The programme of regionally coordinated

Household International Migration Surveys in the Mediterranean Countries

MED-HIMS PROGRAMME

November 2013

MANUAL 5

SAMPLING PLANS FOR MED-HIMS SURVEYS

MED-HIMS

Mediterranean Household International Migration Survey

Manual 5

Sampling Plans

for MED-HIMS Surveys

November 2013

Foreword

The 'Mediterranean Household International Migration Survey' (MED-HIMS) is a regional, coordinated programme of household surveys developed for the countries of the Southern and Eastern Mediterranean region. The programme is designed to overcome the lack of data on international migration for the region by collecting reliable and representative multi-topic, multi-level, retrospective and comparative data on the characteristics and behaviour of migrants and the determinants and consequences of international migration and mobility. The programme has its origin in the European Commission's MEDSTAT Programme, and since its initial inception in 2009, has gone through a wide preparatory and consultative process with support from the European Commission, the World Bank, the UNHCR and the UNFPA and later on the other partner organisations.

The main objectives of the MED-HIMS Programme are:

(i)   to study the recent trends, causes, determinants, dynamics and consequences of international migration and mobility, and the inter-linkages between migration and development; and

(ii)   to explore scenarios for a closer cooperation in the area of migration and development between the sending countries in the Southern Mediterranean region and receiving countries, particularly the European Union.

The objectives and design of the MED-HIMS are guided by the vision of the 2004 Marrakech Action Plan for Statistics (MAPS), the 2009 Dakar Declaration on the Development of Statistics (DDDS), the 2011 EC Communication on the Global Approach to Migration and Mobility (GAMM), and the various strategies and recommendations of the United Nations Global Forum on Migration and Development (GFMD).

The MED-HIMS methodology is designed to deal with the various dimensions of international migration and mobility by carrying out specialized national household surveys in the sending countries that aim to capture current and recent developments in the Southern Mediterranean region.

In order to maximise the quality, utility and comparability of the data collected through the MED-HIMS surveys, a set of Model Questionnaires has been developed by the international organisations together with the countries of the region. These questionnaires are designed to collect representative data on out-migration, return migration, forced migration, intention to migrate, circular migration, migration of highly-skilled persons, irregular migration, type and use of remittances, behaviours, attitudes, perceptions and cultural values of people with regard to international migration and mobility, and the inter-linkages between migration and development, as well as relevant information on the individuals, households and local communities involved

In addition to the Model Questionnaires, a series of manuals, guidelines and computer systems, covering the different phases of the survey from the initial organisation to

tabulations, data analysis and reporting results, has been conceived and already partially developed to provide countries with guidance on the design and implementation at the national level of the MED-HIMS survey or equivalent operations, whether under internationally coordinated programmes or national stand-alone programmes, within or outside the Southern and Eastern Mediterranean region.

The set of manuals and guidelines comprises the following:

- Manual 1: MED-HIMS Model Questionnaires

- Manual 2: Survey Design and Organization

- Manual 3: Instructions to Supervisors

- Manual 4: Instructions to Interviewers

- Manual 5: Guidelines for Sample Design

- Manual 6: Data Dictionary and Recode Specifications

- Manual 7: Guidelines for Country Report

Given the greatly varying availability of sources of information for sampling encountered in the first three participating countries and the timetable of the different national surveys, this current version simply presents the case studies of those first three countries Egypt, Jordan and Morocco that have undertaken this phase of the survey design, with some introductory comments on differences compared to a theoretically ideal approach.

The sampling reports presented here are the results of the technical assistance missions undertaken by Richard Bilsborrow (University of North Caroline, MED-HIMS Sampling Expert) in the framework of MEDSTAT III to Central Agency for Public Mobilization and Statistics of Egypt, (15-28 September 2012), the Department of Statistics of Jordan (28 September 28 - 11 October 2012) and the Haut Commissariat au Plan of Morocco (25-29 March 2013).

Content

Introduction

1. Introduction to the introduction

This is a brief introduction to the three first sample plans developed under MEDSTAT III for the MEDHIMS project to design, implement, and analyze household surveys on international migration in predominantly Arab-speaking countries in the Mediterranean region. This project had grown out of an urgent need to collect significantly better data on international migration and its linkages with socio-economic development in the region, a need recognized by governments in the region as well as international agencies, as described elsewhere.

To collect this data, it is necessary to have a plan for identifying households to interview in the participating countries. This is the purpose of a sample, which needs to be nationally representative, cost-efficient, and draw on existing data sources, for better or worse, in each country. Before briefly reviewing and assessing the experiences to date in the first three countries (of the anticipated eight total), it is useful to indicate what the usual desirable approach is, which begins with a recognition that international migrants of interest for this project (in all countries, this refers to emigrants, return migrants, and forced migrants, and not to immigrants, though a few of the participating countries also have significant numbers of immigrants) are both (a) "rare elements" in the statistical sense, and (b) not distributed randomly in a population.

Therefore, the next section of this introduction briefly describes the "ideal" approach based on principles of sampling rare elements (Kish, 1965). This is followed by a short summary of the very different actual sampling designs developed for real world applications in Egypt, Jordan and Morocco, and a short concluding section, before the three country sampling approaches are described in some detail.

2. The theoretically Ideal approach—sampling rare elements

As described in detail in Bilsborrow et al. (1997), there are basically two key steps in designing a sample of households to investigate international migration as "rare elements": (a) stratify areas of the country according to the prevalence of (households with) international migrants and oversample areas with higher prevalence using stratified sampling; and (b) in the last-stage small area units, conduct two-phase sampling, which involves, first, listing all households in the sample areas selected to classify them as migrant and non-migrant households, then oversampling households with international migrants; and second, conducting interviews in the sample of migrant and non-migrant households selected. Steps (a) and (b) are further elaborated below.

Regarding (a), the first and fundamental requisite is a source of data that identifies households with international migrants of interest and others (non-migrant households). This requires that there have been a previous census, large nationally representative household

survey, or a universal continuous population register that keeps track of all persons in the country and their location (as in most countries of the European Union). This data source, to be useful for present purposes, should have obtained data on either households--whether households have an international migrant (including a former member who left to live abroad)--or individuals, whether each person is an international migrant. As very few developing countries have high quality population registers that cover the whole population, the ideal source of data is a (recent) national census of population, provided it have included questions on, e.g., whether a previous member of the household has moved to live abroad in the last X years and not returned (to identify households with emigrants), and on whether some current member had lived abroad but returned (return migrant household).

If the census is able to obtain more-or-less complete and accurate data from these questions, tabulations may be prepared based on these questions, to classify households (or individuals) as migrant households and non-migrant households. With this data, the prevalence (or proportion) of migrant households can be calculated for all administrative-political divisions and sub-divisions of the country. All administrative units can then be categorized or grouped into strata, such as high, medium and low. The high stratum will include areas which had high proportions of households with one or more migrants of interest (such as an emigrant), medium proportions, and low proportions (the vast majority, with zero or near zero). Then stratified sampling is used to select (sample) areas from each of the three categories or strata, selecting a much higher *proportion* of areas from the high stratum compared to the medium one, and a much lower proportion of areas from the low stratum compared to the medium one. The result of this stratified sampling is that fieldworkers will concentrate their time interviewing households in areas expected to have more migrant households, and will spend much less time searching for them than in a random sample.

However, once the sample areas (Primary Sampling Units, etc.) are selected using disproportionate sampling or oversampling of areas with higher proportions of international migrants, just going up to randomly selected households in the sample areas selected does not ensure that the field work will not spend most of its effort interviewing non-migrants. Hence it is also necessary to undertake two-phase sampling in the last, smallest stage area units, commonly called census sectors or census enumeration areas (EAs). Once these are selected, a preliminary field operation involving quick visits to all households in each sample EA is needed, to ask just a few questions to identify if it is indeed a residential household and if it has had any prior household member leave to live abroad and remain abroad, or leave and subsequently return. Interviewers spend only minutes asking any adult member or even a neighbor in a single quick visit for this basic data, and record the data, one line per household, on lists. From these lists for each EA, households with migrants are oversampled, but some few non-migrant households are also selected. Then in phase 2, the sample of migrant and non-migrant households is interviewed. Since at all stages, higher proportions of migrants (areas, then households at the EA-level stage) are selected than non-migrants, the sample is not self-weighting, so careful records need to be kept for all EA on the numbers of households of each type found, the number selected into the sample, and the number actually successfully interviewed.

3. Brief comparison and contrast of approaches in Egypt, Jordan and Morocco

The approach of section 2 above is the ideal. In reality, shortcomings are common, both in the region and globally. The first is that there may be no (recent) census at all, or that the census did not include the right questions, or that even if it did, they were not administered well by census enumerators or responded to honestly by respondents. In the case of the three countries, a fairly recent population census was available (2004-2006), and reasonable questions were used to obtain data on at least one of the two main types of migrant households (return migrants, or emigrants); but in each case, there were serious deficiencies in the administration and coverage of the questions, and/or respondents often did not want to reveal the household had an international migrant. The result is that the censuses all had serious under reporting of international migrants. The other main source of data in each country was some kind of household survey, but all three were too small to capture large enough numbers of households with one or more *recent* international migrants.

Thus the two sources of data in Egypt were the 2006 population census and the National Labour Market Panel Survey (NLMPS). Despite all efforts, underreporting of emigrants in Egypt in the census was both extremely high and concentrated in particular areas, especially the large cities. As a consequence, it was decided to *not* follow the recommended approach in (2) above but instead select census enumeration areas as the PSUs distributed across the country in proportion to the population by taking advantage of a recently updated National Master Sample. However, in each EA, although there would be no two-phase sampling in sample EA, still households with migrants would be oversampled compared to non-migrant households through a field procedure of interviewing any households (from the usual list of 84-88) with a migrant, but only a few of the many households with no migrant.

In the case of Jordan, the two main sources were the 2004 population census and the 2012-13 Job Creation Survey (JCS), based on a large, nearly nationally representative sample of sub-districts. Each was thought to have significant under reporting of households with emigrants and/or return migrants. But an experiment was carried out, to compute the (relative) prevalence of migrants across sub-districts and compare the results, to see if, despite their deficiencies, they tended to indicate the same sub-districts as having high, medium, and low prevalence rates. Therefore, both sources together were drawn upon to create three strata for the 89 sub-districts in Jordan, and stratified sampling was used to select 30 sub-districts. Based on the 2004 population census, tabulations of prevalence rates of both emigrants and return migrants for the 30 sampled sub-districts were prepared and used to oversample localities, and then to oversample EAs within localities.

