Migration Analysis Using Demographic Surveys and Surveillance Systems

10

Philippe Bocquier

Introduction

Most demographic data sources contain information on mortality and fertility, very few on migration. Nonetheless, some demographic sources can be quite helpful for the study of migration, even though migration analysis was not a core objective of the data collection. In some cases it is only a single question, in some others it is a set of more or less complex questions on migration.

This chapter endeavors to review the possibilities of including migration in demographic analysis, either as a dependent variable or as an independent variable, using existing demographic data that were not initially collected to study migration. By giving scholars basic advice regarding the potential and limitations of two major demographic data sources in developing countries, namely Demographic and Health Surveys (DHS) and Health and Demographic Surveillance Systems (HDSS), this chapter is intended to encourage scholars to dare use these sources for migration study. It will therefore not cover two other main demographic sources, census data (cf. Chap. 8 by Sobek) and migration-specific surveys (cf. Chap. 9 by Beauchemin and Schoumaker).

This chapter is organized as follows. Criteria to evaluate data sources as regard to migration analysis are presented in a first section. The second section is devoted to the use of migration data as a determinant, taking DHS in selected West African countries as examples. The third section shows how to use HDSS data to study migration as both an event and a determinant. These sections are illustrated with examples of analyses by the author and by others. The chapter concludes with a synthesis and way forward.

Criteria to Evaluate Data Sources as Regard to Migration Analysis

Because most of the demographic sources were not meant to produce migration indicators, the data collection tools and sampling procedures used to produce these data are not necessarily adequate for migration analysis. This section aims at reviewing the main criteria that can help evaluate the quality of a demographic source as regard to migration analysis. Ability of data sources to estimate migrants' economic, social and demographic characteristics well is not sufficient. The data sources should be gauged on their ability to reliably estimate flows between origins and destinations (migration matrices) and the corresponding rates, which are at the core of migration analysis.

P. Bocquier (✉)
Université Catholique de Louvain, Louvain-la-Neuve, Belgium

University of the Witwatersrand, Johannesburg, South Africa
e-mail: philippe.bocquier@uclouvain.be

M.J. White (ed.), *International Handbook of Migration and Population Distribution*, International Handbooks of Population 6, DOI 10.1007/978-94-017-7282-2_10

Sampling Procedures

Sampling aims at representing as best as possible a living population at a given point in time. Sampling involves random choice of locations, households and sometimes individuals in a given spatiotemporal universe. The sample is meant to represent first current characteristics or behavior of sampling units. How the sample represents past characteristics or behavior, including migratory itineraries, is more challenging and depends a lot on how the current situation (e.g. repartition of migrants) depends on past behavior (e.g. migration histories). The sample will obviously not be representative of the dead if their behavior differed from the living. It will not be representative of the international emigrants either if their behavior differed from the population still in the country. These biases are obvious and well known although difficult to evaluate to the point that, despite evidence of the contrary, most migration analyzes assume independence of mortality and migration or international and internal migration. However, other biases may originate in the sampling procedure even after discarding biases due to death or international migration.

Considering that most demographers use surveys as their main data sources, it is important to consider how migration analysis is restricted or enhanced by sampling choice such as sample size, stratification, and oversampling. Although HDSS are usually not based on samples as such, since they usually involve exhaustive data collection on a population limited by clear geographical boundaries, HDSS will be included in the discussion as they involve a choice in the target population, in terms of geographical area, time-frame, population size, etc. The quality implications are not the same but the same criteria may be used to evaluate these sources.

Apart from censuses and population registers, demographic data collection usually involves the choice of a target population. Ideally, surveys should be representative of both sending and receiving areas (Bilsborrow et al. 1984; Courgeau 1988). Although many national surveys pretend to be representative at the national level, the sample is not necessarily drawn in such a way to be representative of all relevant geographical units in the national space. Typically a national sample will be representative of both rural and urban strata, including perhaps the major urban agglomeration. Very rarely a national sample is representative of all administrative regions of a country, since this involves a substantial increase in sampling size. For the same reason, even rarer are samples that are representative of regions and area (urban versus rural) of residence within these regions.

Because most surveys employ stratified sampling to reduce costs of data collection, household clusters are drawn in large strata. Usually urban strata are overrepresented as main destination areas so that more detailed information is collected on migrants where they live. Typically, clusters of urban households that represent urban areas are drawn within a limited number of cities that are themselves drawn from a list of all cities in the country. Ideally urban households should represent the whole spectrum of city size (from small towns to the capital city) but very often only two or three cities of different sizes are represented. Stratification is also used within rural areas, although the main criterion to classify rural households is not the size but rather the ecological environment. For both urban and rural areas, the stratification strategy results from a trade-off between representativeness and cost that implies that the sample is not often representative at a very low geographical scale.

The choice of strata represents therefore an important limitation for the analysis of migration flows. A sample based on rural and urban strata will be sufficient to analyze urban-rural flows in both directions but not interregional flows. Suppose that a country have ten regions but only four have been sampled. Even if successive residences in all ten regions are collected for all migrants in the sample of four regions, these migration itineraries will not be representative of migration flows to and from the six regions that were not sampled. The same holds if only two cities are sampled among urban areas: the inter-urban migration matrix will not be representative of all flows between the hierarchy of urban areas. To analyze urban-rural migration at the regional level, samples will have to be representative at the regional level, preferably in addition to the urban-rural level. This was done in the Network of Surveys on Migration and Urbanization in West Africa (NESMUWA) conducted in eight West African countries simultaneously in 1993 (Bocquier 2004; Bocquier and Traoré 1998). NESMUWA was a unique set of surveys aiming at measuring both internal and international migration flows in the 5 years preceding data collection.

As shown in Table 10.1, in seven out of eight countries[1] between 20 and 30 strata were defined of which 50–68 % were urban (58 % on average, against 30 % urbanization rate in the population). The sampling of a larger proportion of urban areas was necessary to better evaluate migration flows toward and between different categories of cities classified by size. As a consequence of the high number of strata, sample size was quite large varying from 7364 to 13,292 households. The mean number of households per stratum varied from below 250 (for three countries) to around 500 (for another three countries), Guinea (818) excluded. The number of individuals per stratum varied from 1370 to 3729 with an average of 2455, Guinea (6083) excluded.

Other surveys do not offer such large samples and a sampling procedure that ensure good representation of all urban areas by size category. Nonetheless most DHS surveys oversample urban areas because they are deemed more heterogeneous. This is rather good news as in most countries urban areas are also destinations for migrants. DHS samples may therefore be sufficient to provide origin-destination migration matrices by large rural and urban categories. However a minority will give reliable estimates of migration flows by both administrative region and urban-rural areas.

The Table 10.2 takes the examples of DHS for NESMUWA countries. Very few DHS samples combine a high number of strata (>7), a large sample (>7000 households), strata defined by both administrative region (RG) and urban-rural areas (UR), and oversampled urban areas (higher than 1.1 ratio of sampled urban households share to percent urban in population as estimated by the UN). Only DHS-Guinea 1992, DHS-Mali 1995–1996, and DHS-Niger 2006 fulfill all four criteria. However, all DHS from Mali and Niger fulfill at least three criteria, making time-comparison of three DHS possible for these two countries albeit with some grouping of geographical areas (strata were not defined the same way from one survey to the next). By contrast, analysis of migration matrices will be rather poor for Burkina Faso and Côte d'Ivoire. There were missed opportunities in Guinea where sample size was reduced and urban oversampling abandoned for DHS conducted in 1999 and 2005. For DHS-Burkina Faso 2003 and DHS-Senegal 1999, it might still be possible to construct matrices by both administrative regions and urban-rural areas after careful examination of the mean number of households per urban and rural areas (not available from DHS-MEASURE web site).

