the risk set after the first migration, so that a person who migrated before 18 can also be included in the computation of migration estimates after age 18. This illustrates the results one would obtain if the question in the household survey were about the first migration from age 18, instead of the first migration regardless of age at migration.

Figure 11: Comparisons of cumulative probability of 1st departure (18–40), computed with and without the filtering effect of age (DR Congo, Ghana, Senegal) – all destinations (90% confidence intervals).



Population: Surviving children of respondents. Data: MAFE network modules from the Biographic Surveys in Africa.
Notes: Filtering effect of age (18): same method as with household data.
No filtering effect of age: results as if the question was asked about the first migration after age 18.

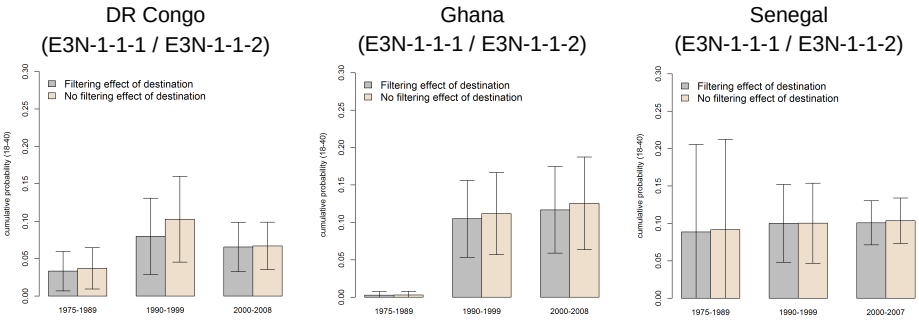
Overall, the trends look broadly similar, but one difference is consistently found in the three countries (Figure 11): the filtering effect of age tends to underestimate the level of recent migration by around 20%. By removing people who had migrated before age 18 from the risk set, people with a greater risk of migration in the recent period are excluded. As a result, the increase in migration may be underestimated with baseline estimates, reinforcing the effect on trends of removing deceased children.

4.3.2 Filtering effect of destination

Analyzing migrations by destination with the household survey data on the first migration may also be affected by a filtering effect. Since only the date and destination of the first migration was collected in the household survey, any migration to a destination other than the one under study leads to censoring. We discuss this issue by considering migration to Europe. If a person moves from the DR Congo to South Africa in 1996, only that migration will be recorded in the household survey. For the event history models of migration to Europe, the person will be in the person-period dataset until the time s/he moves to South Africa (1996), and the dependent variable will be equal to 0 for all the years until the date of censoring. However, the person may come back to DR Congo two years later (1998) and make a move to Europe in 2000. In a similar way, the person may move from South Africa to Europe in 2000. In both cases, the first migration to Europe would be 2000, but this migration does not appear in the household data. This issue is potentially a problem if migration to other destinations is not independent of migration to Europe.

As for the filtering effect of migrations before age 18, the use of network data in the biographic survey allows testing the impact of this factor in the three countries covered by the project. We measure the trends as if the date of the first migration were a destination-specific question (i.e., "In what year did [name] move to [Europe] for the first time for at least one year?"). In this way, any migration to Europe would be identified regardless of whether or not the person had migrated to another region before. This is compared to the trends measured with the network data as if the data were collected in the same way as in the household survey (i.e., "In what year did [name] leave [country] for the first time for at least one year?").

Figure 12: Comparisons of cumulative probability of 1st migration (18–40), with and without the filtering effect of destination (DR Congo, Ghana, Senegal) – migrations to Europe (90% confidence intervals).



Notes: Filtering effect of destination: same method as with household data.

No filtering effect of destination: results as if the question were asked about the first migration to Europe

Population: Surviving children of respondents.

Data: MAFE network modules from the Biographic Surveys in Africa

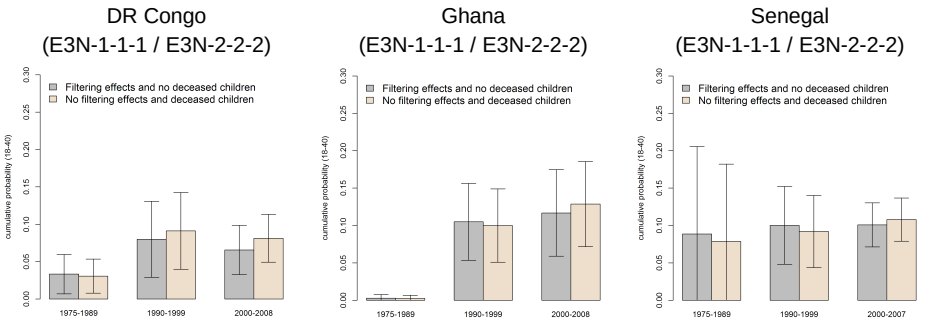
Reassuringly, the general trend is broadly similar to the one with the filtering effect. Migration from Congo to Europe is a little higher in the 1990s without the filtering effect, but results are very close in the 1980s and the early 2000s. This suggests that in the 1990s, people who had lived in another African country before were more likely to move to Europe. In both Ghana and Senegal differences are very small, and chances of migrations to Europe are slightly lower when removing the filtering effect of migrations to other destinations. This results from a lower propensity of migration to Europe among those who had migrated within Africa before. All in all, results for migration to Europe appear quite robust to this filtering effect for the first migration.