Finally, for Morocco, two sources of data were also examined, the 2004 population census and the longitudinal survey of households (ENDPR) involving three visits over a 12-month period in 2009-2010. Tabulations were prepared from each on available data for households with emigrants and households with return migrants (far more of the latter were found), Since Morocco has already decided to postpone the Morocco-HIMS to after the 2014 census, that made it possible to evaluate the census questions used in 2004 and make recommendations for minor improvements, in order to collect data with less underreporting of migrants. Then a multi-stage sampling design based on the ideal principles of (2) above

was outlined for the planned 2015 Morocco-HIMS, which should be conducted as soon as possible after the census (months, not years).

4. Conclusions

Each country is different in many important ways apart from its sources and quality of data, types and predominant levels of international migrants, and a host of cultural, infrastructural, political-administrative, and developmental factors. Major lessons can be learned from the implementation of each sample. Thus following the survey implementation, it can be checked in each country whether the sampling approach worked or not. In Egypt, for example, the proportions of migrant and non-migrant households found in EAs in different areas could be compared with those in 2004 to see if developing strata based on the 2004 data (with adjustment for the know urban extreme undercount) could have been useful, or not. In the case of Jordan, was the actual (relative) prevalence of households with migrants found in the new Jordan-MIMS close to that anticipated—i.e., was the stratification useful? And for Morocco, was it worth waiting until a new census was carried out, with (even) better census questions and training of enumerators regarding the migration questions? In each case, there should be an ex post assessment of the sampling approach-- how well did the procedures work, or not work, and why? Were problems due to the data available for stratification, deficiencies in training, phrasing of survey questions, administration of questionnaires, sampling households (or individuals) in the last stage, reluctance of respondents to provide data on international migrants, etc. Would it have made any difference if a non-government agency administered the survey, such as a university research center or a private firm?

This early experience with three countries—each choosing a different path—shows the need for flexibility, but with an eye on the theoretically ideal approach, to adapt its principles to fit country conditions and possibilities. Comprehensive, national-level surveys on international migration constitute a new form of data collection, but one certain to be repeated many times in the years to come. It is thus a learning process at this early stage, and the goal is to strive for improvement.

Design of Sample for the Egypt Household International Migration Survey (Egypt-HIMS)

1. Introduction

This document describes the development and implementation of the design of the Egypt Household International Migration Survey (Egypt-HIMS), planned to be carried out by CAPMAS in early 2013. It is well known that international migrants are relatively "rare elements" in a population, especially recent migrants. This is generally true throughout the world, whether one is interested in emigrants from primarily sending countries or immigrants inprimarily receiving countries. The usual recommendation for selecting a sample in these circumstances is to use stratification to create strata to use to allow oversampling of areas with higher proportions of international migrants (see Kish, 1965, and Bilsborrow et al. 1997, for adaptation to surveys of international migration). It is proposed that the MED-HIMS project, involving household surveys on international migration proposed for eight countries in the Arabic Mediterranean region (Morocco, Algeria, Tunisia, Egypt, Palestinian Territories, Jordan, Lebanon, and Syria), will therefore generally seek to use stratification to improve the efficiency of fieldwork in locating households with international migrants. Since Egypt collected data on international migrants in its 2006 population census and also has other national surveys with data on international migrants, it was anticipated that the sample in Egypt would be able to employ stratification of areas according to the proportion of international migrants (emigrants) in the population, followed by disproportionate sampling to select the primary sampling units (PSUs) as well as areas within sample PSUs, always oversampling areas with higher proportions of households containing migrants. However, for reasons described below, this procedure was only partially followed.

The organization of this report is as follows: First, overall objectives of the Egypt-HIMS are summarized, followed by a summary in section 3 of the sample design eventually adopted. Section 4 explains the selection of the sample in detail, but first reviews existing sources of data that could possibly provide the information needed to create a sampling frame, that is, data on the spatial distribution and prevalence of households in Egypt with international migrants and therefore the feasibility of developing a stratified sample. Based on the shortcomings of the existing data sources, the actual sample design adopted is described and the sample is selected. Section 5 completes the discussion of sampling procedures by describing procedures for selecting households in the final stage, followed by short sections on sampling weights and concluding observations and recommendations in sections 6 and 7.

2. Summary of Survey Objectives

The Egypt-HIMS is a national research programme designed to overcome the lack of data on international migration from Egypt by collecting reliable and representative data on the determinants and consequences of international migration and mobility. The Egypt-HIMS will be implemented as a part of the internationally coordinated MED-HIMS programme in collaboration with the EC, the World Bank, the ILO, the UNFPA, and the UNHCR.

The main objectives of the Egypt-HIMS are:

(i) to study the recent trends, causes, determinants, dynamics and consequences of international migration and mobility, and the inter-linkages between migration and development; and

(ii) to explore scenarios for a closer cooperation in the area of migration and development between Egypt and European and other destination countries.

Three definitions of migrant are relevant, for an emigrant or out-migrant living abroad, a return migrant, and a forced migrant A household containing any of the three is defined as a migrant household, with all other households defined as non-migrant households. For a migrant to be eligible for the individual interview (and for a household to be called a migrant household), there are two specific requirements, that the person's (last) migration occurred since January 1, 2000, and that the person was at least age 15 at the time of that migration. Thus for a household to be classified as a household with an emigrant, that emigrant must have left the household to live abroad at age 15+ since January 1, 2000, and continue living abroad. A return migrant, to be eligible for the individual interview, must have returned since that date and have been age 15+ at the time of return. A forced migrant to be interviewed must have been forced to leave his/her country of origin and come to live in Egypt at age 15+ since January 1, 2000, though the person may have first been in a different household/ residence or refugee camp than the one at the time of the survey.

3. Summary of Sample Design

Given the rareness of international migrants (though less so in Egypt than many developing countries), and the availability of sufficient funding, it was decided to select a large, nationally representative sample of around 80,000 households. The availability of a nationally representative Master Sample (MS) recently updated in 2010, covering 5024 enumeration areas (EAs), made it appropriate to use it for a sampling frame. The Egypt-HIMS was based on a sample of 1000 Primary Sampling Units (PSUs), which were taken to be the EAs from the master sample. The MS provides a sample of PSUs representative of both urban and rural populations and is of sufficient size to provide statistically reliable results for the five major regions of Egypt, and for urban and rural populations in each (with only the sample in the large but sparsely populated mostly desert frontier region too small).

The final Egypt-HIMS sample was selected in *two stages*: (1) 1000 PSUs (EAs) were selected from the MS in proportion to the estimated population size (PPES) of the primary administrative units in Egypt, governorates; and (2) households were selected at random

from existing (2010) lists of households in selected sample EAs, aiming to achieve an average of 80 households in each EA, which would be screened to determine migration status. Finally, two-*phase sampling* will be used, with the first phase being a listing of households to identify those containing an out-migrant living abroad (emigrant), a return migrant, a forced migrant, or no migrant; any household containing any of the first three is a migrant household, with all other households defined as non-migrant households. In the first phase, households with migrants would be oversampled compared to non-migrant households. The initial plan is to select in the sample all households containing any of the three types of migrants, plus a small number of non-migrant households (selected randomly). In the second phase, sample households are interviewed.

The total anticipated sample of migrant households is estimated *a priori* to comprise about 6,000 out-migrant households, 6,000 return migrant households, and several hundred forced migrant households. This is to be complemented by the selection of 3,000 non-migrant households, three in each sample EA. Taking into account that some small percentage of migrant households will contain more than one migrant, and/or more than one type of migrant (data will be collected for each migrant), the total number of individual migrant questionnaires to be implemented is expected to be somewhat larger, around 7,000 out-migrants and 7,000 return migrants. Since data will be obtained in each migrant household for one non-migrant (selected randomly), the number of non-migrant questionnaires to be completed is estimated to be about 6,000 + 6,000 + 3,000 = 15,000.

In the selection of households for screening, since the MS was updated less than three years prior to the planned implementation of the Egypt-HIMS in 2013, it was decided there would be no need for a second field operation to update the household lists in each sample EA—resulting in a considerable saving in field costs.

4. Development of Sample Design

The first need for selecting a sample is to have an adequate sampling frame, which requires having sufficient data to create strata. This is usually desirable in the case of surveys of international migrants since they are rare elements (Bilsborrow et al. 1997), but it requires having reliable data on households with and without international migrants of interest. For the MED-HIMS, the focus in all countries is on households with emigrants and return migrants, that is, which have adult members of the household which migrated to live abroad (out-migrant or emigrant households) and those which contain persons who migrated abroad and then returned to their origin household (return migrant households). The first step, therefore, was to examine the data on international migration from the 2006 census. It was already known that the census greatly undercounted (the stock of) Egyptians living abroad, counting only about 500,000 compared to estimates from independent sources ranging from 3 to 10 million, out of an enumerated population of 72 million in the 2006 census. This population had risen to an estimated 81 million people in 2010. If the undercount of migrants abroad were distributed randomly, the 2006 census data could still serve as an adequate sampling frame to develop strata based on the prevalence of migrants in administrative areas, but it was known that the undercount was far higher in urban areas, especially the large cities of Cairo and Alexandria.

To seek more reliable (and recent) data on international migrants, it was hoped that appropriate questions could be added to one or more rounds of the 2012 Egypt national labor force survey, but this was unfortunately not done, so an effort was made to examine the possibility of using data from a much smaller survey, from the (third round of the) National Labor Market Panel Survey (NLMPS), carried out by CAPMAS in 2011-12 (in collaboration with Professor Ragui Assad of the University of Minnesota). This survey included questions to identify if the household had any former member living abroad (both the total number and those aged 15+ when they left), as well as to identify current household members who had returned from living abroad. The total sample size of this survey was 12,500 households, from 476 enumeration areas. The total number of out-migrants living abroad was found to be only 803 (of any age and with no time cut-off limit on when they left), while 1367 return migrants aged 15+ were found. Since there is no time or age cut-off in the identification of emigrants, in contrast to the MED-HIMS project, so the NLMPS survey data overstate the number of out-migrant *adults* compared to the number of return migrants (which is limited to adults), providing an indicator of the potentially much greater prevalence of recent return migrants in Egypt. This raises an important question: should the planned survey select in the sample for interview every household with a return migrant as well as every household with an out-migrant, which may lead to too many return migrants relative to emigrants? This is returned to in the final recommendations in section 7.