In HDSS, there is no sampling issue as such as regard to the population under surveillance. There might be samples drawn from this population but the universe to which these samples refer would be the population under surveillance (or a subset of this population defined along some demographic or socioeconomic criteria) and not any other population. Yet, migration analysis heavily depends on the choice of the population under surveillance. Whereas representativity is not a concern in HDSS, exemplarity is. Idiosyncrasy is something that HDSS analysts should control when possible. Among the possible tools is the use of some control group external to the HDSS, the systematic comparison with the general population (e.g. using census data), or the comparative analysis of several HDSS controlling for macro characteristics of these HDSS. At the very least, one should be careful to refer to idiosyncrasy issue explicitly in the analyzes.

[1] We exclude the survey in Nigeria from the discussion as it followed a very different methodology.

Table 10.1 Number of strata, household and individuals in the NESMUWA surveys

Country	Strata (urban/ total)	Number of household	Mean number of households per stratum	Number of individuals	Mean number of individuals per stratum
Burkina Faso	10/ 20 = 50 %	10,091	504	64,798	3240
Côte d'Ivoire	20/ 30 = 67 %	13,292	443	69,902	2330
Guinea	5/ 9 = 56 %	7364	818	54,750	6083
Mali	15/ 22 = 68 %	10,890	495	82,042	3729
Mauritania	13/ 28 = 46 %	7385	264	51,337	1833
Niger	16/ 30 = 53 %	6870	229	41,095	1370
Senegal	19/ 29 = 66 %	7635	263	64,601	2228
Average	58 %	9075	431	61,218	2973

Source: Bocquier and Traoré 1998

Another consideration is that pointed by Sankoh and Byass (2012) of "whether the final population is defined as being within a contiguous area or in a collection of small areas (e.g. discrete villages or city quarters) within a wider area". This affects the definition of migration events "since local moves in a non-contiguous population may be classified as in- and out-migrations, whereas similar moves in a contiguous area would amount to within-site migrations" (Sankoh and Byass 2012). Even if the HDSS is defined as a contiguous area, migration with neighboring areas will depend on the varying degree of isolation (or its opposite, integration) of the HDSS in a larger area. For example, if the HDSS area is situated in a rather dense web of villages or in a city, then the chance to cross the HDSS boundaries increases. A good indicator of this closeness to neighboring areas is the intensity of marriage-related migration, or marriages that result in migrations. Marriages in isolated populations tend to be more endogamous.

Data Collection Tools

Provided that the sampling procedure is adequate, questionnaires have obviously to include questions on migration. Migration histories from birth to the time of interview as collected in the NESMUWA surveys would be ideal but this is not standard in non-migration-oriented surveys (see Chap. 9 by Beauchemin and Schoumaker in this volume for a review of these surveys). However many surveys include questions on place of birth and on place of previous residence.

Matrices cross-tabulating place of birth with current place of residence may be used to form indicators summarizing lifetime migration. Exposure time to the risk of migration varies from one individual to the other depending on the age of the individual at the time of data collection. In other words, the indicator is heavily right-censored. The indicator will therefore be very dependent on the age structure, unless it is computed by cohort, i.e. for specific age groups. Lifetime migration cohort indicators may be computed and compared over several censuses or surveys, e.g. proportion whose place of residence is different from place of birth, or proportion living in urban area born in rural areas, etc. Apart from the problem of age-control, lifetime migration indicators make the implicit assumption that only one migration occurred (from place of birth to place of current residence). A consequence is an underestimation of migration intensity since migration is a

Table 10.2 Stratification indicators in Demographic and Health Surveys for a select group of West African countries

Country, year	Number of households interviewed			Strata Type (UR = urban-rural, RG = region, IN = intervention)	Urban household share (a)	Percent urban in population (UN) (b)	Urban over-sampling index (a)/(b)	Number of strata			Mean number of households per stratum			Quality index for migration matrices analysis[a]
	Urban	Rural	Total					Urban	Rural	Total	Urban	Rural	Total	
Burkina Faso														
DHS 1993	2431	3275	5706	UR	42.6 %	14.6 %	2.92	2	1	3	1216	3275	1902	1
DHS 1998–1999	1448	3685	5133	UR, IN	28.2 %	17.0 %	1.66	3	2	5	483	1843	1027	1
DHS 2003	2340	7130	9470	RG	24.7 %	20.0 %	1.23	na	na	13	na	na	728	3
Cote d'Ivoire														
DHS 1994	2982	3366	6348	UR	47.0 %	40.8 %	1.15	3	2	5	994	1683	1270	1
DHS 1998–1999	1562	740	2302	UR, IN	67.9 %	42.8 %	1.59	3	2	5	521	370	460	1
Guinea														
DHS 1992	3990	3164	7154	UR, RG	55.8 %	28.6 %	1.95	5	4	9	798	791	795	4
DHS 1999	1714	3751	5465	UR, RG	31.4 %	30.7 %	1.02	5	4	9	343	938	607	2
DHS 2005	1800	4680	6480	UR, RG	27.8 %	32.8 %	0.85	8	7	15	225	669	432	2
Mali														
DHS 1995–1996	3099	6413	9512	UR, RG	32.6 %	25.8 %	1.27	7	6	13	443	1069	732	4
DHS 2001	3136	10,581	13,717	UR, RG	22.9 %	28.7 %	0.80	7	6	13	448	1764	1055	3
DHS 2006	4413	9282	13,695	UR, RG	32.2 %	31.7 %	1.01	9	8	17	490	1160	806	3
Mauritania														
DHS 2000–2001	3979	2779	6758	UR, RG	58.9 %	40.0 %	1.47	5	4	9	796	695	751	3
Niger														
DHS 1992	2261	3558	5819	UR, RG	38.9 %	15.6 %	2.50	8	7	15	283	508	388	3
DHS 1998	1766	4611	6377	UR, RG	27.7 %	16.0 %	1.73	6	5	11	294	922	580	3
DHS 2006	2508	5910	8418	UR, RG	29.8 %	16.9 %	1.77	8	7	15	314	844	561	4
Senegal														
DHS 1992–1993	1587	2148	3735	UR, RG	42.5 %	39.3 %	1.08	4	4	8	397	537	467	2
DHS 1997	1681	3359	5040	RG	33.4 %	39.9 %	0.84	na	na	12	na	na	420	1
DHS 1999	2724	5137	7861	RG	34.7 %	40.2 %	0.86	na	na	50	na	na	157	2
DHS 2005	3320	4539	7859	UR, RG	42.2 %	41.1 %	1.03	11	11	22	302	413	357	3

[a] 1 = poor, 2 = fair, 3 = good, 4 = excellent. This index is the sum of the following 4 criteria favorable to migration flow analysis: a high number of strata (>7), a large sample (>7000 households), strata defined by both administrative region and urban-rural areas (UR, RG), and oversampled urban areas (>1.10)

renewable event. Another consequence is that return migration is not taken into account since place of current residence will be the same as place of birth. In sum, lifetime migration matrices are seriously biased and are not favored by demographers although they are the most widely available data on migration.