Similar comparisons were done for migrations within Africa and to other regions (results not shown). Overall, migration trends from the DR Congo to African destinations are not affected by the filtering effect (people whose first migration is to Europe or North America are not more likely to move back to Africa). By contrast, results in Senegal and Ghana show that the filtering effect tends to slightly underestimate migrations to African destinations, especially in the most recent periods. This suggests that return migrants from Europe are more likely to move to another African destination. However, the probabilities of moving to African countries from Senegal and Ghana are quite small, and actual differences are thus also small.

4.4 Relaxing several assumptions together

We now turn to a comparison of the trends obtained when the three assumptions (deceased children not taken into account, filtering effects of age and destination) are relaxed together. The comparisons are made using the network data, and the focus is on migrations to Europe (Figure 13). The 'filtering effects and no deceased children' option corresponds theoretically to the trends that can be computed when using the three basic questions on first out-migration in the household questionnaire (Figure 2). The other option signals results that can be obtained with additional questions (especially on deceased children and first migration by destination rather than first migration regardless of the destination). In all three countries, using data on the first migration of surviving children without filtering effects tends to overestimate the level of migration in earlier periods (by around 15%–20%) while at the same time underestimating it in recent periods (by around 5%–20%), which results in underestimating any increase in the probabilities of migration. The qualitative conclusions about migration trends are not affected by these differences, in part because the probabilities of migration are relatively small and the sampling errors are large. However, these comparisons indicate that the migration trends are affected by the simplifying assumptions, and they need to be interpreted accordingly.

Figure 13: Comparisons of cumulative probability of 1st migration (18–40), with and without relaxing three assumptions together (DR Congo, Ghana, Senegal) – migrations to Europe (90% confidence intervals).



Population: children of respondents.
Data: MAFE network modules from the Biographic Surveys in Africa.

5. Conclusions

We have shown how simple data collected from household surveys in origin countries allow reconstructing migration trends. In the absence of a 'gold standard' (i.e., representative data that would allow us to estimate the accuracy of migration trends based on the MAFE results), we have evaluated the reliability of the estimates along two axes: their precision and their sensitivity to various methodological choices. Such data provide useful rough estimates of trends in departure that are currently crucially lacking for developing countries. However, the estimates also have some limitations.

First, the confidence intervals are quite large, especially for earlier periods. With sample sizes of 1,000 to 1,500 households, as in the MAFE surveys, one should not expect to be able to reconstruct migration trends in great temporal detail. Confidence intervals would be even larger for analyses on subpopulations (by gender, level of education). In short, these data provide approximate levels of migration and broad trends, which should be taken as qualitative findings. The precision of the estimates could be improved with larger samples. However, if these questions are to be included in existing household surveys (with samples of typically 5,000 to 10,000 households), gains of precision may be partly offset by less frequent migration in rural areas and by the difficulty of oversampling households with migrants.

Secondly, the estimates are sensitive to several methodological choices and assumptions. The estimated levels and trends of migration are influenced by the data collection tool, the inclusion or not of deceased children, and the filtering effects of age and destination. While the broad qualitative conclusions (e.g., overall increase, stability, relative importance of broad destinations) are relatively robust in our examples, these analyses highlight that methodological choices and assumptions bring additional uncertainty to the estimates. In summary, using the three simple questions in household surveys of surviving children should be viewed as a means of providing rough estimates. These estimates can be improved by taking into account deceased children by removing the filtering effect of age (e.g., by asking the question on the first migration after age 18) and by collecting data on the first migration by destination. These improvements would, however, necessitate more questions in the household survey, and they would make data collection more complex. The most complete and complex approach would consist of collecting full migration histories of children, as in the network module of the MAFE biographic surveys. Such data prove to be very flexible for analyzing migration trends. However, the cost and the complexity of collecting such data may limit the feasibility of this approach in multi-topic surveys. The quality of such data also needs to be assessed in detail. Even though the trends obtained from full migration histories of children are broadly consistent with those from the three questions in household surveys, the results of these two approaches are far from perfect

in their matching. Collecting more complex data from proxy respondents may in fact lead to greater underreporting of migration.

Thirdly, estimates are influenced by the reference population from which migration data is collected. We used data collected on children of household heads in order to obtain retrospective migration data on a fixed set of individuals, regardless of their place of residence at the time of the survey. We have also shown that migrations of children at adult ages and their parents are not strongly correlated. However, a drawback of this approach is that little information is available for people beyond age 40, especially in earlier periods. Therefore, migration trends can only be computed for ages 18–40 over the last 30–35 years. In low fertility contexts another potential issue is that the number of children would be low and sampling errors greater. An alternative or complementary approach would be to collect data on the respondent's siblings. This would allow better coverage of higher ages and earlier periods; and the number of siblings is also expected to be greater than the number of children. Sibling survival data is already collected in many Demographic and Health Surveys (Masquelier 2013), and including a few additional questions on migration (such as in Figure 2) may be a cost-effective way of obtaining new data on international migration.

In summary, our experience with the MAFE surveys show that a few simple questions in small-scale household surveys can improve our knowledge about migration levels and trends. However, these estimates are not foolproof and should, wherever possible, be contrasted with other estimates. Further research is also needed to test alternative data collection methods (e.g., among siblings) in various contexts. Testing a few simple approaches in existing household surveys may contribute to tackling the persisting challenge of measuring migration in developing countries (Jensen 2013).

References

Abel, G. (2013). Estimating global migration flow tables using place of birth data. *Demographic Research* 28(18): 505–546. doi:10.4054/DemRes.2013.28.18.

Beauchemin, C. (2011). Rural–urban migration in West Africa: towards a reversal? Migration trends and economic situation in Burkina Faso and Côte d'Ivoire. *Population, Space and Place* 17(1): 47–72. doi:10.1002/psp.573.