Data in the NLMPS can also be used to determine whether the data from the 2006 census on Egyptians living abroad may be a useful indicator of the actual numbers of households with migrants (both abroad and return migrants). Thus part of the recent third round of this survey in 2011-12 is based on a new supplement of 200 EAs (of the 476 total) selected from the 5024 EAs of the national Master Sample, specifically from subsample number 1 of the four subsamples (each has 1256 EAs). Thus data on migrants and return migrants found in these 200 EAs could be compared with how these EAs would have been classified into (two) strata based on the 2006 census data (high and low expected prevalence of emigrants) to see if expectations based on the census stratification would have been born out in the 2011-12 survey once it was implemented. The question is, were significantly more migrants actually found per EA in the 200 in those EAs classified from the 2006 census in the high migrant stratum than in those classified in the low migrant stratum? If so, that would suggest that the 2006 census could serve as a basis for creating strata on the prevalence of international migrants, despite the known overall undercount. But if the answer is no, then there would be no basis for stratification based on the 2006 census data.

Returning to the NLMPS, stratification into high and low strata based on the prevalence of migrants in EAs from the 2006 census was used to oversample enumeration areas (EAs) from the 2010 Master Sample. Thus three times as many EAs were selected from the high stratum ratio of 3 to 1 (150 EAs in the high stratum and 50 in the low stratum), to increase chances of finding households with international migration experience in the NLMPS. Then the NLMPS was carried out, leading to the question, was the prevalence (or percentage of households with) of migrants found per EA in the high stratum significantly higher than that of the low stratum? As noted above, if so, that could provide a justification for using the 2006 census data to stratify all areas in Egypt according to the

prevalence of migrants and then select a stratified national sample, to improve prospects of finding international migrants in the proposed Egypt-HIMS.

Accordingly, the numbers of out-migrants and return migrants together found in the high and low strata of the NLMPS were tabulated, without regard to urban or rural area. In the high stratum, no migrants were found in 17 EAs and 558 migrants were found in the other 131 available EAs. The second step was to estimate the total population of the 148 EAs by multiplying 148 by the number of households in the sample cluster per EA in the NLMPS (10) times mean household size per cluster. With mean household size estimated by CAPMAS staff as 4 in urban areas and 5 in rural ones, and one third of the high stratum sample being urban and two-thirds rural[1], the weighted mean household size is calculated to be 4.67. Thus the total population in the survey in the 148 high stratum EAs is 148 x 10 x 4.67 = 6,869, and the proportion migrants is 558/6869 = .081. Meanwhile, in the low migration stratum, in 21 EAs, no migrants were reported while in the 21 with a migrant reported, the total was 94. The low stratum comprised 44% urban and 56% rural areas, so the weighted household size is .44 (4) + .56 (5) = 4.56, and the total population = 42 x 10 x 4.56 = 1915.2. Accordingly, the proportion of migrants found in the low stratum is 0.0491. While this is lower than the 0.081, it is over half the value in the high stratum. A rule of thumb is that the ratio of the two proportions should be at least 4 (versus 1.65 here) to justify using stratification rather than random allocation of areas. Therefore it was decided to select a non-stratified probability sample of PSUs, in Egypt, which is what CAPMAS desired to do from the beginning to take full advantage of its 2010 Master Sample preparation.

The MS of 5024 EAs is divided into four subsamples. As the other subsamples had been used for other recent surveys since 2010, it was decided to use subsample number 2. But first, the 24 EAs added to the nationally representative sample of 5000 EAs in the original MS were excluded (they had been added to provide a larger sample for the two relatively small urban areas of Port Said and Suez). Then 5000 EAs remained, so each of the four sub-samples had 1250 EAs providing a nationally representative sample. Once subsample 2 was selected, one of five of the EAs was randomly deleted, to leave a desired sample of 1000 EAs. This comprised 440 urban EAs and 560 rural ones, proportional to the 44% urban distribution of the population.

The sampling frame for the Egypt-HIMS was thus the national Master Sample, selected with PPES. In the first stage of the two-stage sample, a sample of 1000 EAs was drawn, with these EAs constituting the PSUs. In the second stage, within each sample EA, a sample of over 80 households (88 in urban areas and 84 in rural areas, to allow for likely non-response of 10% in urban areas and 5% in rural ones) was selected at random from the lists (EAs averaged 200 households at the time of the 2006 census, so with population growth, this was expected to be 230 on average at the time of the survey in early 2013).

---

[1] For stratification from the census, in the NLMPS 4% (of the population living abroad) was used to divide the rural EAs into high and low (migration prevalence) strata, while 1% was used as the threshold to divide the urban population into high and low strata. No one believes that the prevalence of households with international migrants in rural areas is multiple times that of urban areas. Thus it is suspected that the underreporting of Egyptians living abroad is far more in urban than rural areas.

The complete sample of enumeration areas and households was drawn by the head of sampling, Mr. Shaker El-Naggar, and his staff at CAPMAS, and was also provided to the consultant. A page showing the sampling approach is presented in Appendix C (with the full sample available only from CAPMAS).

5. Two-phase sampling in enumeration areas

The households selected in each EA will be visited by the field team to administer a short household questionnaire (from MQ1) to screen for households with and without migrants of the three types. Interviewers will administer this short questionnaire to the head of the household or his/her proxy respondent, or any adult member present. Meanwhile, the supervisor will complete two sheets, the first being a complete *listing sheet* to record, on one line per household (using landscape format), the name of the household head, number of household members, whether the household has a recent emigrant, a recent return migrant, or a recent forced migrant (see Appendix A). If none of the above three situations exists, the household is a non-migrant household. The supervisor thus records the data for the 84-88 households (of the 230 in the sample EA) listed by the interviewers he/she is supervising. In addition, at the end of the listing operation, the supervisor must record the total number of households of each type listed in the EA from among the 84-88 households drawn in the sample, the number of each type selected in the sample, and, when the fieldwork is completed in the sample EA, the number of each type successfully interviewed (see Supervisor Summary sheet for each EA, in Appendix B). Even though the preliminary plan is to select for interview every household with a defined migrant, the listing is still vital to draw the random sample of non-migrant households as well as to determine the weights applicable to each household type successfully interviewed in each sample EA.

In practice, whenever interviewers conducting the screening interview encounter a household found to have any of the three types of migrants (emigrant, return migrant, or forced migrant—see below), they will immediately continue to administer the rest of the household questionnaire and the appropriate individual questionnaires This latter will include administering the detailed individual questionnaires for each international migrant (emigrant) who left to live abroad (the last time) since January 1, 2000 if he/she was at least age 15 at the time of emigration; for each return migrant who returned to the origin household from living abroad since that date if he/she was at least age 15 on return; and to any forced migrant who came to live in this survey household in Egypt at age 15+ since the millennium cut-off date. In addition, in every migrant and non-migrant household, the interviewer will randomly select on the spot one non-migrant household member aged 15+ to interview, using digits provided by CAPMAS.

The summary listing sheet to be completed by the supervisor in each EA should include data on not only households but also on the total numbers of *individuals* of each of the four types found in each of the 84-88 sample households listed, the numbers selected in each sample household for interview, and the number of individual interviews successfully completed. This leads to additional weights for each type of individual interviewed in the EA, as well as the household weights.
.

6. Sampling weights

The final sample involves six household sampling weights based on the following (each weight must be the inverse of the probability of selection): (1) the sampling rate of the MS (proportion of the country's population covered by the MS, based on the estimated population in 2010, which will be the same as the proportion in 2013: about 1 million households were in the Master Sample in 2010 representing 17 million households); (2) the selection of second stage area units (*shiakhas*) from the MS (almost exactly one-fifth); (3) the weight for the sample EAs in the sample *shiakhas* (proportion of population), almost always 0.01 to 0.02; (4) the sampling rate of households for listing (sampling 84-88 per EA, from EAs averaging around 230 households in early 2013, so on average this will mean weights of about 3; (5) the weights to adjust for any sampling of households with different types of migrants (initial plan is to select 100% of all three migrant types in sample EAs, from the screening lists, implying no adjustment, but this could change after the pretest or even during the survey, provided this is carefully monitored), plus three non-migrant households from the mean of 80 listed in each sample EA (usually will have a weight of 15 to 25 per household); and (6) adjustments for non-response of each of the four types of sample households (there being NO replacement of non-responding or absent sample households). In addition, there will be (7) weights for *individuals* successfully interviewed within sample households in each of the four possible categories. For this reason, on the listing sheet *for each household listed*, data are needed on (a) the numbers of adults 15+ of each of the four possible types (the three migrant types plus non-migrants) as well as (b) the numbers sampled and (c) the numbers actually successfully interviewed. This results in household-specific weights for each type of individual migrant and non-migrant. For example, if there are 2 return migrants, but only one is successfully interviewed, then that one has an individual weight of 2. A complication arises with non-migrants. If the randomly selected non-migrant refuses, with no replacement, there is no respondent to weight and the household becomes a non-response.

7. Conclusions and recommendations

A question arises with respect to how much information to obtain in the short initial screening household questionnaire, about migrants. On the one hand, will households be willing to generally indicate, to virtual strangers at their door, not only if they have a migrant but also whether the migrant arrived since January 1, 2000. This is desirable to properly identify households which will have persons for detailed individual interviews. Otherwise, the screening survey will not distinguish households with recent migrants from those with earlier/lifetime migrants until the household survey is already completed and the individual interview has begun, at which time the interview will be cut short. Is this desirable? This remains an undecided issue. However, if in the pretest, or during the first month of the actual survey, households with a recent immigrant are rare, then there should not be a cut-off date for emigrants since too few will be being encountered (but still the age cut-off of 15+ at the time of migration must be used).

A more likely issue to be resolved is how common will households with return migrants be compared to households with emigrants? If a much larger number of households is found with return migrants, it may be desirable to not select 100% of them but rather some other

proportion for interview using some random selection procedure, such as one of every two. This would require field interviewing procedures as described above (automatically interview any household with a migrant) to be altered.