The question on place of residence $(t - n)$ years before the survey is sometimes asked in censuses (place of residence at previous census or n years ago). It is not a good question to estimate the number of migrations, because it assumes only one migration over the period. In particular, return migrations are not taken into account. Yet, this question is good to estimate net migration rates over the period n. The question on place of previous residence that captures recent migration (including return migration) was preferred in DHS. Last migration would be representative of all migrations under the strong assumption that migrants migrated only once over the reference period. This assumption is reasonable when the reference period is short, because the shorter the period of reference the higher the probability to migrate only once. Data from the REMUAO surveys may be used to estimate the optimal reference period before survey for the most recent migration. Figure 10.1 represents the distribution of last migration by year before survey and the kernel density of this distribution. Last migration is clearly skewed to the right. Figure 10.2 shows how the distribution spreads to the left the higher the rank of the migration before the survey.

Figure 10.3 shows the distribution of the last migration in four REMUAO countries (both sex, weighted samples). This Figure provides a justification for considering that indicators of place of previous residence should be based preferably on a 3-year reference period before survey. Migration matrices on longer periods will lead to higher underestimation of migration rates. In the four chosen REMUAO countries (Table 10.3, pooled samples), last migrations represent 79.4 % of total migrations recorded in the 3 years before surveys, but their share drops from 89.0 % in the year immediately before the survey to only 69.7 % in the third year before the survey. Moreover, on longer than 3-year reference period, migration indicators based on last migration only will be seriously biased, since they will under-represent migrations of frequent movers, i.e. individuals who migrated more than once in the reference period.

The question on place of previous residence is asked to members of the households at the time of the survey. Members who moved out of the household are not taken into account. This is not a problem for internal migration since movers not counted in households at origin are supposed to be represented in other households at destination. Still, international migrations will be missed since their households at international destinations will not be included in the national sample. In surveys and censuses, international emigration flows can be estimated using a specific questionnaire about members of the household who migrated, say in the 5 past years, to an international destination. This is the technique used in NESMUWA surveys. Under the assumption that emigrants migrated only once during the reference period, this questionnaire on emigrants is a good tool to estimate international out-migration flows, while questions on origin of international immigrant is sufficient to estimate international in-migration flows.

As compared to single question on last migration (or on residence n years ago), migration histories present the obvious advantage of exhaustivity. Indicators based on complete residential itineraries from birth to time of survey will necessarily give unbiased results, barring selection bias from mortality. HDSS usually record all residential moves within the surveillance area and across its boundaries since the beginning of surveillance (first enumeration). Unfortunately, complete migration histories from birth are rarely available for those who were not born in the surveillance area, for the period before the inception of the HDSS, and for in-migrants before their first move in the HDSS. While complete histories are not necessary to compute migration rates, their absence limits the use of past migration experience to explain migrations under surveillance.

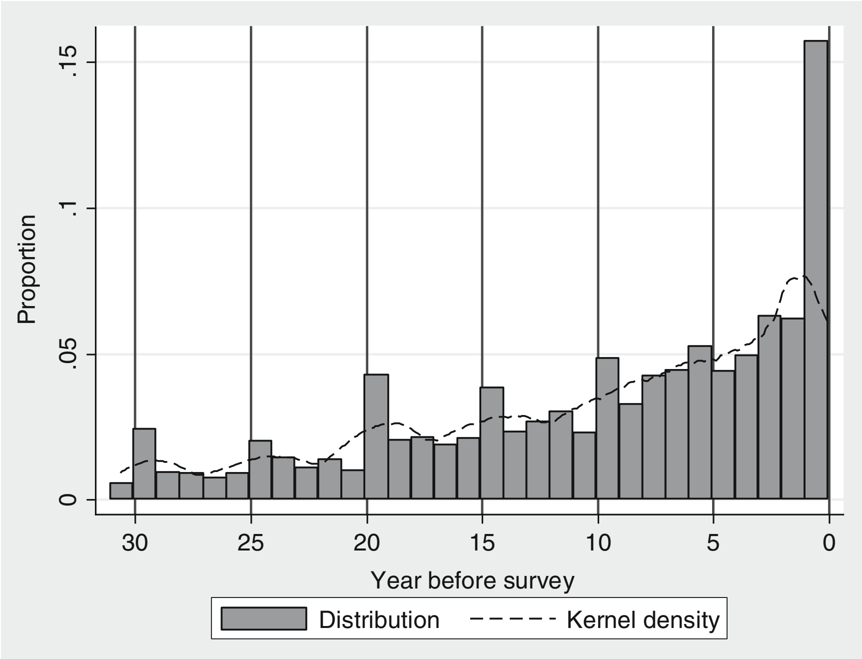

Fig. 10.1 Distribution of last migration– Mali, Female, 15 years and more (Source: REMUAO 1993, unweighted sample, our own computation)

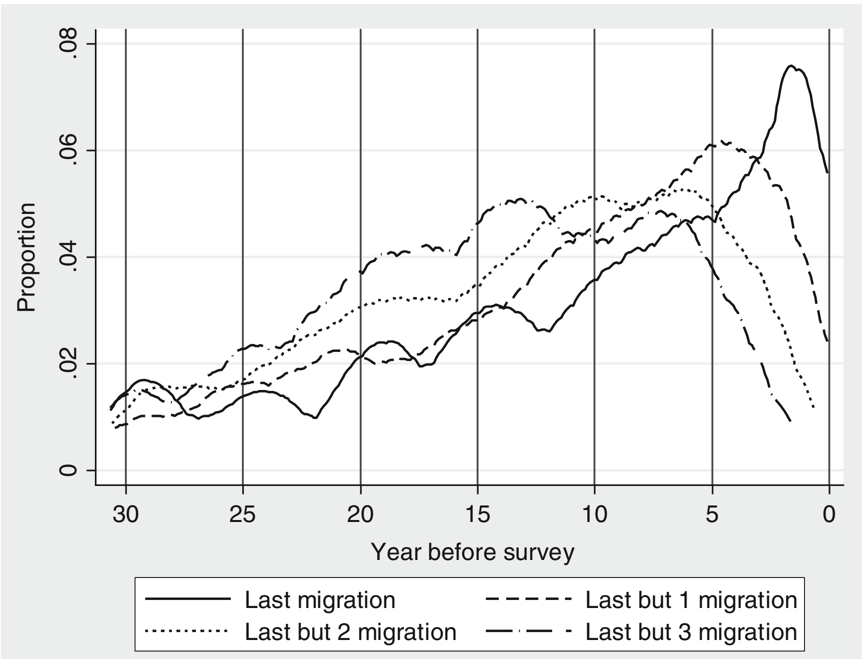

Fig. 10.2 Distribution of migration by rank – Mali, Female, 15 years and more (Source: REMUAO 1993, weighted sample, our own computation)

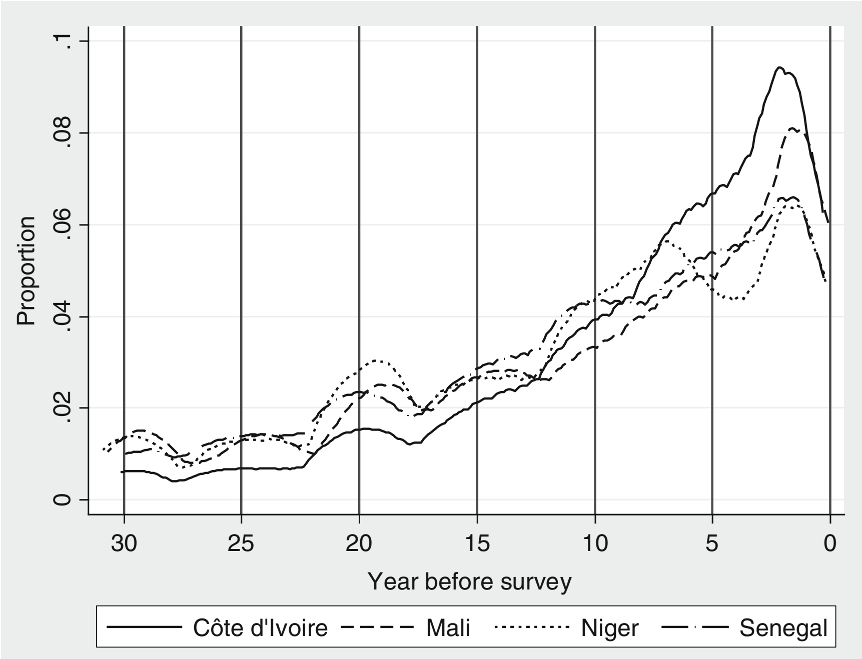

Fig. 10.3 Distribution of last migration by country (Source: REMUAO 1993, weighted samples, both sex aged 15 years and more, our own computation)