Beauchemin, C. (2015, forthcoming). Migrations between Africa and Europe: Rationale for a Survey Design. *Population*.

Bilsborrow, R.E. (2007). *Surveys of International Migration: Issues and Tips*. Paper presented at the Sixth Coordination Meeting on International Migration United Nations, New York, November 26–27 2007.

Bilsborrow, R.E., Hugo, G., Oberai, A., and Zlotnik, H. (1997). *International Migration Statistics: Guidelines for Improving Data Collection Systems*. Geneva: International Labour Office, United Nations Population Fund.

Bocquier, P. (2003). Analyzing Urbanization in Africa. In: Hugo, G. and Champion, A. (eds.). *New Forms of Urbanisation*. Aldershot, UK: Ashgate: 133–150.

Donato, K. (1998). Factors that Influence Migration. Mexican Migration Project Data. In: Mexican Ministry of Foreign Affairs and U.S. Commission on Immigration Reform (eds.). *Migration Between Mexico and the United States: Binational Study. Vol. 3. Research report and background materials*. Mexico City and Washington, D.C.: Mexican Ministry of Foreign Affairs and U.S. Commission on Immigration Reform: 903–964. (https://www.utexas.edu/lbj/uscir/binpapers/v3a-5donato.pdf)

Jensen, E. (2013). A Review of Methods for Estimating Emigration. Washington D.C.: US Census Bureau (Population Division Working Paper 101).

Kasnauskienė, G. and Igoševa, R. (2010). The Practice of Using Household Surveys for Estimating Undeclared Migration. *Ekonomika* 89(3): 122–137.

Masquelier, B. (2013). Adult Mortality From Sibling Survival Data: A Reappraisal of Selection Biases. *Demography* 50(1): 207–228. doi:10.1007/s13524-012-0149-1.

Massey, D., Goldring, L., and Durand, J. (1994). Continuities in Transnational Migration: An Analysis of Nineteen Mexican Communities. *American Journal of Sociology* 99(6): 1492–1533. doi:10.1086/230452.

Massey, D. (1987). The Ethnosurvey in Theory and Practice. *International Migration Review* 21(4): 1498–1522. doi:10.2307/2546522.

OECD (2008). *Perspectives des migrations internationales*. Paris: OECD Publishing. doi:10.1787/migr_outlook-2008-fr.

Parsons, C., Skeldon, R., Walmsley, T., and Winters, A. (2007). Quantifying international migration: a database of bilateral migrant stocks. Washington D.C.: The World Bank (Policy Research Working Paper 4165).

Piché, V., Gregory, J., and Lavoie, J.-P. (1984). L'analyse historique des migrations: la pertinence de l'analyse longitudinale à partir des histoires rétrospectives. *Genus* XI(3–4): 25–45.

Poulain, M., Perrin N., and Singleton, A. (eds.) (2006). *Towards Harmonised European Statistics on International Migration*. Louvain-la-Neuve: Presses universitaires de Louvain.

Schoumaker, B. and Mezger, C. (2013). Sampling and Computation of Weights in the MAFE Surveys. Paris: INED (MAFE Methodological note 6).

Ünalan, T. (2005). Definition of Household Membership in International Migration Surveys. *Journal of Social Sciences* 1(4): 220–225.

Woodrow-Lafield, K.A. (1996). Emigration from the USA: Multiplicity Survey Evidence. *Population Research and Policy Review* 15(2): 171–199. doi:10.1007/BF00126136.

Zaba, B. (1987). The Indirect Estimation of Migration: A Critical Review. *International Migration Review* 21(4): 1395–1445. doi:10.2307/2546519.

Appendix

Cumulative probabilities of first migration between ages 18 and 40 are computed from regression coefficients of the discrete-time event history models. The following example shows a model with two age groups (18–24 and 25–39) and three periods (1975–1989, 1990–1999, and 2000–2008). The age effects and period effects are estimated with dummy variables; the reference categories are the 18–24 age group and 1975–1989 period. The model is written as follows, where log-odds are a linear function of regression coefficients and dummy variables:

$$\log\left(\frac{p_t}{1-p_t}\right) = \alpha + b_1.D_{25-39} + b_2.D_{1990-99} + b_3.D_{2000-08} \quad \text{A.1}$$

The log-odds can be transformed into conditional probabilities of migrating at age t (p_t) in the following way:

$$p_t = \frac{\exp(\alpha + b_1.D_{25-39} + b_2.D_{1990-99} + b_3.D_{2000-08})}{1 + \exp(\alpha + b_1.D_{25-39} + b_2.D_{1990-99} + b_3.D_{2000-08})} \quad \text{A.2}$$

For instance, the probability of migration at age 19 in the first period ($D_{25-39} = 0$, $D_{1990-99} = 0$ and $D_{2000-08} = 0$) is:

$$p_{19} = \frac{\exp(\alpha)}{1 + \exp(\alpha)} \quad \text{A.3}$$

The probability of migration at age 35 in the period 2000–2008 ($D_{25-39} = 1$, $D_{1990-99} = 0$ and $D_{2000-08} = 1$) is:

$$p_{35} = \frac{\exp(\alpha + b_1 + b_3)}{1 + \exp(\alpha + b_1 + b_3)} \quad \text{A.4}$$

Probabilities of migration are computed for each period and each age. For a given period, the "survival" probability between ages 18 and 40 (probability of not migrating) is obtained by multiplying the survival probabilities at each age (between 18 and 39).