Appendixes

- Appendix A – Supervisor Listing Sheet for Enumeration Areas
- Appendix B – Supervisor Summary Sheet for Sampling in EAs
- Appendix C – Sample of Enumeration Areas (excerpt)

Egypt Household International Migration Survey (Egypt-HIMS)

Appendix A to the Sampling report
Supervisor Listing Sheet for Enumeration Areas

Interviewer: _____   Supervisor _____   Sheet _____ of _____

Governorate _____ Code _____ District/Markaz _____ Code _____ Shiakha/Village _____ Code for survey _____

List of Occupied Households in Sample Shiakha/Village

Complete one line for each household in the community. List on a separate line households that function as separate economic units even if they occupy the same dwelling and share kitchen or toilet facilities (as long as they do not pool income and eat together). Do not list households in which the head is a non-Egyptian (by birth or naturalization) except in the case of forced mogrants.

#	Name of household head	Total population living in household currently	Total number aged 15-59	No. of return migrants *Stratum 1*	No. of forced migrants 15-59 *Stratum 2*	No. who left who have not returned from abroad *Stratum 3*	No. of non-migrants aged 15-59 *Stratum 4*	2 digit random number assigned to select NM	If non-migrant interviewed completed, mark with x	Household selected and interviewed completed, mark with x	Total number of individual interviews completed *Ma*
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	!11)	(12))

Egypt Household International Migration Survey (Egypt-HIMS)

Appendix B to the Sampling Report
Supervisor Summary Sheet for Sampling in EAs

Governorate:_____Code _____Markaz:_____Code_____

Shiakha/Village _____Code_____EA_____Code____

Directions on how to get to Shiakha/Village/EA: _____

Supervisor: _____Regional coordinator_____

(1) Stratum for household	(2) Total number in EA	(3) Number in sample	(4) Probability of selection	(5) Number actually completed (record in parentheses hhs with no non-migrant)	(6) Household weight (office)
1 Return migrant			1.0		
2 Forced migrant			1.0		
3 Emigrant			1.0		
4 Non- migrant					
5 Not at risk		0	0	0	0
	Total Number of households at risk: 1+2+3 +4 =	Total Number of households in sample 1+2+3 +4=		Total Number of households completed 1+2+3 +4=	XXXXXXXXX
(7) Stratum for individuals	(8) Total number of individuals in EA in stratum	(9) Total number of individuals in EA in stratum in sample	(10) Probability of selection	(11) Number actually completed	Individual weight (office)
1 Return migrant			1.0		XXXXXXXX
2 Forced migrant			1.0		XXXXXXXX
3 Emigrant			1.0		XXXXXXXX
4 Non- migrant	XXXXXXXXXXX		XXXXXXX		XXXXXXXX

Steps for calculating probability of selection of households by stratum:

1. In the listing of households, households are *not at risk* of being sampled if the head is non-Egyptian, or if there is no adult household member aged 15-59 (i.e., only people over age 60).

2. Record the total number of households listed by stratum in the EA in column 2. A household containing any type of migrant is a migrant household and will be automatically selected in the sample and have a probability of selection of 1.00. The number in the sample in column (3) will therefore always be the same as in column (2).

3. However, the number completed may be smaller due to refusals, absences, incomplete interviews, etc. (column 5). In this case, those households actually completed in that stratum will have a weight more than 1.0 (computed in office).

4. For non-migrants, three households are to be selected in each EA randomly. Even if one or more of these refuse, are not available, etc., it is never replaced. The weights may thus be over 50.

5. Following interviews, record the data from the listing sheets on the total numbers of *households* listed, sampled, and actually completed in strata 1-4 plus the column totals.

6. (Office calculation) Calculate probability of selection of households in each stratum = (Number of households completed)/(Number of households in the stratum). Inverse is the weight.

7. Then repeat the procedure for *individuals*: First, record data on the total numbers of individuals in each of the four types or strata *but only for the households actually completed in the EA* (from column 5 above, the sum at the bottom), in column (8).

8. Any type of migrant (strata 1-3) is automatically selected into the sample and has a probability of selection of 1.00. Therefore, the number in the sample in column (9) will be the same as in column (8).

9. Record the number of individuals in strata 1-3 actually completed in column (11). Whenever anyone selected is not successfully interviewed, weights are computed in the office to adjust, but sinve this is done separately for each household, it cannot be computed from these data.

10. For non-migrants, the procedure is the same, except the total number of non-migrants in the sample of *households completed* (column 9 for stratum 4) is the *sum of* (a) the number of eligible non-migrant adults aged 15+ (excluding non-Egyptians) living in the (maximum of 3) non-migrant households actually completed (see row 4, column 5 entry); and (b) the sum of the numbers in column 5 for strata 1-3, since one non-migrant is sampled from each migrant household. The only exceptions are households in strata 1 and 2 which have no adult non-migrants.

11. A note should be made on this Supervisor sheet when this occurs, in column (5) recording in parentheses the number of households in the stratum that have no adult non-migrant (e.g., a whole family returns from abroad, or a whole household is comprised of forced migrants).

12. As per point 9 above, record the total number of non-migrant individuals in the sample (again without replacement) actually interviewed, in column (11). As in 9 above, weights cannot be computed from columns (9) and (11) since they must be computed separately for each household.

Egypt Household International Migration Survey (Egypt-HIMS)

Appendix C to the Sampling Report
Sample of Enumeration Areas (excerpt)

The Master Sample PSU's (5024)

Gov_Code	Gov_NAME	_Kism Code	ORDER	KISM_NAME	SHIAKH A_Code	SHIAKHA_SER	SHIAKHA_NAME	URBAN / RURAL	PSU	HH_size	F1	F2	F3	HH- size adj	Main cluster	F4	F3*F4	D.Weaght
1	القاهرة	27	1	قسم اول السلام	2	436	السلام الشرقية	1	6	241	0,070	0,202	0,014	220	88	0,400	0,006	177,878
1	القاهرة	27	1	قسم اول السلام	2	437	السلام الغربية	1	10	176	0,051	0,202	0,010	234	88	0,376	0,004	259,072
1	القاهرة	27	1	قسم اول السلام	2	437	السلام الغربية	1	14	217	0,063	0,202	0,013	244	88	0,361	0,005	219,103
1	القاهرة	27	1	قسم اول السلام	3	438	العصارة	1	18	232	0,067	0,202	0,014	204	88	0,431	0,006	171,340
1	القاهرة	27	1	قسم اول السلام	3	438	العصارة	1	26	216	0,062	0,202	0,013	255	88	0,345	0,004	230,040
1	القاهرة	27	1	قسم اول السلام	04	438	العصارة (بركة تشمل الأنفص)	1	30	201	0,058	0,202	0,012	252	88	0,349	0,004	244,299
1	القاهرة	33	1	قسم ثانى السلام	2	527	منطقة النهضة	1	34	200	0,058	0,202	0,012	222	88	0,396	0,005	216,292
1	القاهرة	33	1	قسم ثانى السلام	2	445	منطقة النهضة	1	38	190	0,055	0,202	0,011	213	88	0,413	0,005	218,446
1	القاهرة	31	2	قسم المرج	01	496	المرج البحرية	1	46	345	0,100	0,202	0,020	250	88	0,352	0,007	141,201
1	القاهرة	31	2	قسم المرج	1	496	المرج البحرية	1	50	198	0,057	0,202	0,012	297	88	0,296	0,003	292,287
1	القاهرة	33	2	قسم المرج	1	496	المرج البحرية	1	54	181	0,052	0,202	0,011	249	88	0,353	0,004	268,064
1	القاهرة	31	2	قسم المرج	02	497	المرج القبلية	1	58	207	0,060	0,202	0,012	252	88	0,349	0,004	237,218
1	القاهرة	31	2	قسم المرج	2	497	المرج القبلية	1	62	194	0,056	0,202	0,011	232	88	0,379	0,004	233,026
1	القاهرة	31	2	قسم المرج	4	499	عزبة النخل الشرقية	1	70	180	0,052	0,202	0,010	256	88	0,344	0,004	277,131
1	القاهرة	31	2	قسم المرج	04	499	عزبة النخل الشرقية	1	74	181	0,052	0,202	0,011	254	88	0,346	0,004	273,447
1	القاهرة	31	2	قسم المرج	7	502	كفر الشرف	1	78	188	0,054	0,202	0,011	253	88	0,348	0,004	262,229
1	القاهرة	20	3	قسم المطرية	1	316	العرب	1	82	220	0,064	0,202	0,013	221	88	0,398	0,005	195,743

Design of Sample for the Jordan Household International Migration Survey (Jordan-HIMS)

1. Introduction

This document describes the development and implementation of the sample design for the Jordan Household International Migration Survey (Jordan-HIMS), by the Department of Statistics (DOS), Hashemite Kingdom of Jordan, which is planned to be carried out in 2013 after the elections planned for January. The sample design was developed in collaboration with the DOS staff.

International migrants are generally "rare elements" in a population, particularly recent migrants. The usual recommendation for selecting a sample in these circumstances is to oversample areas with higher proportions of the rare elements, meaning in the present context households with international migrants (Kish, 1965; Bilsborrow et al. 1997). The general recommendation for the MED-HIMS project, involving household surveys on international migration proposed for eight Arabic-speaking countries in the Mediterranean region, is therefore to use stratification to classify areas according to the prevalence of international migrants (especially emigrants), then oversample areas with high prevalence of migrants, and finally to oversample households with migrants in sample areas already selected.

The organization of this document is as follows: The next section summarizes overall survey objectives for Jordan, followed by a concise statement of the sample design adopted. Then the major section of this document, section 4, first reviews existing sources of data that could potentially provide the data needed to create a sampling frame for selecting the sample for the Jordan-HIMS. The political-administrative divisions and subdivisions in Jordan are described, including the unusual, extreme variations in population sizes at each administrative level. Nevertheless, it was possible to develop an adequate sampling frame, create strata, and select a stratified sample, using subdistricts as the primary sampling units. Section 4 also describes recommended procedures for sampling smaller areas within subdistricts. Following this, three short sections conclude the document. Once the sample of areas is selected, section 5 describes methods for selecting households in the last stage, from small area units (blocks or clusters), while section 6 describes the importance of keeping track of sampling methods to derive sampling weights at each stage--since stratified samples are not self-weighting. A concluding section 7 offers

recommendations on things to be alert for during the pre-test and the beginning of fieldwork that may call for adjusting methods of sampling households within the final stage clusters.

2. Brief Summary of Survey Objectives and Definitions

The Jordan-HIMS is the centre of a national research programme to overcome the scarcity of data on international migration from Jordan by collecting detailed data on international migration from a representative sample of households in Jordan. The main objectives of the survey are:

(1) to collect, process and analyse detailed data to study recent trends in migration, processes and pathways of migrants, and characteristics of migrants (compared to non-migrants), as well to investigate the causes, determinants, dynamics and consequences of international migration in Jordan, and linkages between migration and development; and

(2) to use similar methodologies and questionnaires to facilitate closer cooperation in the area of migration and development between the Jordan and the other seven participating Arab Mediterranean countries and the destination countries of Europe and elsewhere.