Table 10.3 Proportion of migrations by rank and by completed years before survey in Côte d'Ivoire, Mali, Niger and Senegal

Completed years before survey	Last migration	Last but 1 migration	Last but 2 migration	Last but 3 migration	Last but 4 or more migration	Total	Migrations (N)
0	88.98	9.69	1.00	0.34	0.00	100	5937
1	80.75	16.02	2.02	0.73	0.48	100	7705
2	69.74	23.50	4.36	1.37	1.04	100	7052
0–2	*79.36*	*16.75*	*2.52*	*0.84*	*0.53*	*100*	*20,694*
3	63.77	24.61	7.43	2.38	1.81	100	6988
4	54.15	30.49	8.67	3.97	2.73	100	6780
0–5	*71.24*	*21.05*	*4.73*	*1.76*	*1.22*	*100*	*34,462*
5–9	50.53	27.23	10.92	5.54	5.78	100	32,911
10–14	40.12	27.47	13.61	8.16	10.64	100	25,766
15–19	35.07	24.75	14.14	9.48	16.56	100	18,542
20–24	32.14	24.11	13.72	9.90	20.12	100	13,284
25–29	31.36	23.93	14.31	8.91	21.50	100	8811
30+	32.07	24.95	13.00	9.58	20.40	100	18,087
Total	46.66	24.83	11.05	6.67	10.80	100	151,862

Source: REMUAO 1993, weighted samples, both sex aged 15 years and more, our own computation

Another issue is with the minimum duration of residence criterion used to consider an individual either a resident or non-resident. The criterion is necessary to discard short-term visits either in or out of the surveillance area. The 3-month criterion means that an individual has to reside in the surveillance area for at least 3 months to be designated as an in-migrant, while a resident in the system is designated as an out-migrant when away for at least 3 months. To make life easier to fieldworkers, this minimum duration of residence is often chosen to coincide with average time between household visitations, i.e. rounds of data collection. In principle, the precision for data collection should always be higher than the precision for data analysis. For example, if 6 months is the minimum duration of residence used for analysis, then the criteria used in the field should be less than 6 months (e.g. 4 or 3 months). That way, the analysis will not be influenced by variations in the implementation of minimum duration of residence in the field.

Last but not least, reasons for migration are important to collect. They cannot substitute for the analysis of determinants, which will be based on other covariates such as individual and household characteristics or events. Reasons for migration will however help to qualify individual migration. They complement well information on origin and destination. Use of reasons for migration is comparable in many ways to the use of causes of death in mortality analysis. Yet, reliability of reasons for migration is an issue. Contrary to origin or destination, reasons given by respondents are often subjected to recall bias (i.e. when the respondents does not quite remember the circumstances of migration) and to conformity or rationalization bias. The circumstances prior to migration will be redeemed or reinterpreted by the respondent under the light of what happened as a consequence of this migration. For example, if the migration was motivated by economic reasons but failed in that regard, then the respondent may be tempted, consciously or not, to shed a good light on this migration by evoking family or other reasons. Of course, it is impossible to know using retrospective data such as DHS what were the exact circumstances and views of respondents at the time of migration. Ideally, one could ask about intention to move prior to migration on a continuous basis in HDSS, but this has actually never been done.

The Case for the Demographic and Health Survey

Questions on Migration and Residences

Some DHS collect information on previous residence of adult respondents (female aged 15–49, and sometimes males aged 15–54 or 15–59) but never on their children. Residential histories have been collected to improve accuracy of recording other information, notably calendar data on contraceptive use. However, few countries chose to include this module. A total of 28 surveys in 16 countries collected monthly residence histories in a migration calendar: Bolivia (1993–1994), Brazil (1991–1992; 1996), Colombia (1990; 1995; 2000; 2005), Dominican-Republic (1991; 1996), Egypt (1992–1993), Guatemala (1995), Jordan (1997), Kenya (1998), Morocco (1992), Nicaragua (1997–1998), Paraguay (1990), Peru (1991–1992; 1996; 2000; 2004–2006, 2007–2008),[2] Philippines (1993; 1998), Turkey (1993; 1998), Vietnam (1997, 2002), and Zimbabwe (1994). The module was anyway collected only for women who used contraception in the past 5 years, which is an important limitation in most developing countries. Therefore, migration analysis using DHS must rely on information on previous residence collected on adult of reproductive age only, and that for about half the countries that conducted a DHS survey.

Information on the duration of stay in the current place of residence (urban or rural) and on the previous place of residence of the mother is available through two questions asked in the women's questionnaire: "How long have you been living continuously in (name of locality, town or city of current residence)?" and, if the person was not born in place of current residence, "Just before you moved here, did you live in a city, in a town, or in the countryside?" Country of previous residence is never asked, but sometimes the category "abroad" is added to the type of area. Duration of stay is available in years. Place of

[2] From 2004, Peru conducted continuous surveys.

current residence is available in fairly high resolution and its reliability depends on the stratification used for sampling. Place of previous residence is available in large categories, most of the time capital city, other urban (sometimes divided into large city and town), rural and, sometimes, abroad.

Limitations

Perhaps the most important limitation of using information on previous residence has to do with the temporal ordering of migration event vis-à-vis its covariates. The DHS may provide information on previous residence but the respondents' socioeconomic situation in this previous residence is not informed. Most of the socioeconomic variables are not time variant and thus reflect the respondent or household situation at the time of the survey. In particular, time-varying and area-specific indicators of wealth and access to services, as well as household characteristics and composition, would certainly provide better determinants of residential change than current situation indicators. This limitation is not particular to migration analysis and is in fact a serious limitation to any contextual approach of fertility and child mortality too. The principle of anteriority of the cause on the effect is breached when covariates are not strictly referring to the period before the event be it migration, birth, or death.

Another limitation is the number of years that can be reasonably covered by the question on previous residence. This question captures for each respondent last migration only. As mentioned in the previous section, migration matrices will only be reliable under the strong assumption that respondents have only been migrating once over a preferably short reference period. To determine the optimal length of this reference period is not easy without prior knowledge on migration intensity in the country. Because DHS main aim was not to collect information on migration, answers on duration of current residence and on place of previous residence may not be as reliable as in NESMUWA which aimed at collecting information on migration. Using NESMUWA 1993 surveys and DHS done around the same year, we compare rates of last migration by year before survey (Figs. 10.4 and 10.5). Because NESMUWA surveys recorded the calendar year of migration the rates are very much dependent on the month of interview in the survey year (1993). Côte d'Ivoire and Senegal conducted their survey later in the 1993 year than Mali and Niger. Therefore rates for 1992 and 1993 are combined into a rate representing 1 year duration of residence.