$$\hat{S}_{40} = \prod_{t=18}^{39}(1 - p_t) \quad \text{A.5}$$

Since the conditional probability of migration (p_t) is constant within each age group (18–24 and 25–39), the survival probability is also equal to:

$$\hat{S}_{40} = (1 - p_{18-24})^7 . (1 - p_{25-39})^{15} \quad \text{A.6}$$

The cumulative probability of migration (CPM) is equal to:

$$\text{CPM}_{18-40} = 1 - \hat{S}_{40} \quad \text{A.7}$$

Table Appendix 1 illustrates the computation of cumulative probabilities of migration from the DR Congo to all destinations by three broad periods using household survey data (model W3H-1). Regression coefficients are used to compute log-odds by age group and period (Equation A.1). The log odds are transformed into probabilities of migration as in Equation A.2. Probabilities of migration at each age are then transformed into cumulative probabilities of migrating between ages 18 and 40, using Equations A.6 and A.7. In this example, the cumulative probabilities of an international migration between ages 18 and 40 increased from 0.13 (1975–1979) to 0.32 (2000–2008).

Table A-1: Discrete-time event history model of first departure (18–40) from DR Congo by destination, three broad periods. Illustration of the transformation of the regression coefficients into cumulative probabilities of migration.
Population: Children of household heads.
Data: MAFE Household Surveys.

	All destinations			Period		
	(W3H-1)		Age group	1975–1989	1990–1999	2000–2008
Constant	-5.41	Log odds	18–24	-5.41	-4.58	-4.36
18–24 (REF)	–		25–39	-4.96	-4.13	-3.90
25–39	0.46					
1975–1989 (REF)	–	Probabilities (p_t)	18–24	0.0044	0.0101	0.0126
1990–1999	0.83		25–39	0.0070	0.0159	0.0198
2000–2008	1.05					
		Cumulative probabilities	18–40	0.13	0.27	0.32

Table A-2: Discrete-time event history models of first departure (18–40) from DR Congo by destination, three broad periods. Results presented as regression coefficients and cumulative probabilities (18–40). Population: Children of household heads. Data: MAFE Household Surveys.

	DR Congo – 3 periods: regression coefficients and standard errors			
	All destinations (W3H-1)	**Africa (A3H-1)**	**Europe (E3H-1)**	**Others (O3H-1)**
Constant	-5.41 (0.25)	-6.02 (0.35)	-6.41 (0.38)	-9.62 (0.57)
18–24 (REF)	–	–		
25–39	0.46 (0.16)	0.37 (0.19)	0.93 (0.30)	0.14 (0.60)
1975–1989 (REF)	–	–	–	–
1990–1999	0.83 (0.28)	0.86 (0.39)	0.41 (0.41)	3.15 (0.72)
2000–2008	1.05 (0.27)	1.52 (0.36)	-0.50 (0.42)	2.59 (0.67)
N (person-years)	38859	38859	38859	38859
N (individuals)	3697	3697	3697	3697
AIC	5209	4028	1270	606
BIC	5243	4068	1305	640

	DR Congo – 3 periods: cumulative probabilities (18–40) and standard errors			
	All destinations (W3H-1)	**Africa (A3H-1)**	**Europe (E3H-1)**	**Others (O3H-1)**
1975–1989	0.13 (0.03)	0.07 (0.02)	0.07 (0.03)	0.00 (0.00)
1990–1999	0.27 (0.03)	0.15 (0.03)	0.11 (0.02)	0.04 (0.02)
2000–2008	0.32 (0.03)	0.27 (0.03)	0.04 (0.01)	0.02 (0.01)

Note: AIC: Akaike information criterion; BIC: Bayesian information criterion.

Table A-3: Discrete-time event history models of first departure (18–40) from DR Congo by destination, seven 5-year periods. Results presented as regression coefficients and cumulative probabilities (18–40). Population: Children of household heads. Data: MAFE Household Surveys.

	DR Congo – 7 periods: regression coefficients and standard errors			
	All destinations (W7H-1)	**Africa (A7H-1)**	**Europe (E7H-1)**	**Others (O7H-1)**
Constant	-7.55 (1)	-4.33 (0.2)	-7.68 (1.01)	
18–24 (REF)	–	–	–	
25–39	0.45 (0.16)	0.35 (0.19)	0.97 (0.3)	
1975–1979 (REF)	–	–	–	
1980–1984	1.89 (1.12)	–	1.26 (1.26)	
1985–1989	2.39 (1.04)	-2.21 (0.66)	1.36 (1.08)	
1990–1994	3.02 (1.02)	-1.34 (0.42)	2.03 (1.04)	
1995–1999	2.94 (1.02)	-1.13 (0.36)	1.35 (1.05)	
2000–2004	3.13 (1.01)	-0.65 (0.28)	1.24 (1.04)	
2005–2008	3.26 (1.01)	-0.32 (0.22)	-0.2 (1.11)	
N (person-years)	38859	38859	38859	
N (individuals)	3697	3697	3697	
AIC	5211	4019	1257	
BIC	5279	4079	1326	

	DR Congo – 7 periods: cumulative probabilities (18–40) and standard errors			
	All destinations (W7H-1)	**Africa (A7H-1)**	**Europe (E7H-1)**	**Others (O7H-1)**
1975–1979	0.02 (0.02)	–	0.02 (0.02)	
1980–1984	0.10 (0.05)	0.04 (0.02)	0.07 (0.05)	
1985–1989	0.16 (0.04)	0.09 (0.03)	0.08 (0.03)	
1990–1994	0.28 (0.05)	0.11 (0.04)	0.15 (0.04)	
1995–1999	0.26 (0.04)	0.17 (0.04)	0.08 (0.02)	
2000–2004	0.30 (0.03)	0.23 (0.03)	0.07 (0.02)	
2005–2008	0.34 (0.04)	0.31 (0.04)	0.02 (0.01)	

Table A-4: Discrete-time event history models of first departure (18–40) from Ghana by destination, three broad periods. Results presented as regression coefficients and cumulative probabilities (18–40). Population: Children of household heads. Data: MAFE Household Surveys.