The Jordan-HIMS is designed to collect data on a representative sample of households and individual migrants and non-migrants, to study out-migration to international destinations (hereafter referred to as emigrants or out-migrants), return migrants, and forced migrants. Topics covered include migrant characteristics, processes of migration, intentions to migrate (of both migrants and non-migrants), circular migration, migration of highly-skilled/educated persons, irregular migration, remittances and their use, and relationships between migration and well-being and links between migration and development.

Migrants are defined in many different ways by countries and investigators—a major problem in the entire field of migration data collection and research. To ensure the comparability of data across the participating MED-HIMS countries, representatives from national statistical offices of these countries have agreed on common definitions and questionnaires to collect data. All countries are focusing data collection on emigrants, return migrants and forced migrants, as stated above. Emigrants are former members of the survey household who have left to live abroad for at least six months and not returned, while those that have emigrated and then returned to the same origin household are considered return migrants. Finally, forced migrants are persons who have left their country of origin to seek asylum or refuge elsewhere, to escape political

or religious persecution or the fear of it, trafficking or coercion, civil strife (including the loss of family members), or environmental disaster. Those migrants who left (out-migrants), returned (return migrants), or arrived in the survey country (Jordan) (a) since January 1, 2000, and (b) were at least age 15 at the time of (last) migration will then be interviewed using a detailed individual questionnaire.

3. Summary of Sample Design

Given the availability of sufficient funding, it was decided to select a large, nationally representative sample of around 50,000 households to seek sufficient numbers of the three types of migrants. While both data sets with potentially useful data on international migration—the 2004 census of population and the new, just completed, Job Creation Survey (JCS) of 2012-- had serious limitations, the fact that they led to almost identical stratification of districts and subdistricts in Jordan led to accepting them as providing a reasonable basis for stratification.

Given the numbers of districts (51) and subdistricts (89) and the fact that large areas of the country appeared to have very few migrants, it was considered desirable to select a geographically concentrated sample of areas both (a) to find as many households with international migrants as possible, and at the same time (b) to concentrate fieldwork to make it more efficient. Therefore it was decided *a priori* that a sample of around 30 subdistricts would be sufficient to constitute the primary sampling units or PSUs. Three strata appeared to be the most reasonable number to use, based on a preliminary review of the two data sources. The main data source used was the JCS, since the data, albeit only on out-migrants, were very recent compared to data from the 2004 population census. However, for subdistricts in the JCS with prevalence values that were borderline in terms of which stratum to classify them in (high vs. medium, or medium vs. low), census data were used to break the tie, taking into account census data on the prevalence of both return migrants and out-migrants. In addition, the JCS covered only 60 subdistricts, so the remaining 29 had to be examined and placed into strata based only on census data. Prevalence values of migrants in these 29 were compared with the census values of the 60 subdistricts already classified based on the JCS combined with census data to ensure that the stratification criteria were as close as possible with those used for the other 29, to combine the 60 with the 29 into strata for the whole country.

The final result was a high stratum comprising 14 subdistricts, a medium stratum with 25 subdistricts, and a low stratum with 50 subdistricts. Based on estimates of the mean proportion or prevalence of migrants in the three strata and the number of subdistricts, the best sample selection was taken to be 13 (of 14) in the high stratum, 11 of 25 in the medium stratum, and six of 50 in the low stratum.

The second stage in the three-stage sample design was the selection of units called localities from the 30 sample subdistricts. Data were computed on the prevalence of migrants in all localities in the 30 subdistricts, using only the census, since the JCS does not have a large enough sample size and number of migrant households for meaningful computations (only 686 households out of 40,000 had one or more out-migrants). Once tabulations on the prevalence of emigrants and return migrants were obtained from the census, disproportionate sampling of localities and blocks was carried out, involving selecting a higher proportion of localities in each sample subdistrict in the high stratum compared to the other two, and similarly, selecting more blocks per locality in the high stratum compared to the other two strata. The final sample is indicated in Table 2 below.

Table 1. Final sample distribution across strata

Stratum	Subdistricts	Localities per Subdistrict	Blocks per Locality	Total Blocks in Stratum
High	13	8	4	416
Medium	11	4	4	176
Low	6	4	2	48

DOS indicated it desired that the sample localities be selected from the national master sample, which is unusually large relative to the size of the population of Jordan, covering just over 40% of the population (6000 clusters of the approximately 14,100 clusters in Jordan). Thus once the localities were tabulated according to the proportions of the population they had which were emigrants or return migrants, the next step was to determine which localities were in the master sample. On the basis of a sample of three very different provinces—Amman, Ajloun, and Ma'an—it was determined that the Master Sample covers 100% of the localities of the country (i.e., part of the population of each locality is included in the Master Sample). Procedures for selecting localities and blocks are described in section 4 below.

Once the PSUs, Secondary Sampling Units (Localities) and Tertiary Sampling Units (blocks or clusters) were selected, in the last stage households were to be selected for interview using two-phase sampling, involving listing households in sample blocks and oversampling those with migrants.

The number of households in the sample expected to have one or more emigrants currently living abroad was estimated *a priori* to be about 2,000, along with about 4,000 return migrant households, and under 200 forced migrant households. This was to be complemented by the selection of two non-migrant households per cluster, or 640 x 2 = 1,280 non-migrant households. Since one non-migrant is to be selected at random for the individual interview in each migrant household as well as in each non-migrant household, the *total number of non-migrant individual interviews will be about 7,280*. Meanwhile, if the number of migrants per migrant household is 1.2, then the total numbers of *individual* out-migrant interviews would be about 2,400, along with 4,800 return migrants, and perhaps 600 forced migrants (assuming a mean number of three persons ages 15+ per forced

migrant household), for a total of around 7,800 individual interviews of migrants altogether. (Section 7 describes possible deviations in this design.) *The total number of households in the final sample interviewed in detail will then be about 7,480, and the total number of individual interviews around 15,080.*

4. Development of the Sample

The first need for selecting a sample is to have an adequate sampling frame, which means having reasonably reliable data on the prevalence of migrants by administrative area to create strata. This requires having data on households with and without international migrants of interest. For the MED-HIMS project, the focus is on households with emigrants and return migrants, that is, which have adult members of the household which migrated to live abroad (out-migrant households) and those which contain persons who migrated abroad and then returned to their origin household since 2000 (return migrants). The first step, therefore, was to examine the data available on international migration from the just completed Job Creation Survey (JCS), which asked the household respondent whether the household included a question on former household members living abroad (with no time reference). The JCS was expected by DOS to have sufficient migrants to provide the data necessary to create the strata on the prevalence of migrants in Jordan to serve as the sampling frame. The second potential source of data, though already known to undercount Jordanians living abroad, was the 2004 census. The advantages and limitations of each are reviewed below, beginning with the census.

The census, of course, has the unique advantage of covering the entire country and therefore being able to provide data, without sampling error, on all 12 governorates, 51 districts, 89 subdistricts, 1,043 localities (some divided into areas and subareas), and 14,418 (census) blocks, with a mean of 66 households per block/cluster at the time of the census (5.1 million persons divided by 14,418 divided by a mean of 5.4 persons per household, in 2004). While 70 is the number of households routinely used by DOS for mean cluster size recently, with population growth, and in the absence of creation of more blocks, mean cluster/block size should have increased to about 82 currently. This is based on the official DOS population estimate of about 6,356,000 on October 9, 2012, in turn based on an estimated annual population growth rate of 2.1% (see DOS, *Jordan in Figures*, Amman: 2012). By the expected time of the survey in February-May, 2013, if household size were to remain at the same mean size of 5.4, cluster/block size would rise

to 83.5. However, it will be higher than this (say about 88) if mean household size is falling at all since 2004, as is likely. Thus around the world such a decline is occurring due to declining fertility, higher longevity, and rising divorce rates. It should be recalled that the *minimum* block or cluster size in Jordan was fixed by DOS at 30 households following the census, with a maximum of close to 300 households.

The census reported only 29,397 Jordanians as living outside the country in 2004, or only 0.6% of the Jordanian population (see footnote 1, Tables 6.1-6.4). Of these, only 20% were female, 71% had left within the previous 5 years and 80% within the past 10 years. Fully half reported having left for work reasons, 31% to study, 13% to accompany someone, and 5% for tourism. These data used a definition of emigrants (see census definitions provided in the 2004 Jordan census publications) abroad as "Jordanians living abroad, temporarily absent and expected to return....with students, military and diplomats abroad excluded?[1] This definition likely resulted in a significant undercount of Jordanians living abroad but this undercount may not have been as large as one would expect, since the census was carried out 3 years after the Iraq invasion of Kuwait, by which time vast numbers of Jordanians left the Gulf states to return to Jordan.

Correspondingly, the number of return migrants in the 2004 census—reported as having their *previous residence abroad* (up to any year in the past before moving to their current residence), was 425,000, or 9.1% of the Jordanian census population. This value was likely inflated in 2004, compared to a "normal" or current level in 2012, precisely due to the flood of Jordanians from the Gulf back to Jordan in the years right before the census. However, the large number is also partly due to the lack of any time cut-off in the question, on when the person returned to Jordan, so it would also capture those who fled the Gulf states in the early 1990's due to Iraq's invasion of Kuwait and the resulting American attack on Iraq in 1991.

Unfortunately, there is no other recent source of data available on return migrants, and the only other source of data on emigrants is the 2012 JCS. The JCS was carried out by DOS in 2012, with data entry just finished at the time of the consultant's mission at the end of September. The survey was a national survey, conducted using hand-held computers. Funded by the government of Jordan, it was designed to provide additional employment data that would go beyond the data in the national labour force survey, which is carried out quarterly on 12,500 households (50,000 per year). The

---

[1] DOS, *Main Results of Population and Housing Census, 2004. Volume 4. Jordanians Abroad, Non-Jordanians Abroad, and Persons with Special Needs* (Amman: DOS). Evidently the definition does not seem to have been applied.