DHS offers estimates that are heavily marked by age heaping and seems to overestimate rates for 1 year duration of residence as compared to 2 years duration of residence. In both DHS and NESMUWA sources Côte d'Ivoire has the highest rates, while the three other countries have comparable rates (Table 10.4). Yet, rates are significantly different in the two sources for the 3 or 5 years before the surveys, except for 3-year rates in Niger. DHS produced generally higher rates than NESMUWA, except in Mali where the opposite holds. These differences may be attributed to differences in definitions of residence (any change of residence in NESMUWA surveys; village, town, or city where respondent was interviewed in DHS surveys) and in collecting duration of residence (counted in months in NESMUWA surveys; in years in DHS, which might explain the heaping: see Figs. 10.4 and 10.5), but also in sampling as mentioned before.

Analysis of international migration is limited to in-migration flows (provided that information on previous foreign residence is available) since no emigration questionnaire is available in DHS. International migration flows could in principle be measured through surveys in destination countries by computing the numerator using destination countries and the denominator in origin country. However, pre-coded responses to the question on previous residence in DHS do not include the country of previous residence. Therefore, recent international migration flows and determinants can only be analyzed using DHS for foreign origin as a whole and not by country of origin.

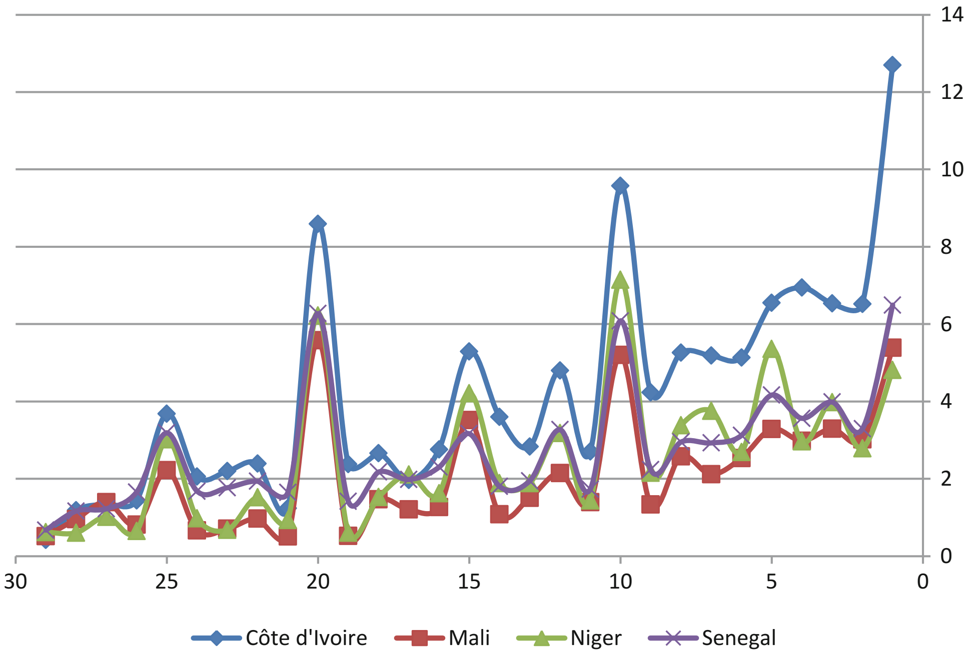

Fig. 10.4 Hazard rates of last migration (in %) by country (Source: DHS weighted samples, female aged 15–49, our own computation)

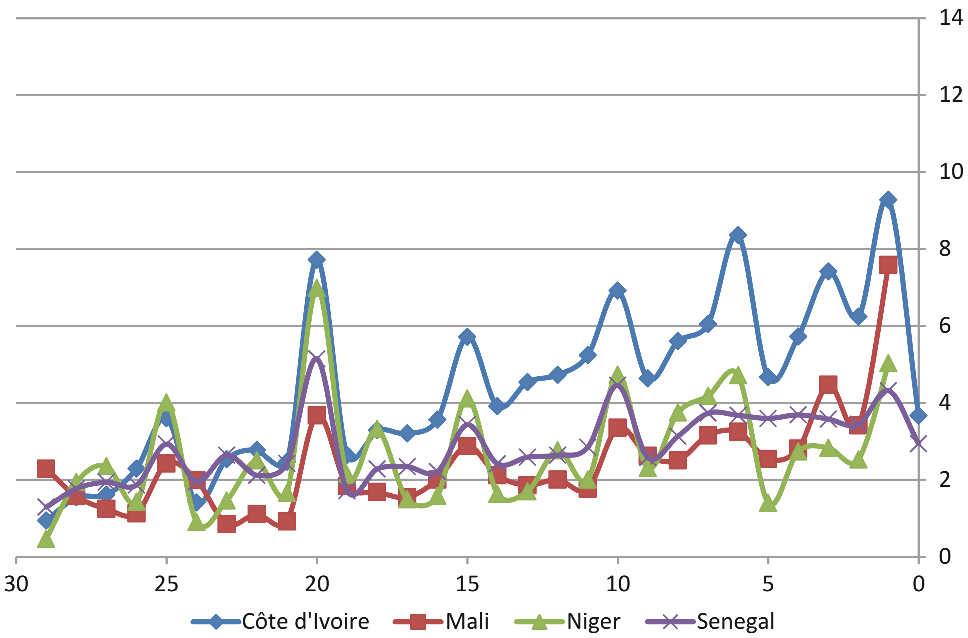

Fig. 10.5 Hazard rates of last migration (in %) by country (Source: REMUAO weighted samples, female aged 15–49, our own computation)

Table 10.4 Rates of last migration (in %) in the 3 and 5 years preceding survey by country

Country	Survey (date)	Rate of last migration 3 years before survey [95 % CI]	Rate of last migration 5 years before survey [95 % CI]
Côte d'Ivoire	REMUAO 1993	**6.31**	**6.41**
		[5.93–6.72]	**[6.10–6.75]**
	DHS 1994	**8.65**	**7.98**
		[8.24–9.08]	**[7.66–8.31]**
Mali	REMUAO 1993	*5.20*	*4.25*
		[4.87–5.56]	*[4.01–4.50]*
	DHS 1995–1996	*3.67*	*3.46*
		[3.44–3.92]	*[3.28–3.65]*
Niger	REMUAO 1993	3.49	**2.95**
		[3.18–3.84]	**[2.72–3.21]**
	DHS 1992	3.77	**3.92**
		[3.49–4.09]	**[3.68–4.17]**
Senegal	REMUAO 1993	**3.56**	**3.58**
		[3.33–3.80]	**[3.40–3.78]**
	DHS 1992–1993	**4.50**	**4.26**
		[4.19–4.83]	**[4.03–4.51]**

Source: DHS and REMUAO weighted samples, female aged 15–49, our own computation

Existing and Potential Analyzes

Migration analysis using DHS surveys has rarely been conducted. There is certainly here an opportunity for conducting a systematic cross-country, multi-year analysis of migration flows and migration determinants, for the 3 years preceding the survey. Migration flows by region *and* area (urban-rural) of residence might not be easily estimated because of sampling constraints (size and stratification), but migration by area of residence is possible using most DHS surveys. At the very least the urban-rural divide will be available while in most DHS surveys urban hierarchy also is.

The analysis of the determinants of migration may be done using fixed covariates, such as the basic age and sex variables, or others like ethnicity, and obviously place of residence. This would already a great progress in migration analysis as this type of analysis has actually never been done so far on a systematic basis for each DHS survey for which migration variables are available. Of course, international migration analysis will be limited to immigration since information on emigrants is not available.