	Ghana – 3 periods: regression coefficients and standard errors			
	All destinations (W3H-1)	**Africa (A3H-1)**	**Europe (E3H-1)**	**Others (O3H-1)**
Constant	-5.58 (0.32)	-6.27 (0.61)	-7.10 (0.52)	-6.60 (0.48)
18–24 (REF)	–	–	–	–
25–39	0.76 (0.24)	-0.13 (0.67)	0.90 (0.38)	0.99 (0.38)
1975–1989 (REF)	–	–	–	–
1990–1999	0.68 (0.37)	-0.43 (0.76)	1.39 (0.58)	0.64 (0.56)
2000–2008	0.71 (0.34)	0.04 (0.70)	1.43 (0.52)	0.51 (0.54)
N (person-years)	17198	17198	17198	17198
N (individuals)	1684	1684	1684	1684
AIC	2285	473	1226	1011
BIC	2316	504	1257	1042

	Ghana – 3 periods: cumulative probabilities (18–40) and standard errors			
	All destinations (W3H-1)	**Africa (A3H-1)**	**Europe (E3H-1)**	**Others (O3H-1)**
1975–1989	0.14 (0.04)	0.04 (0.02)	0.04 (0.02)	0.06 (0.03)
1990–1999	0.25 (0.05)	0.02 (0.01)	0.13 (0.04)	0.11 (0.04)
2000–2008	0.26 (0.04)	0.04 (0.02)	0.14 (0.03)	0.10 (0.02)

Table A-5: Discrete-time event history models of first departure (18–40) from Ghana by destination, seven 5-year periods. Results presented as regression coefficients and cumulative probabilities (18–40). Population: Children of household heads. Data: MAFE Household Surveys.

	Ghana – 7 periods: regression coefficients and standard errors			
	All destinations (W7H-1)	**Africa (A7H-1)**	**Europe (E7H-1)**	**Others (O7H-1)**
Constant	-4.29 (0.52)	-4.54 (0.7)	-6.10 (0.85)	-6.21 (1.02)
18–24 (REF)	–	–	–	–
25–39	0.82 (0.25)	0 (0.7)	0.95 (0.38)	0.99 (0.38)
1975–1979 (REF)	–	–	–	–
1980–1984	-1.53 (0.66)	-2.12 (0.87)	-0.60 (1.14)	-1.16 (1.23)
1985–1989	-1.87 (0.72)	–	-1.86 (1.18)	-0.25 (1.16)
1990–1994	-0.4 (0.62)	-3.47 (1.03)	0.94 (0.97)	0.13 (1.12)
1995–1999	-0.86 (0.57)	-1.83 (0.95)	-0.43 (0.92)	0.32 (1.06)
2000–2004	-0.24 (0.55)	-1.21 (0.84)	0.83 (0.88)	0.31 (1.07)
2005–2008	-1.22 (0.56)	-3.08 (0.85)	-0.37 (0.92)	-0.11 (1.06)
N (person-years)	17198	17198	17198	17198
N (individuals)	1684	1684	1684	1684
AIC	2253	450	1198	1016
BIC	2315	503	1260	1078

	Ghana – 7 periods: cumulative probabilities (18–40) and standard errors			
	All destinations (W7H-1)	**Africa (A7H-1)**	**Europe (E7H-1)**	**Others (O7H-1)**
1975–1979	0.43 (0.16)	0.21 (0.14)	0.10 (0.08)	0.09 (0.09)
1980–1984	0.11 (0.05)	0.03 (0.02)	0.06 (0.04)	0.03 (0.02)
1985–1989	0.08 (0.04)	–	0.02 (0.01)	0.07 (0.04)
1990–1994	0.31 (0.09)	0.01 (0.01)	0.23 (0.09)	0.10 (0.05)
1995–1999	0.21 (0.05)	0.04 (0.02)	0.06 (0.02)	0.12 (0.05)
2000–2004	0.36 (0.06)	0.07 (0.04)	0.21 (0.06)	0.12 (0.03)
2005–2008	0.15 (0.03)	0.01 (0.01)	0.07 (0.02)	0.08 (0.02)

Table A-6: Discrete-time event history models of first departure (18–40) from Senegal by destination, three broad periods. Results presented as regression coefficients and cumulative probabilities (18–40). Population: Children of household heads. Data: MAFE Household Surveys.

	Senegal – 3 periods: regression coefficients and standard errors			
	All destinations (W3H-1)	**Africa (A3H-1)**	**Europe (E3H-1)**	**Others (O3H-1)**
Constant	-4.76 (0.2)	-5.92 (0.33)	-5.34 (0.28)	-6.90 (0.56)
18–24 (REF)	–	–	–	–
25–39	0.10 (0.16)	0.38 (0.32)	-0.06 (0.20)	0.14 (0.38)
1975–1989 (REF)	–	–	–	–
1990–1999	-0.18 (0.23)	-0.33 (0.4)	-0.16 (0.33)	0.09 (0.59)
2000–2007	-0.10 (0.23)	-0.36 (0.41)	0.10 (0.31)	-0.45 (0.6)
N (person-years)	31264	31264	31264	31264
N (individuals)	2940	2940	2940	2940
AIC	2614	949	1676	412
BIC	2647	983	1710	445

	Senegal – 3 periods: cumulative probabilities (18–40) and standard errors			
	All destinations (W3H-1)	**Africa (A3H-1)**	**Europe (E3H-1)**	**Others (O3H-1)**
1975–1989	0.18 (0.03)	0.07 (0.03)	0.10 (0.03)	0.02 (0.01)
1990–1999	0.16 (0.02)	0.05 (0.01)	0.08 (0.02)	0.03 (0.01)
2000–2007	0.17 (0.02)	0.05 (0.01)	0.11 (0.01)	0.02 (0.01)

Table A-7: Discrete-time event history models of first departure (18–40) from Senegal by destination, seven 5-year periods. Results presented as regression coefficients and cumulative probabilities (18–40). Population: Children of household heads. Data: MAFE Household Surveys.