JCS was intended to collect more in-depth data on unemployment, so areas of the country which had high unemployment rates were oversampled. It also included a question asking households if they had a former member living abroad. While it covered all 12 governorates and 51 districts, it found 0 migrant households in the governorate of Aquaba and only one in Ma'an, both in the south. Overall, only 686 (possibly 685 as one was possibly miscoded as 98) households with an emigrant were found out of 39,938 total households responding to the question in the JCS, containing 1,023 migrants, or 1.5 migrants on average per household with an emigrant. Tabulations on the proportion of households with emigrants by district for the 51 districts indicated a clear pattern of higher proportions of migrants in most of the districts in the center of the country (which contains Amman, which by itself had 179 of the migrant households), very low proportions in the south, and intermediate values in the northern region (each region has four governorates, with greatly varying population sizes, as well as prevalence of migrants).

In sum, both existing sources of data on international migration in Jordan— the 2012 JCS and the 2004 population census—have serious deficiencies: the small number and proportion of households with emigrants reported in the JCS, and the tiny number of emigrants (and proportion of the population) as well as the large (possibly exaggerated) number of return migrants in the census. But given their totally different data collection times, population coverage and methodologies, if they were to lead to similar conclusions about where migrants are concentrated in Jordan, that would lend credence to using them to classify areas in the country according to the prevalence of international migrants. On the other hand, if they failed to coincide often in the classification of areas, that would suggest that neither should be used to establish strata for sampling, and that the sample of areas for the proposed Jordan-HIMS could be a simple random sample based on population size, with probabilities of selection of areas proportional to estimated population size (PPES). While this would also have the advantage of being self-weighting, avoiding the need to develop and keep track of sampling weights at each stage of selection. But it would also result in a small proportion of households with migrants, equal to the proportion in the country, and a huge survey cost (like the JCS) of collecting data on migrants.

A simple statistical method was then devised to compare the classification of areas into strata using each of the two data sources, to see if they yielded similar classifications or not. For the JCS, the proportion of sample households with one or more emigrant was computed for the 51 districts.

Similarly, the proportions of population that were emigrants and return migrants were computed from the census for the 51 districts. In each case, stratum boundaries were created, independently, to create three strata based on the frequency of values from the source, with evidently very different *sui generis* stratum boundaries in the data sources.

For the JCS, a classification of the 51 districts into three strata seemed the most useful, with the low stratum being those with under 1% of the households reporting an out-migrant (emigrant), the high stratum being those with over 2% having an emigrant, and the middle stratum being those with 1-1.99% having an emigrant. The maximum value reported was 6.7% reporting an emigrant (Irbid, district no. 9), while 10 of the 51 reported zero emigrants.

In the case of the census, the small number of emigrants reported ("proportion of Jordanian households with a former member living abroad in 2004") led to the decision to take into account simultaneously also the incidence of return migrants by district ("proportion of Jordanian households with a member whose previous residence was outside Jordan"). For emigrants, only 3 cases had over 1%, with 8 more cases being over ½ of 1%, and the balance low. For the prevalence of return migrants, seven districts had proportions over 10% and over 25 under 1%. Using the two criteria together led to a classification of districts from the census data of 12 in the high stratum, 23 in the medium, and 16 in the low.

All 51 districts available from both sources were thus independently classified into three strata according to the JCS and the 2004 population census. The data were put into a Chi-square contingency table, Table 1 below. Data in this table were then used to determine if the classifications of districts provided by the two sources were independent, in which case neither could be considered to provide a reliable basis for stratification of areas in Jordan according to the prevalence of international migrants (since each would contradict the other)On the other hand, if they provide similar classifications of areas, they would be said to be "contingent" upon each other.

From the column and row totals of the table one can compute expected values for the nine cells by multiplying the percentages of the total in each row by the percentage of the total in each column (e.g., for the top left row-column entry, $10/51 \times 12/51 = 0.046$) to get the expected percentage of frequencies for that cell, which when multiplied by the total number of districts (51) yields the expected frequency for that cell (i.e., $0.046 \times 51 = 2.35$). This is then subtracted from the observed value (7), and squared. All such values are summed, which in the present case

Table 2. Contingency table for comparing stratification of two data sources

Job Creation Survey stratification, 2012		Census stratification, 2004			
		High	Medium	Low	Total
	High	7	2	1	10
	Medium	4	14	3	21
	Low	1	7	12	20
	Total	12	28	16	51

= 117.31, the Chi^2 value. Since the chi-square value for 4 degrees of freedom is 13.28 at the 0.01 level, independence of the cells is powerfully rejected, viz., the two data sources do not provide different estimates but rather are *highly related* to each other. This means they provide similar conclusions regarding stratification, which in turn provides a strong justification for using either or both together for stratification. While it is known that both have their limitations, those limitations are independent of each other, so the two can be considered to provide similar information about the geographic concentration of international migrants in Jordan. This, then, provides a statistical justification for using them to stratify districts. However, it was our *a priori* desire to select smaller administrative units as Primary Sampling Units to produce a more concentrated and efficient sample for fieldwork.

Extending the Chi-square study to subdistricts at this stage was not done since it would have provided little additional information as, unfortunately, the JCS only covered 60 of the 89 subdistricts anyway. So once these 60 were stratified drawing on both the JCS and the census proportions, the remaining 29 subdistricts in Jordan still needed to also be classified into strata, but this could be done only from using data from the census. This was done by examining the proportions of the population that were return migrants and out-migrants for the 29 subdistricts and comparing them with the values for the other 60 which had already been classified, i.e., by comparing the census (prevalence) values of the 29 with the census values of the 60 which had already been classified into high, medium and low strata. None of the 29 warranted classification into the high stratum. The

final result was the creation of three strata with the following frequencies: High 14 subdistricts; Medium 25 subdistricts, and Low 50 subdistricts.

The final step was to determine the allocation of the 30 PSUs (subdistricts) across the three strata. Proportional allocation (essentially what would result on average from random selection) would have led to selection of 6, 9 and 15 subdistricts from the three strata, respectively, resulting in an inefficient allocation of field workers to find migrant households. It is more efficient to allocate PSUs according to the expected proportion of migrant households to be found in the three strata. Based on the (albeit too low) proportions found in the two sources for households with emigrants, it was estimated that mean values would be approximately 0.032, 0.016, and 0.004 in the three strata, illustrating the potential value of stratification. When the 30 PSUs are distributed across strata taking into account these values (multiplying each mean value by the number of subdistricts observed in each stratum), the result is selecting 13 (of the 14) in the High stratum, 11 of the 25 in the Middle stratum, and 6 of 50 in the Low stratum. Other methods of allocation involving oversampling across strata could have been used, but allocation, as done here, according to the (mean expected) values of the key dependent variable, proportion of households with migrants, is a widely recommended approach for sampling across strata. This allocation of PSUs across strata should make for an efficient allocation of resources for finding households with migrants, if the PSUs are generally correctly assigned to the three strata. It does require, however, keeping track of the different weights for PSUs in the three strata resulting from the unequal proportions of subdistricts selected (see section 6 below), in this first stage of the sample selection.

Subdistricts were then sampled by the DOS staff and the consultant, using random numbers generated by computer. The resulting sample follows (with the first two digit number indicating the governorate code, the second the district code, and the third and fourth digits the subdistrict code:

High Stratum (13)

1111, 1121, 1131, 1141, 1151, 1161, 1191, 1251, 1411, 2111, 2161, 2191, 2411

Medium Stratum (11)

1193, 1214, 1231, 1241, 1311, 1321, 1331, 2151, 2181, 2211, 3111

Low Stratum (6)

1181, 2213, 2232, 3311, 3331, 3412

It should be noted that the sample includes almost all the subdistricts of Amman (governorate code 11); has at least one entry from every governorate except the two smallest ones, which also happened to have very few migrants, according to both data sources, Jerash and Tafeila; and has few subdistricts from the South region (four). The latter is to be expected since the region accounts for only a very small part of emigration and return migration from Jordan, according to the two data sources available.

The second stage administrative units in Jordan are called localities, of which there are 1,043, or a mean of 11.7 per subdistrict. Then in the third stage, the normal procedure would be to select blocks or clusters, of which there are 14,418 in Jordan, or a mean of 13.6 per locality. Then once the prevalence of migrants is computed for localities and blocks from either or both the JCS and the census, the goal would be to select a sample of localities from the 30 sampled subdistricts using stratification and then oversampling, again in proportion to the prevalence of migrant households in sample localities of each sample subdistrict; this procedure would then be repeated by obtaining data to estimate the prevalence of migrant households in census blocks among all localities selected in the second stage. However, this logical method is complicated in Jordan by the *vast difference*s in the population sizes of localities as well as of blocks within localities, which was not known until tabulations were obtained for subdistricts in the master sample on population sizes and proportions of Jordanians living abroad and return migrants by locality. Indeed, even subdistricts vary enormously by size, with some (3) rural ones with under 4000 people while some (2) urban oneshad over 450,000 people (the largest being 477,000, in Amman). These data, moreover, are for 2004, so the urban population sizes will be generally even higher in 2013. The number of localities per subdistrict varies from two to over 40, including 7 cases under four, and dozens with over 30. Similarly, the number of blocks per locality varies from one to over 70, with most rural localities having only one (precise figures on the ranges were unknown due to not having time to go through the entire census files). These blocks had a mean population size of 63 households at the time of the 2004 census, though DOS staff described them as generally having 70 in 2012. However, based on estimated population growth from 2004, the mean number of households per block can be estimated as 83.5 (if mean household size remained at the census level of 5.4—see above).

The fact that the population sizes of subdistricts, localities and blocks vary so widely—and that the numbers of localities per sample subdistrict, and numbers of blocks per sample locality, also vary so dramatically—complicates drawing the sample at the second and third stages considerably. To continue

the process of oversampling, it was thussuggested that DOS select 8 localities and then 4 blocks per locality from each subdistrict in the high stratum, 4 localities and 4 blocks per locality in the medium stratum, and 4 localities and 2 blocks per locality in the low stratum. This leads to the following numbers per subdistrict.