In-migration rates and determinants analysis can be performed by reversing analysis time. I describe here the method in much the same terms as Béguy et al. (2010). It consists in running the time of analysis in reverse (Baydar and White 1988). This produces tables that will formally have the same structure and properties as decrement tables but which will be interpreted as increment tables. Analyzes in the preceding section are actually produced using this method. It allows conducting descriptive and multivariate analysis on the determinants of in-migration in the same way as for out-migration. For out-migration analysis, the starting time of analysis is often birth or any specific age (often 15 or 18) at which the migrant is likely to migrate of his/her own volition. For in-migration analysis, a specific age can also be used, but reversing time will prevent us from using date of death (the mirror equivalent of date of birth) as a starting time of analysis. This is because death occurs at very different ages and therefore would introduce unnecessary age heterogeneity in the analysis time. In addition, death is in most cases not independent from migration behavior. For these reasons, in order to produce age-specific in-migration rates, the analyst has to choose the age for starting the time in reverse, depending on the size of the population at risk at older ages and the scope of the analysis. In the DHS case, one has to choose the upper age limit below that chosen for the sample selection (45 or lower instead of 50 for females, 55 or lower instead of 55 or 59 for males).

As mentioned earlier, analysis is limited by the unavailability of household characteristics and composition and other contextual variables at the time of migration. Education level might be used upon some hypothesis about the correspondence between age and level of education. For example, if completed primary education is the respondent's level of education declared at time of survey, then a reasonable assumption is to consider the respondent as being at school from age 6 to 12, if this is the age when primary education ends in the country, and attribute primary level to the respondent from age 12 onward. However, other individual time-varying characteristics, such as occupation, cannot be deduced from characteristics at the time of the survey.

Considering their limitations for migration analysis, most DHS users may prefer to use migration not as the dependent variable but rather as a determinant in fertility or child mortality analyses. It seems that migration has not been so often used as a determinant so far essentially because of technical constraints. Many early uses of migration in fertility or mortality were limited to migrant status in the destination area. Brockerhoff pioneered the comparison of migrants (differentiated by their origin) and non-migrants as regard to their fertility (Brockerhoff and Yang 1994) and mortality of their children (Brockerhoff 1994, 1995). Migrants were considered from the time they arrived at current place of residence and period in previous residence was discarded from analysis. In some other studies (e.g. Ssengonzi et al. 2002; Van de Poel et al. 2007) migration status is used directly as a determinant, without actually controlling for the duration in the current place of residence. In some others, the beginning of exposure time is compared with the time of migration to identify categories of migration exposure (e.g. Omariba and Boyle 2010). In yet other papers (e.g. Chattopadhyay et al. 2006), comparison is made between lifetime behavior and behavior prior to migration. However, these studies do not directly use migration as a time-varying covariate.

Now that time-varying covariates and left-censoring are easily handled with available statistical software, the entire risk period can be analyzed. The paper by Bocquier et al. (2011) on child mortality by area of residence is an example of such event history analysis on DHS data using the migration event as a time-varying covariate. The urban-rural differential was the main interest of this paper which concentrated its attention on urban-rural migration effects by comparing under-5 mortality before and after migration. However, urban-to-urban, rural-to-rural or more complex migration stream effects could as well be analyzed although not in all countries due to sampling issues (sample size and stratification: see above).

The limitations of migration as a determinant will be very similar to those listed above regarding migration analysis per se. The period of reference should be chosen with care and most covariates are not reliable to contextualize the period before migration. To note, these limitations are not particular to the migration variable. However imperfect the indicator may be, the use of migration as a determinant should be encouraged even if the focus is not on migration. Capturing changes in place of residence and the heterogeneity of respondents as regard to residency can only improve the quality of the demographic analysis. Neglecting this important heterogeneity in the life of respondents may actually bias effects of other determinants.

The Case for Health and Demographic Surveillance Systems

As much as a survey sample is meant to represent a (usually) national population, a health and demographic surveillance system (HDSS) is a "*geographically-defined population under continuous demographic monitoring with timely production of data on all births, deaths and migrations*" (INDEPTH founding documents, 1998, http://www.indepth-network.org/).

Originally implemented to collect demographic data on the catchment area of hospitals or dispensaries to complement public health data on specific diseases, HDSS objectives now go beyond finding the right denominator to computing epidemiological rates. In addition to providing information on the burden of diseases in the population, they serve as early alert systems and platforms to implement and evaluate the impact of health interventions, sometimes using randomization of cases and controls. Besides the continuous and exhaustive monitoring of vital and migration events, HDSS serve as sampling frames to draw samples for a range of health, economic, social and behavioral nested studies. Data collection rounds take place annually or even more frequently. Several years of continuous data collection are necessary for return to investment. HDSS are notoriously expensive to implement but are also irreplaceable tools for epidemiological, demographic and socio-economic studies.

HDSS are not valued for their representativeness but rather for their ability to generate reliable, longitudinal, community-based and well-contextualized health data. Here exemplarity takes over representativity. In a survey, the sampled population is drawn from a universe, which is the total targeted population. Each unit in this population is interchangeable and randomness of the draw ensures that sampled units taken as a whole represent the universe well. Confidence intervals are computed using simple and reasonable laws from the Gaussian family. These confidence intervals account for both sampling errors (associated with sample size, stratification, clustering, etc.) and data collection errors (due to respondents, interviewers, data entry clerks, etc.) as long as they are random, i.e. unbiased. In HDSS framework, the population of a geographically limited area is interviewed but is not meant to represent the whole population of a country. On the contrary, an HDSS is considered exemplar or illustrative of a particular, sometimes marginal situation, monitored through a careful examination of contextual, environmental and community-level information. HDSS are usually situated in deprived rural, semi-urban or urban areas. A given HDSS population is then considered as a unique draw from a hypothetical universe of all possible similarly deprived situations. The fact that the population follow-up is exhaustive does not mean that there is no random component in this population. Sampling errors are absent and random data collection errors are supposed to be reduced to a minimum through regular waves of data collection and complex consistency checks, but randomness may occur from behaviors themselves however close-to-perfect data collection may be. Therefore, confidence intervals are still needed though computation techniques may differ from sample analysis. Resampling methods (bootstrap, jackknife…) will be preferred to Gaussian-based methods.

Rather than aiming at representing the behavior of the whole population, HDSS aim at identifying causal relationships in sequences of events in great details, including for rare events (e.g. maternal death, neglected diseases) that hold in similar contexts. The incidence of an event is of higher interest in HDSS than the prevalence of this event. Also, the causal relationships between events at community, household and individual levels are of higher interest in HDSS than the precise description of each event at a given time. In other words, HDSS analysis seeks at generalizing processes rather than states. Moreover idiosyncrasy inherent to HDSS is compensated by comparative analyses of HDSS data, which is encouraged by a unique (in all acceptance of the term) network – the International Network for the continuous Demographic Evaluation of Populations and Their Health in developing countries (INDEPTH). Triangulation with administrative, hospital, census and survey data may also help generalization of HDSS results.

Migration Registration System and Their Limitations

As regard to migration, HDSS offer exceptionally rich data on residential history albeit limited to small populations. In principle, all residential moves (changes of household) within the surveillance area are captured in addition to moves in and out of the surveillance area. The time criterion varies among HDSS from 1 month to 1 year. It is advisable to use less than 6-month criterion in the field, since 6-month duration tends to be the standard duration for residence in migration analysis.