	Senegal – 7 periods: regression coefficients and standard errors			
	All destinations (W7H-1)	**Africa (A7H-1)**	**Europe (E7H-1)**	**Others (O7H-1)**
Constant	-4.56 (0.52)	-5.39 (0.87)	-5.55 (0.71)	
18–24 (REF)	–	–	–	
25–39	0.09 (0.16)	0.37 (0.32)	-0.07 (0.2)	
1975–1979 (REF)	–	–	–	
1980–1984	-0.18 (0.66)	-0.61 (1.03)	0.44 (0.89)	
1985–1989	-0.25 (0.57)	-0.61 (0.98)	0.12 (0.78)	
1990–1994	-0.46 (0.56)	-1.09 (0.97)	-0.13 (0.78)	
1995–1999	-0.32 (0.55)	-0.73 (0.94)	0.15 (0.74)	
2000–2004	-0.32 (0.54)	-0.85 (0.94)	0.25 (0.73)	
2005–2007	-0.27 (0.55)	-0.92 (0.98)	0.38 (0.73)	
N (person-years)	31264	31264	31264	
N (individuals)	2940	2940	2940	
AIC	2621	956	1683	
BIC	2688	1023	1750	

	Senegal – 7 periods: cumulative probabilities (18–40) and standard errors			
	All destinations (W7H-1)	**Africa (A7H-1)**	**Europe (E7H-1)**	**Others (O7H-1)**
1975–1979	0.22 (0.10)	0.12 (0.1)	0.08 (0.05)	
1980–1984	0.19 (0.07)	0.07 (0.04)	0.12 (0.06)	
1985–1989	0.17 (0.04)	0.07 (0.03)	0.09 (0.03)	
1990–1994	0.14 (0.03)	0.04 (0.02)	0.07 (0.02)	
1995–1999	0.16 (0.03)	0.06 (0.02)	0.09 (0.02)	
2000–2004	0.16 (0.02)	0.05 (0.02)	0.1 (0.02)	
2005–2007	0.17 (0.03)	0.05 (0.02)		

Table A-8: Discrete-time event history models of first departure (18–40) from DR Congo, three broad periods. Model comparisons estimated with various types of data. Results presented as regression coefficients and cumulative probabilities (18–40).
Population: surviving children of household head (household data) or of respondent (network data).
Data: MAFE Household Surveys and MAFE network modules from the Biographic Surveys.

	DR Congo – 3 periods : regression coefficients and standard errors						
	All destinations Household data, without deceased children	All destinations Household data, with deceased children	All destinations Network data, without deceased children, filtering effects of age and destination	All destinations Network data, without deceased children, no filtering effect of age, filtering effect of destination	Europe Network data, without deceased children, filtering effects of age and destination	Europe Network data, without deceased children, filtering effect of age, no filtering effect of destination	Europe Network data, with deceased children, no filtering effects of age and destination
Model name	(W3H-1)	(W3H-2)	(W3N-1-1-1)	(W3N-1-2-1)	(E3N-1-1-1)	(E3N-1-1-2)	(E3N-2-2-2)
Constant	-5.41 (0.25)	-5.52 (0.24)	-5.75 (0.71)	-5.63 (0.66)	-7.42 (0.54)	-7.48 (0.54)	-7.36 (0.48)
18–24 (REF)	–	–	–	–	–	–	–
25–39	0.46 (0.16)	0.44 (0.16)	0.31 (0.22)	0.14 (0.19)	1.20 (0.53)	1.39 (0.52)	0.98 (0.41)
1975–1989 (REF)	–	–	–	–	–	–	–
1990–1999	0.83 (0.28)	0.89 (0.27)	0.98 (0.73)	0.98 (0.69)	0.90 (0.59)	1.06 (0.55)	1.15 (0.54)
2000–2008	1.05 (0.27)	1.16 (0.26)	1.33 (0.70)	1.57 (0.66)	0.70 (0.57)	0.61 (0.55)	1.1 (0.51)
N (person-years)	38859	40705	20659	21343	20659	22237	23207
N (individuals)	3697	3851	2178	2269	2178	2252	2344
AIC	5209	5309	3371	4026	910	1038	1249
BIC	5243	5343	3403	4058	942	1070	1281
	DR Congo – 3 periods: cumulative probabilities (18–40) and standard errors						
1975–1989	0.13 (0.03)	0.11 (0.03)	0.08 (0.06)	0.05 (1.60)	0.03 (0.02)	0.04 (0.02)	0.03 (0.01)
1990–1999	0.27 (0.03go)	0.26 (0.03)	0.21 (0.04)	0.21 (0.04)	0.08 (0.03)	0.10 (0.03)	0.09 (0.03)
2000–2008	0.32 (0.03)	0.32 (0.02)	0.28 (0.03)	0.34 (0.03)	0.07 (0.02)	0.07 (0.02)	0.09 (0.02)

Table A-9: Discrete-time event history models of first departure (18–40) from Ghana, three broad periods. Model comparisons estimated with various types of data. Results presented as regression coefficients and cumulative probabilities (18–40).
Population: surviving children of household head (household data) or of respondent (network data).
Data: MAFE Household Surveys and MAFE network modules from the Biographic Surveys.