High stratum: 8 localities X 4 blocks = 32 blocks

Medium stratum: 4 localities X 4 blocks = 16 blocks

Low stratum: 4 localities X 2 blocks = 8 blocks

To test how this would work, data from the master sample for several high stratum subdistricts in Amman (code 1121 above) and Ajloun (2411) and low stratum subdistricts in Amman ((1181) and Ma'an (3311) were printed out and examined to see how localities might be selected at the second stage. In the 2004 census, the sample subdistrict in Ajloun had 30 localities with populations varying from 2 to 992 in rural areas (apparently 24) and 4668 to 17,414 in the apparent six urban areas. Only 9 of the 30 had any emigrant, while 14 had a return migrant. A key practical issue is then how much weight to give to the proportion of the population with emigrants vs. the proportion with a return migrant. Giving some subjective priority to the former, a trial sample was selected working with the DOS sampling chief. This led to selecting 8 localities, which included 5 of the six urban ones, so then automatically 4 blocks would later be selected from each of those, yielding 20 blocks of the desired 32. Note this means 4 blocks averaging 83.5 households, which with a mean population of 5.4 per household means blocks of 451 persons on average (though using the 2004 census mean of 63 would mean only 340 persons per locality on average, for comparison with the census figures above, and in the printout). Selecting 4 blocks would mean thus about 1800 persons in 2013, out of a whole population of the town or city of say 5000 to 20,000 in 2013. For the former, this would be a large portion of the population, perhaps too large (it would be harder to explain why only a few households are *not* selected). Meanwhile, the first rural localities selected, three, had only 218, 221, and 43 persons, or far fewer than the mean of 340 (2004 values) expected. The locality with 43 persons would have had only 8 households. In any case, all of the households would be listed (see below) by the field team in each of these 3.

But now with the 8 localities, we would only have 23 blocks selected vs. the 32 desired, so nine more would need to be selected. To do this would require selecting some combination of (a) all the other rural localities in the PSU with a single emigrant or return migrant in 2004, (b) adding the other

urban area not initially selected due to its lower proportion of migrants, and/or (c) selecting *more* than 4 blocks in some of the original 5 (urban) localities already selected. It is interesting, moreover, that the one locality in the subdistrict with a high proportion of emigrants is in fact a rural one (with 17%), while no other locality (urban or rural) had over 2%, and only 2 other localities had as many as 2% return migrants, oneb being the rural area with only 43 persons (meaning it had one person!). This illustrates another problem in sampling at this stage, the small numbers problem.

Opposite problems occur in sampling from the high migration intensity subdistrict of Amman (1121). It has only four localities, with populations of 47,000 to 190,000 in 2004, and low proportions of emigrants (only one with over 1%) and high proportions of return migrants (from 11 to 19%!). Selecting 8 blocks from each of the four could be done easily, but should not localities with higher proportions of migrants have larger numbers of blocks selected relative to other localities, which would be more consistent with the procedures above for Ajloun?

Finally, in a low stratum mostly rural, low population subdistrict (3311), there were 15 localities with populations from 9 to 1936, with four under 340. The initial selection of four localities, the two with the highest values for emigrants and the two with the highest values for immigrants (the maximum proportion for emigrants being 0.005 and for return migrants 0.043), led to selecting localities with populations of 216, 631, 592, and 1168 persons. From each, two blocks are to be selected, but three of the four do not have populations sufficient for that (under 680 persons).

Thus sampling localities and blocks will require some additional procedures (certainly including aggregating populations of localities under 340 persons or 63 households in 2004 so that all are at least that large, before the sample is drawn), and establishing criteria for weighting the proportion of emigrants compared to the proportion of return migrants in the locality in the selection process. While households with emigrants should be favoured, since they are much rarer than households with return migrants in Jordan, by how much should they be favoured, e.g., by a factor of 2 to 1, or 4 to 1? This is something that could be done *a priori* by computer from the lists of localities in the Master Sample for the 30 subdistricts in the sample before the sample of localities is selected. Even so, the list of localities in each sample subdistrict (by population size and proportion of the population which is emigrants or return migrants) will still need to be examined, along the lines of the discussion of the case subdistricts above, to select the sample of localities, and the number of blocks per locality.

The more disaggregated the data, the more unstable are the figures on migrant prevalence due precisely to the mobility of migrants—many living in localities in 2004 will surely have moved elsewhere by 2013. It is therefore recommended to *not* obtain data on the prevalence of migrants in census blocks to select blocks using oversampling, but rather to simply select them randomly from sample localities, using the number of blocks per locality and stratum, as indicated above. However, these rules can be altered if desirable, as long as careful track is kept of the procedures; the population of the blocks sampled must be known as well as the (2004) census population of all blocks in sample localities to compute weights for the blocks (see section 6 below).

5. Sampling households in the last stage: two-phase sampling

Once the last stage clusters or blocks have been sampled, two-phase sampling will be used. In the first phase, a listing operation is to be performed in each sample block in order to identify those households with international migrants of interest, and the vast majority of households which do not contain migrants of interest (non-migrant households). With this information, households with migrants will be oversampled. On average, there will be about 83.5 households per cluster or block in early 2013, which is an excellent number for listing since it can be managed by an average size interviewing team of 3-4 interviewers and one supervisor in one day, at least in urban areas. In rural areas it may take two days since the dwellings are not as close, and more walking time from one dwelling to another is necessary. But most of Jordan is urban, and close to a road.

A listing sheet will identify at the top the subdistrict name and code, locality name and code, and block name and code, the date, name of interviewer compiling the list, and supervisor. While the exact nature of the listing operation remains to be established, the usual procedure is to use one line per household to record the data, but this may be altered for Jordan (see Appendix A to Egypt report above for an example). The first column to the left is used to record the name of the head; then the columns moving to the right will be used to record the number of household members usually sleeping there; whether the household has any member living abroad who was at least age 15 when he/she left to live abroad; whether the household has any current member who has returned from living abroad since January 1, 2000 who was at least age 15 when he/she returned; and whether the household contains someone who has come to live in the household to escape civil strife or persecution in his/her origin country. The households listed are then

grouped into four strata, to use for oversampling migrant households in this last stage: i. households with an emigrant; ii. households with a return migrant; iii. households with a forced migrant; and iv. non-migrant households (all the rest). Households which are comprised exclusively of foreigners that are not forced migrants should be listed on the listing sheet but excluded from the sampling frame as not eligible households.

Using a landscape orientation, it is possible to list up to 30 or more households on a single listing sheet, which makes it possible for each interviewer to work independently and simultaneously under a supervisor. Once the listing sheets are combined for the sample cluster/block, a code can be assigned to each household to indicate its stratum. Then those households in the strata with migrants are sampled (selected for interview) with a high probability (probably select all households with an emigrant or forced migrant, and a large proportion of households with a return migrant), while the number of non-migrant households in each of these ultimate area units (UAUs) is fixed a priori at two households. The actual sample in this last stage may be drawn in the field by a (trained) supervisor, or the data from all the listing sheets may be brought back to DOS headquarters in Amman to draw the sample. In the latter case, it is desirable to also record on the listing sheet the exact street address of the dwelling where the sample household is located, to ease finding it later. In any case, it is desirable to send the same interviewing team back to the cluster.

The second phase of two-phase sampling is to conduct interviews in the households selected. Note that weights are needed to weigh the data for households since they are sampled using oversampling. Weights are also needed to adjust for non-response. Sample households should *not* be replaced by another household, e.g., one nearby or the next one on the listing sheet.

6. Sampling weights

The final sample comprises five sample weights for households: (1) a weight to adjust for the overall sampling rate of the Master Sample; (2) three weights for the inverse of the probability of selection of PSUs (subdistricts) from the three strata; (3) weights to adjust for the proportion of localities (or population) selected into the sample from each sample subdistrict; (4) weights to adjust for the probability of selection of blocks from sample localities, different for each locality; and finally (5) weights at the last stage, for differentially selecting households with migrants and without from the listing sheets prepared in the UAUs. All population figures below refer to the 2004 census of population.

For (1), the weight to generate data for the whole country is about 5/2 since the master sample constitutes 40% of the population of Jordan.

For (2), the weights are the inverse of the probability of selection of the subdistrict from each stratum: this is 14/13 for each of the 13 subdistricts in the high stratum; 25/11 for the 11 subdistricts in the medium stratum; and 50/6 for the six subdistricts in the low stratum. Actually, the weights should be modified to be the total population in the census in the subdistricts (for example, for the low stratum, this weight is the census population of the 50 subdistricts combined divided by the population of the six sample subdistricts combined).

For (3), the weight will be population of the sample subdistrict divided by the sum of the populations of all the sample localities selected from the subdistrict. This will vary across subdistricts, and may be called the locality-specific weight.

For (4), similarly, the weight will be the population of the sample locality divided by the sum of the populations of the blocks actually selected in the sample.

Finally, for (5), separate weights are needed for each type of household selected, thus four weights for the three types of migrant households and non-migrant households. These four weights are specific for each sample block, so careful records need to be maintained on (a) the number of households of that type listed, the number selected in the sample, and (c) the number actually successfully interviewed. The weights will always be (a)/(c) for each of the four types of households. If all migrant households are sampled, (a) = (b) so the weight for the migrant household will only adjust for non-response. For non-migrant households, the weight will be simply (a)/2 if two are always selected per block or cluster, as suggested here.

Each household will thus have attached to it five weights multiplied together for weighting the data so that the weighted data of the sample households represents the population of Jordan (excluding households comprising those foreigners who are not forced migrants, about 8% in 2004). Additional weights are needed for each of the four types of (adult) individuals—emigrants, return migrants, forced migrants, and non-migrants, which will vary according the the number found, the number selected for interview (usually all of the emigrants and forced migrants at least), and the number actually successfully interviewed.

7. Precautions and possible adjustments

There are two issues that need to be resolved (apart from the details of selection procedures for localities, described above). One is the fact that the Master Sample has not been updated for most of the country, and the other is the apparent much larger number of households with return migrants that may result in the fieldwork compared to households with emigrants.

The sample of blocks above, based on clusters of size 70 as DOS has been using, will yield a total population for listing of 44,800 households, but if population growth is taken into account, the mean population of a block or cluster will be 83.5, meaning the initial sample selected for screening will be 53,400 households. Nevertheless, is this the desired sample? There are areas around Amman which have experienced very rapid population growth since 2004, as the metro area is estimated to have grown by over 25%. It would seem important to include these new areas them in the sample frame, but how is this possible without incurring (as soon as possible) the cost of updating the sample frame and maps, at least around Amman?

Second, 2004 census data found over 10 times as many return migrants as Jordanians living abroad (emigrants), though the former was likely inflated by the timing of the census and the latter underreported. But even if the prevalence of the two differed by only a factor of 2 to 1 in the field in 2013, the survey wouldl yield twice as many return migrants and emigrants. Is that desirable? In the case of any doubt, the pre-test may help, but it will be only once the fieldwork has begun that this may be accurately observed. If so, and if there are judged to be too many return migrants being found compared to emigrants, then the proportion of return migrants sampled in the last stage in the UAUs may be reduced to half or less. This can be done even after the survey begins provided it is carefully kept track of when and where the switch is made.