At each round of data collection *n*, information on these moves is collected retrospectively covering the time since the last round *n-1*. Migration status of in- and out-migrants is confirmed at the following round $n + 1$. This is because of the so-called "hanging cases", i.e. when a person has been declared a migrant on the basis of intentions to leave or stay in the household but has not yet completed the minimum duration in or out of the household to be considered a true migrant. For this reason the data covered by the last round are usually discarded from analysis. Also, the migration status at the onset of surveillance is not often well-known for the whole population. Data covering events from initial census (i.e. round 0) to round 1 or even round 2 are usually not reliable and discarded from analysis. This is because migration is essential to define the population at risk. Unreliable residency status of individuals in the system creates a serious bias in the computation of the population at risk leading to the over- or underestimation of all demographic rates.

Another challenge can be "the reliable re-identification of an individual on in-migration as being the same person who previously moved out" in the same HDSS (Sankoh and Byass 2012). Within-HDSS moves were often ignored in identification system, leading to the attribution of two different identifiers for the same individual moving from one household to another. This double identification is not an issue as regard to computing the population at risk (there is no double count for the same time period) but this leads to several imprecisions in the analysis. Within-HDSS moves may be confused with migration in and out of the HDSS. The continuity in biographical record is artificially broken, leading to a loss of information. Extra time will be necessary to record information that was already asked to the respondent at former place of residence, leading to interviewer's and respondent's fatigue. Fortunately, procedures are now put in place among cooperating INDEPTH sites to avoid this type of double identification by asking precise questions aimed at reconciling identifiers.

Identification and understanding of migration processes are essential to both management and analysis of the whole data. Once precautions taken on the quality of information HDSS offer a unique tool to compute the complete basic demographic equation (mortality, fertility and migration rates) in countries where vital registration is lacking or deficient. HDSS can routinely produce a core minimum longitudinal micro-dataset containing all vital events for each individual under surveillance. This is sufficient to compute precise exposure and gross demographic rates as well as more complex statistics such as life tables, life expectancy, age-specific fertility rates, migratory and natural rate of increase, etc. Expanding this core longitudinal dataset with event attributes, other status event and individual, household and community characteristics enables more complex event history analysis.

Embedded in the INDEPTH network, the Multi-local Analysis of the Dynamics of Internal Migration And Health (MADIMAH) initiative follows this event history analysis (EHA) perspective and aimed at improving capacity of HDSS to produce the required datasets, i.e. at promoting EHA-oriented data management. This initiative contributed to the production of core longitudinal micro datasets containing all vital and migratory events (available for free-of-charge download through the iShare micro data repository platform, http://www.indepth-ishare.org/), as well as to the production of all mortality, fertility and migration indicators (available for display and for download through the INDEPTHStats aggregated data platform, www.indepth-ishare.org/indepthstats/).

However important is the monitoring of migration in HDSS, it must be acknowledged that migration has been rather neglected in the analysis of HDSS data. Despite the value of exhaustive recording of migration events over the surveillance period, HDSS usually do not involve the collection of migration histories from birth to first enumeration (for those born before the surveillance started) or from birth to first in-migration. This has no implication on the computation of demographic rates over the surveillance period but it is a serious and often overlooked limitation as it prevents using migration history as a predictor of future migration. To note, the same problem arises in HDSS that do not record reproductive history, or union formation. Analysis of reproductive health determinants is then limited.

Another important limitation is to do with the time of data collection on contextual factors. Even though they are longitudinal data collection systems, HDSS are not free of the issue of temporal order of migration event vis-à-vis its covariates. In much the same way as for cross-sectional surveys, household or community (e.g. village) characteristics are often collected once every *X* years, the assumption being that these characteristics do not change much over time. However plausible this assumption is, collecting this kind of information every *X* years creates a discontinuity in the otherwise longitudinal nature of individual-level data. Dates of changes in household or community characteristics are not captured. Therefore, these changes often cannot be situated before or after demographic and other events in causal analysis thus limiting causal analysis. Assumption on the date of these changes can be made (e.g. by setting changes at the mid-period between two consecutive data collection on household or community characteristics) but this approximation is detrimental to the precision of the analysis.

What about geocoding and spatial analysis of HDSS? As Sankoh and Byass note (2012), "the technological and methodological possibilities for obtaining and using geographical data have advanced considerably, to the point where recording the latitude and longitude of every residential unit, and other salient features, in an HDSS using global positioning system (GPS) technology have become commonplace." Geocoding may not contribute as much to the analysis of migration in HDSS as it would at national level, since the geocoding pertains more to the determinants of migration than to migration events themselves. It certainly helps to get better precision as to the conditions prevailing in households under surveillance, but it does not help characterize places of destination or origin outside the HDSS. In other words, the type of external migration will not be better identified by more precise geocoding within the surveillance area. Geocoding of households and amenities contributes mainly to multilevel analysis since it allows the definition of more precise geographical layers and enhance the possibility to relate household data to some macro characteristics such as rainfall, temperature, etc.

Questions on the Circumstances of Migration

Because HDSS were not initially meant to study migration, information on origin and destination and reasons for migration is rarely collected from the onset of the surveillance, or not in a systematic way. Migration is still considered by many analysts as independent censoring, i.e. as attrition (loss-to-follow-up) or right-censoring in the case of out-migration, or its opposite, left-censoring in the case of in-migration. However, for migration to be considered as a mere censoring event one has to make the strong assumption that it is independent from the event at stake be it death, birth or any other event for that matter. This is contrary to what all migration analyzes show: migration is not random but often motivated by health, economic or social reasons. In sum, it must be acknowledged that migration is a major source of non-independent censoring, an issue that has not been seriously tackled so far in demography. Most analysts blithely consider attrition by migration as independent censoring in standard descriptive and analytical models, i.e. as a nuisance that reduces the population at risk in cohort studies.

To make progress in this matter, it is very important therefore to understand the circumstances of migration better, even to analyze other events than migration. Information on origin and destination and reasons for migration is obviously crucial to analyze migration determinants per se, but it also helps identifying possible selection effects as regard to health and socio-economic events. The few studies on migration-health interactions reviewed in the following section show that behavioral variations by migration status are very high. These interactions have huge consequences on the measurement of demographic rates and, if not taken into account, they may bias the estimation of the effects of other determinants.

When information is collected on migrants, it is usually in the form a question on origin and destination if not on reasons for migration. Because HDSS are limited to small geographical areas, the responses (whether pre-coded or not) involve a hierarchy by geographical distance or by importance of the agglomeration at destination or origin. This hierarchy is particularly important to distinguish between migrants who cross HDSS site boundaries depending on whether they make a close or a distant move. Some migrants may just change residence to neighboring villages or city blocks that are more likely to share characteristics with their place of origin.

Sometimes international destination and origin are coded, which makes international migration analysis possible. It should however be noted that at a HDSS (small) geographical scale, international migration is in most cases too specific to the sites or too small in relative and absolute terms to be relevant at a larger (national) scale. International migration status, when there has been a massive immigration flow due to particular circumstances, may be used for analyzing social, economic or health integration as was done for example in Agincourt HDSS (South Africa) where there had been in the 1980s a major refugee flow from neighboring Mozambique.

Most of the above is relevant to take account of selectivity by in-migration. Selectivity by migration out of the HDSS is usually ignored though it may create a high bias, since out-migration can be regarded as informative censoring as mentioned above. Whereas questions may be asked to in-migrant respondents on their place of origin and the circumstances of their migration, little is known about the destination and circumstances of out-migrations. A great improvement to the analysis of migration determinants could be derived from more precise follow-up of migrants out of the HDSS. Follow-up could help in qualifying out-migration better by identifying out-migration determinants. Retrospective and prospective follow-up may be used to correct for migration bias in the analysis of population behavior. For example, case-control design can be implemented by comparing out-migrants with matched cases of non-migrants in the HDSS. This would help identifying how migration and other behaviors interact in time. Statistical modelling is a strategy to control for selection biases but will never beat hard data on household and migrants' behavior pre- and post-migration. Because mobile phones are now widely and cheaply available throughout the globe, including deprived rural areas, it would be possible to ask HDSS residents their phone number and that of their next of kin for further contact in case of migration. Conditional on respondent's approval, phone interview could be organized in the months following out-migration. This would allow collecting directly from the migrant information on migration circumstances, current living conditions and precise location of destination.