	Ghana – 3 periods: regression coefficients and standard errors						
	All destinations Household data, without deceased children	All destinations Household data, with deceased children	All destinations Network data, without deceased children, filtering effects of age and destination	All destinations Network data, without deceased children, no filtering effect of age, filtering effect of destination	Europe Network data, without deceased children, filtering effects of age and destination	Europe Network data, without deceased children, filtering effect of age, no filtering effect of destination	Europe Network data, with deceased children, no filtering effects of age and destination
Model name	(W3H-1)	(W3H-2)	(W3N-1-1-1)	(W3N-1-2-1)	(E3N-1-1-1)	(E3N-1-1-2)	(E3N-2-2-2)
Constant	-5.58 (0.32)	-5.61 (0.32)	-7.15 (0.53)	-6.75 (0.42)	-9.71 (1.02)	-9.7 (1.02)	-9.66 (1.01)
18–24 (REF)	–	–	–	–	–	–	–
25–39	0.76 (0.24)	0.76 (0.24)	0.98 (0.31)	0.7 (0.27)	1.03 (0.47)	0.97 (0.47)	0.71 (0.41)
1975–1989 (REF)	–	–	–	–	–	–	–
1990–1999	0.68 (0.37)	0.7 (0.37)	2.02 (0.57)	1.81 (0.47)	3.69 (1.04)	3.67 (1.04)	3.77 (1.04)
2000–2008	0.71 (0.34)	0.74 (0.34)	1.89 (0.57)	2.05 (0.45)	3.81 (1.06)	3.77 (1.06)	4.08 (1.04)
N (person-years)	17198	17443	11470	12073	11470	12337	12778
N (individuals)	1684	1704	1254	1311	1254	1292	1329
AIC	2285	2289	1538	1918	857	866	970
BIC	2316	2320	1568	1948	886	896	1000

	Ghana – 3 periods: cumulative probabilities (18–40) and standard errors						
1975–1989	0.14 (0.04)	0.13 (0.04)	0.04 (0.02)	0.04 (0.02)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
1990–1999	0.25 (0.05)	0.25 (0.05)	0.24 (0.06)	0.23 (0.05)	0.11 (0.03)	0.11 (0.03)	0.10 (0.03)
2000–2008	0.26 (0.04)	0.26 (0.04)	0.22 (0.04)	0.28 (0.04)	0.13 (0.04)	0.12 (0.04)	0.13 (0.03)

Table A-10: Discrete-time event history models of first departure (18–40) from Senegal, three broad periods. Model comparisons estimated with various types of data. Results presented as regression coefficients and cumulative probabilities (18–40).
Population: surviving children of household head (household data) or of respondent (network data).
Data: MAFE Household Surveys and MAFE network modules from the Biographic Surveys.

	Senegal – 3 periods: regression coefficients and standard errors						
	All destinations Household data, without deceased children	All destinations Household data, with deceased children	All destinations Network data, without deceased children, filtering effects of age and destination	All destinations Network data, without deceased children, no filtering effect of age, filtering effect of destination	Europe Network data, without deceased children, filtering effects of age and destination	Europe Network data, without deceased children, filtering effect of age, no filtering effect of destination	Europe Network data, with deceased children, no filtering effects of age and destination
Model name	(W3H-1)	(W3H-2)	(W3N-1-1-1)	(W3N-1-2-1)	(E3N-1-1-1)	(E3N-1-1-2)	(E3N-2-2-2)
Constant	-4.76 (0.2)	-4.84 (0.2)	-5.52 (0.56)	-5.44 (0.54)	-5.76 (0.67)	-5.77 (0.67)	-5.85 (0.67)
18–24 (REF)	–	–	–	–	–	–	–
25–39	0.10 (0.16)	0.07 (0.16)	0.37 (0.31)	0.17 (0.29)	0.45 (0.4)	0.42 (0.39)	0.37 (0.37)
1975–1989 (REF)	–	–	–	–	–	–	–
1990–1999	-0.18 (0.23)	-0.10 (0.23)	0.29 (0.69)	0.31 (0.66)	0.06 (0.86)	0.1 (0.86)	0.16 (0.86)
2000–2008	-0.10 (0.23)	0.00 (0.22)	0.29 (0.68)	0.63 (0.64)	0.15 (0.83)	0.16 (0.83)	0.35 (0.82)
N (person-years)	31264	31740	13460	13826	13460	13996	14483
N (individuals)	2940	2983	1377	1419	1377	1400	1443
AIC	2614	2623	1616	1873	1204	1224	1285
BIC	2647	2656	1646	1904	1234	1254	1316
	Senegal – 3 periods: cumulative probabilities (18–40) and standard errors						
1975–1989	0.18 (0.03)	0.17 (0.03)	0.11 (0.07)	0.10 (0.06)	0.09 (0.07)	0.09 (0.07)	0.08 (0.06)
1990–1999	0.16 (0.02)	0.15 (0.02)	0.14 (0.04)	0.14 (0.03)	0.10 (0.03)	0.10 (0.03)	0.09 (0.03)
2000–2008	0.17 (0.02)	0.17 (0.02)	0.14 (0.02)	0.18 (0.02)	0.11 (0.02)	0.10 (0.02)	0.11 (0.02)

Table A-11: Summary table of models tested in the paper, indicating the type of data, time periods, destination and use or not of information on deceased children, and filtering effects of age and destination

Periods for trends	Destination	Household data (H)		Network data (N)							
				Without deceased children (1)				With deceased children (2)			
		Without deceased children (1)	With deceased children (2)	Filtering of age (1)		No filtering of age (2)		Filtering of age (1)		No filtering of age (2)	
				Filtering effect of destination (1)	No filtering effect of destination (2)	Filtering effect of destination (1)	No filtering effect of destination (2)	Filtering effect of destination (1)	No filtering effect of destination (2)	Filtering effect of destination (1)	No filtering effect of destination (2)
3 periods	All destinations (W)	W3H-1	W3H-2	W3N-1-1-1		W3N-1-2-1					
	Africa (A)	A3H-1									
	Europe (E)	E3H-1		E3N-1-1-1	E3N-1-1-2						E3N-2-2-2
	Other destinations (O)	O3H-1									
7 periods	All destinations (W)	W7H-1									
	Africa (A)	A7H-1									
	Europe (E)	E7H-1									
	Other destinations (O)	O7H-1									

Note: Cells with model names indicate models that are fitted and presented in the paper.

Figure A-1: Cumulative probability of 1st departure (18–40) from DR Congo, Ghana, and Senegal by 5-year period and destination* (90% confidence intervals).
Population: Children of household heads.
Data: MAFE Household Surveys

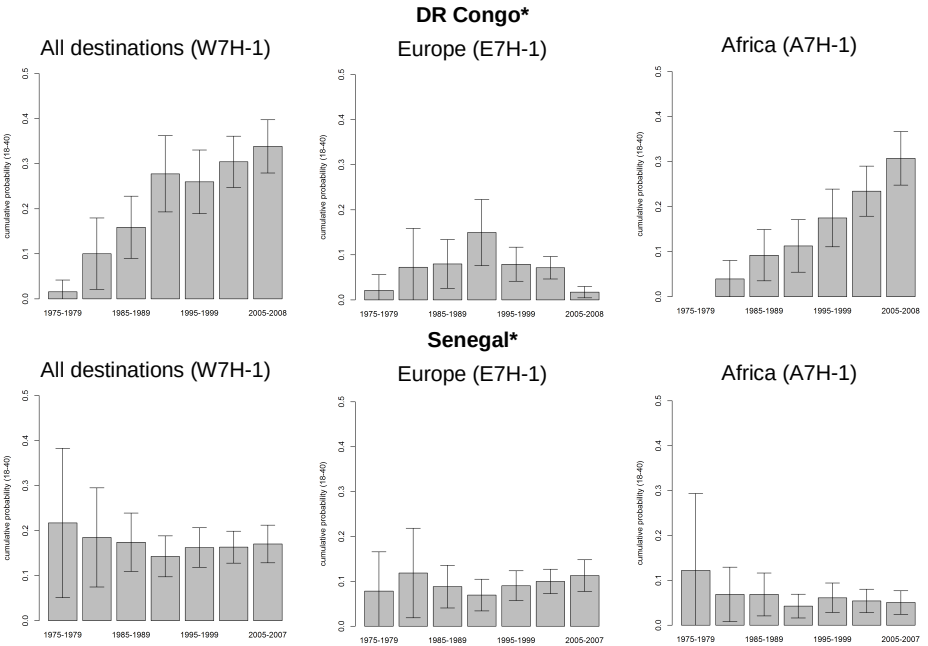
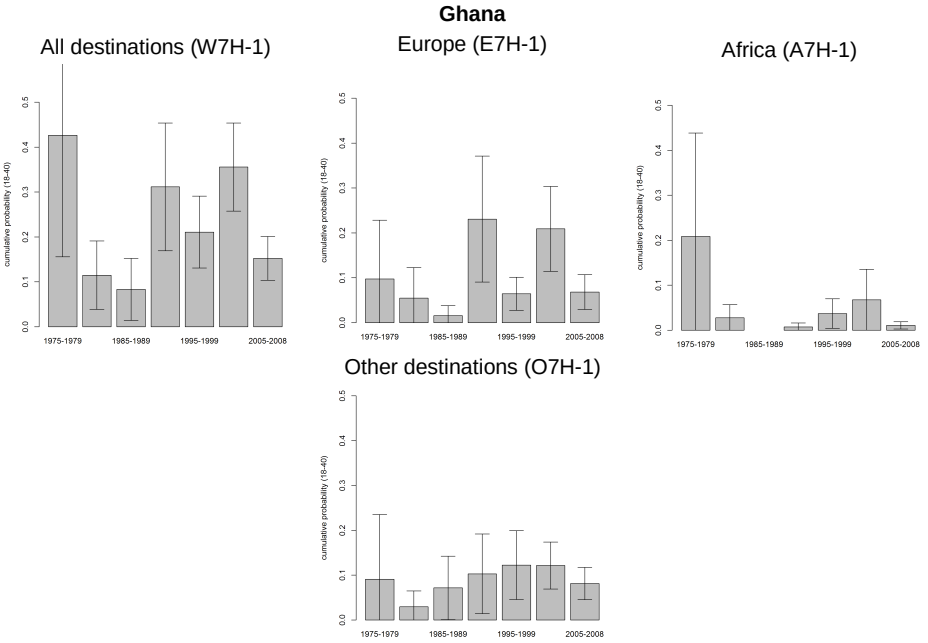


Figure A-1: (Continued)



Notes: * Cumulative probabilities of migration from DR Congo and Senegal to destinations other than Europe and Africa cannot be estimated because of low migration numbers.