Existing and Potential Analyzes

The first initiative regarding analysis of migration determinants and consequences using HDSS data is found in the book "The Dynamics of Migration, Health and Livelihoods – INDEPTH Network perspectives" (Collinson et al. 2009). Summary of findings is available in Gerritsen et al. (2013). All seven participating sites in Africa and Asia showed a relatively regular age structure for migration favoring young adults (aged 20–24) most of them motivated by employment, but also by union formation or dissolution, and sometimes accompanied by their young children. Retirement and access to better health services and care are also motivations at older ages. Return migration is more frequent for males.

A careful analysis of migration flows in Nairobi slums shows an annual turn-over of a quarter of the HDSS population and of a third of those aged 15–30 (Béguy et al. 2010). The circular migration system at play is becoming more intense for women than for men, explaining the long-term decline in male-to-female ratio. Analysis of in- and out-migration determinants show that the high population turn-over in slums is associated with insecurity of livelihoods and tenure, as well as with poor basic amenities and social services. The selection process by which migrants stay or leave urban and rural areas still need to be investigated by closer examination of objective and subjective determinants of migration.

Migration is often associated through remittances in money or in kind with educational improvement of children left at home (rural Bangladesh) or in socio-economic status of the household left behind at large (rural South Africa). How loss of labor employed in agriculture is compensated by remittances depends on the household land resources (rural Thailand).

The burden of child morbidity is higher when the migrating parent is the mother (rural Vietnam). Children born of newly resident mothers have higher mortality risks than those born of long-term migrants in urban slums in Kenya, but the opposite holds for returning migrant mothers from urban to rural areas in Kenya. More detailed analysis comparing children born in and out of the slums (Bocquier et al. 2011) showed that the slum-born have higher mortality than non slum-born, indicating long-term health consequences of delivering in the slums. Also, children born in Nairobi slums to women who were pregnant at the time of migration have the highest risk of dying.

For adults, mortality is higher for returning migrants, essentially because of AIDS/TB (rural Mozambique) thus confirming the “returning home to die” phenomenon observed in another HDSS in rural South Africa (Clark et al. 2007).

To sum up, the findings in the HDSS-based studies highlight the potential negative consequences of migration on health which contrast with the beneficial impacts of migration on livelihoods. What poor populations may economically gain from migration on one hand, they may lose in health on the other hand. The MADIMAH initiative, through analysis of strictly methods and comparable data gathered on a dozen of HDSS, seeks at confirming the direction of the relationships between migration, livelihood and health. A number of other issues are worth analyzing in relation to migration, both as a consequence and as a determinant: reproductive health and fertility, chronic diseases, aging, union formation, etc.

Conclusions

Demographic surveys and surveillance systems can be used for the analysis of migration both as a dependent and as an independent variable. However, a number of limitations have to be borne in mind before conducting both types of analysis. When dealing with last migration as in most surveys, it is important to limit the period of analysis to 3 years before the survey and to limit place of origin to large geographical areas in order to avoid biases. Analyses of interactions between migration and another event should check for the order of these events. When migration is a determinant, it should be a time-varying covariate, which can only improve quality of analysis. Also, information on origin and destination and reasons for migration are important for analysis, since migration is not a random event and is often motivated by health, economic and social issues. Migration is a major source of informative censoring, i.e. not independent from the other events of interest. Analysis should not eliminate migration, migration should rather illuminate analysis. Information prior to migration and follow-up after migration are important improvements that should be encouraged in existing demographic survey programs and surveillance systems.

References

Baydar, N., & White, M. (1988). A method for analyzing backward recurrence time data on residential mobility. *Sociological Methodology, 18*, 105–135.

Béguy, D., Bocquier, P., & Zulu, E. (2010). Circular migration patterns and determinants in Nairobi slum settlements. *Demographic Research, 23*, 549–586.

Bilsborrow, R. E., Oberai, A. S., & Standing, G. (1984). *Migration surveys in low-income countries: Guidelines for survey and questionnaire design*. London: Croom Helm.

Bocquier, P. (2004). Analyzing urbanization in Sub-Saharan Africa. In G. Hugo & A. Champion (Eds.), *New forms of urbanisation*. Aldershot: Ashgate, IUSSP Group on Urbanisation.

Bocquier, P., & Traoré, S. (1998). Migration and urbanization in West Africa: Methodological issues in data collection and inference. In R. E. Bilsborrow (Ed.), *Migration, urbanization, and development: New directions and issues*. Norwell: United Nations Population Fund (UNFPA) and Kluwer Academic Publishers.

Bocquier, P., Beguy, D., Zulu, E., Muindi, K., Konseiga, A., & Yé, Y. (2011). Do migrant children face greater health hazards in slum settlements? Evidence from Nairobi, Kenya. *Journal of Urban Health, 88*, 266–281.

Brockerhoff, M. (1994). The impact of rural-urban migration on child survival. *Health Transition Review, 4*, 127–149.

Brockerhoff, M. (1995). Survival in big cities: The disadvantages of migrants. *Social Science and Medicine, 40*, 1371–1383.

Brockerhoff, M., & Yang, X. (1994). Impact of migration on fertility in Sub-Saharan Africa. *Social Biology, 41*, 19–43.

Chattopadhyay, A., White, M. J., & Debpuur, C. (2006). Migrant fertility in Ghana: Selection versus adaptation and disruption as causal mechanisms. *Population Studies, 60*, 189–203.

Clark, S. J., Collinson, M. A., Kahn, K., Drullinger, K., & Tollman, S. M. (2007). Returning home to die: Circular labour migration and mortality in South Africa. *Scandinavian Journal of Public Health, 69*, 35–44.

Collinson, M., Adazu, K., White, M., & Findley, S. (Eds.). (2009). *The dynamics of migration, health and livelihoods – Indepth network perspectives*. Farnham: Ashgate.

Courgeau, D. (1988). *Méthodes de mesure de la mobilité spatiale – Migrations internes, mobilité temporaire, navettes*. Paris: Editions de l'INED.

Gerritsen, A., Bocquier, P., White, M., Mbacké, C., Alam, N., Beguy, D., Laserson, K., Nhacolo, A., Phuc, H. D., Punpuing, S., & Collinson, M. (2013). Health and demographic surveillance systems: Contributing to an understanding of the dynamics in migration and health. *Global Health Action, 6*, 21496.

Omariba, D. W., & Boyle, M. (2010). Rural–urban migration and cross-national variation in infant mortality in less developed countries. *Population Research and Policy Review, 29*, 275–296.

Sankoh, O., & Byass, P. (2012). The INDEPTH network: Filling some international gaps in epidemiology. *International Journal of Epidemiology, 41*, 579–588.

Ssengonzi, R., De Jong, G. F., & Stokes, S. (2002). The effect of female migration on infant and child survival in Uganda. *Population Research and Policy Review, 2* (21), 403–431.

Van De Poel, E., O'Donnell, O., & Van Doorslaer, E. (2007). Are urban children really healthier? Evidence from 47 developing countries. *Social Science & Medicine, 65*, 1986–2003.