Finally, once the Jordan-HIMS is completed, it will be of considerable methodological value to check the data to determine how the numbers (and prevalence) of households with migrants compared with what was expected, and hence how useful was the stratification.

Design of Sample for the Moroccan Household International Migration Survey (Maroc-HIMS)

1. Introduction

This document describes early and preliminary recommendations for the development of the sample design for the Morocco Household International Migration Survey (Maroc-HIMS), by the Direction de la Statistique (DS) and the Centre d'Etudes et de Recherches Démographiques (CERED) of the Haut Commissariat au Plan, which is now planned to be carried out in 2015 following the national population census planned for September, 2014. The sample design is accordingly only sketched here, since that census should provide data for developing a significantly better sampling frame than is possible at the present time. Nevertheless, given that other preparatory activities are also being currently undertaken in preparation for the 2015 Maroc-HIMS (development of detailed project plan and budget, discussions of changes to the MED-HIMS Model Questionnaires, modification of Arabic translation into Moroccan Arabic, etc.), it was also deemed appropriate to consider the sample design at this time even in a preliminary way. This included reviewing existing sources of data in Morocco to indicate how the new proposed 2014 census could be used to develop the sampling frame. Staff of both DS and CERED collaborated with the Consultant energetically and enthusiastically.

International migrants are generally "rare elements" in a population, particularly recent migrants. The usual recommendation for selecting a sample of rare elements is to oversample areas with higher proportions of the rare elements, meaning in the present context households with international migrants (Kish, 1965; Bilsborrow et al. 1997). The general recommendation for the MED-HIMS project, involving household surveys on international migration proposed for eight Arabic countries in the Mediterranean region, is therefore, whenever possible, to use a stratified sampling approach, involving first classifying areas according to the expected proportion of households with international migrants and then oversampling those with higher proportions of migrants.

The organization of this document is as follows: The next section summarizes overall survey objectives for Morocco, followed by a general description of the proposed sample design suggested for the 2015 Maroc-HIMS. Then section 4 reviews the two main existing sources of data that could have been used to develop a sampling frame if the survey were to be undertaken *before* the 2014 census. This includes an assessment of the advantages and limitations of the two data sources, and what they tell us and don't tell us about the prevalence of households with international migrants in Morocco. In this section also the political-administrative divisions and subdivisions in Morocco are described, which provides fodder for sketching how an adequate sampling frame could be developed for Morocco in section 6, though first section 5 indicates the questions recommended for the 2014 census. Section 6 then suggests a three-stage stratified sampling design to select the sample for the Maroc-HIMS in 2015 using the future 2014 census results. Section 7 indicates the types of sampling weights needed to keep track of and use to weight households at each stage--since stratified samples are not self-weighting —to derive overall weights for all sample households in all sample areas. A brief section concludes the report.

2. Summary of Survey Objectives and Definitions

The Maroc-HIMS is the centre of a national research programme to develop better data on international migration by collecting detailed, up-to-date data on international migration from a representative sample of households in Morocco. The main objectives will be:

(1) to collect, process and analyse detailed data to study recent trends in international migration, processes of migration, and characteristics of migrants (compared to non-migrants), as well to provide data to investigate the causes, determinants, dynamics and consequences of international migration in Morocco, and linkages between international migration and development; and

(2) to use methodologies and questionnaires as similar as possible across the participating countries in the MED-HIMS project to facilitate data comparisons and closer cooperation in the area of migration and

development between Morocco and the other seven participating Arab Mediterranean countries, as well as the destination countries of Europe and elsewhere.

The Maroc-HIMS is designed to collect data on a representative sample of households with and without defined international migrants as well as individual migrants and non-migrants, to study out-migration to international destinations (hereafter referred to as emigrants or out-migrants), return migrants, and forced migrants. Topics covered include migrant characteristics, processes of migration, intentions to migrate (of both migrants and non-migrants), migration of highly-skilled/educated persons, irregular migration, remittances and their use, relationships between migration and well-being of individuals and households, and links between migration and development.

Migrants are defined in many different ways by countries and investigators - a major problem in the field of migration. To ensure the comparability of data across the participating MED-HIMS countries, representatives from national statistical offices of these countries have participated in a number of international meetings in recent years and reached agreement on common definitions and questionnaires to collect data. All countries are focusing data collection on emigrants, return migrants and forced migrants, as stated above. Emigrants are defined as former members of the survey household who have left to live abroad for at least six months and not returned, while those that have emigrated and then returned to the same origin household are investigated as return migrants. Finally, forced migrants are persons who have left their country of origin to seek asylum or refuge elsewhere, to escape political or religious persecution or the fear of it, trafficking or coercion, civil strife (including the loss of family members), or environmental disaster, who are observed at the time of the survey in Morocco. Those migrants who left but have not yet returned (out-migrants, generally referred to as emigrants in the remainder of this report), and return migrants (sometimes referred to as returned migrants) who left and then returned to the origin household (a) within the 10 years prior to the survey date (or since January 1, 2000, in the case of the first countries to undertake the survey - the Palestinian Territories, Egypt, Jordan, and Lebanon), and (b) who were also at least age 15 at the time of their (last) migration, are then to be interviewed in a detailed individual

survey, together with their households. However, for emigrants, since they are rarely available at the time of the interview, another adult in the household (a "proxy respondent") will usually be asked to provide the data. Forced migrants who live in Morocco at the time of the survey and arrived at age 15+ since the cut-off date will also be interviewed.

3. Summary of Sample Approach

Based primarily on the availability of sufficient funding, it was decided to select a large, nationally representative sample of around 15,000 households of migrants and non-migrants in Morocco in 2015. This sample would be expected to collect complete data on about 6,000 households with one or more emigrants, 4,000 households with one or more return migrants, 2,000 households with only non-migrants, and perhaps 200-300 households with forced migrants. It is anticipated that of the 15,000 households with and without migrants sampled from the 2014 census results, about 20% (3,000/15,000) will be lost due to (a) normal refusals, (b) migration of the entire household away from its location at the time of the census by the time of the 2015 survey, or (c) operational difficulties in finding the exact same household as in the census, due to problems with mapping, etc. *It is therefore crucial that the Maroc-HIMS be undertaken as soon as possible after the 2014 census,* if possible, closer to six months than 12 months after the census, and that cartographic files from the census also be well organized and accessible to facilitate easy retrieval of census maps on the location of households selected into the sample for Morocco-HMS.

In the case of Morocco, prior to the arrival of the Consultant, it had been decided that, rather than conduct the survey before the planned September, 2014 census based on existing (known to be deficient) data sources, DS will include appropriate questions in the new census to better identify both emigrants and return migrants. This will then provide *updated data* on households which will then be used to both (a) develop strata for selecting a stratified sample of geographic areas or primary sampling units (PSUs) using oversampling based on the proportion of households with international migrants, and (b) identify households to visit and to interview when they are

confirmed to have emigrants or return migrants, as defined for the MED-HIMS surveys. The fact that the census will provide up-to-date data on all households in Morocco with and without international migrants will make it possible to avoid the usual need to undertake an expensive and time-consuming initial field operation in sample census sectors or enumeration areas to list and identify all households according to their migration status in order to *then* select a sample of migrant and non-migrant households (i.e., two-phase sampling will not be necessary at the last stage). Being able to use the census to identify all households in the country according to migration status will thus lead to a significant savings of time and resources. On the other hand, however, it will require that the survey be undertaken in less than a year following the census, which is likely prior to the creation of a new Master Sample. It will also require that some priority be given in the census data processing to process and tabulate the data from at least one question on international migration from the census (see below) in order to identify households with and without migrants to design the sampling frame.

As it is too complex for a short census to identify forced migrants, they will be covered according to their share in the total Moroccan population (without special sampling) by application of the initial section of the MED-HIMS Model Questionnaires (i.e., MQ-1 – Household Questionnaire) to all sample migrant and non-migrant households. This will identify any households which happen to have one or more forced migrants. These households will then be administered the relatively short forced-migrant module, MQ-5. It is anticipated that the number of such forced migrant households will be modest, not over 200-300 households, which should not present a problem. A second, larger forced migrants sample may be identified by using UNHCR databases, according to methods used in Egypt and Jordan.

To capture emigration in the proposed 2014 population census, the Consultant recommended using different (shorter) questions than those used in the 2004 census to identify (households with) emigrants (see section 5 below). At the same time, for identifying return migrants the questions used in the 2004 census are already excellent and should be used again in 2014.

Before providing details regarding the 2015 sample design, in section 6 below, this sample design report will review the two principal, potentially useful, existing data sources on international migration in Morocco—the latest population census (Recensement Général de la Population et de l'Habitat, 2004) and the latest national longitudinal survey (Enquête Nationale Démographique a Passages Répétés, 2009-2010). This will be instructive in indicating how to use the 2014 census to develop the 2015 sample.

4. Developing the Sample for the Maroc-HIMS: Key Existing Sources of Data

The first need for selecting a sample in a survey of international migrants (based on the principle of seeking "rare elements") is to have an adequate sampling frame, which, whenever possible, means having reasonably reliable data available on migrants by administrative area in order to create strata on the prevalence of international migrants by area. This makes possible oversampling areas with expected high proportions of international migrants of interest. But this process requires, first, being able to identify households with and without international migrants. For the MED-HIMS project, the focus is on households with recent emigrants and return migrants, that is, which have adult members who migrated to live abroad (emigrant/out-migrant households), orwhich contain former members who migrated abroad and then returned to their origin household since the cut-off date, such as 10 years before the census or survey (return migrant households). The first step, therefore, was to examine the data available on international migration from the two main potential sources of data—the 2004 Recensement Général de la Population et de l'Habitat, and the Enquête Nationale Démographique a Passages Répétés, 2009-2010 (ENDPR). The ENDPR followed a representative national sample of 105,000 households for 12 months, initially interviewing them in 2009 and then again 6 and 12 months later. Both data sources include questions seeking data on both emigrants and immigrants, including return migrants (defined as those recent immigrants who were born in Morocco or have Moroccan citizenship).

First, data on (recent) emigrants are reviewed from the 2004 census, and then compared with data from the ENDPR. Then the process is repeated for return migration.

Prior to the current UN round of censuses (2005-14), few countries in the world asked about *emigration* in their census of population. It is therefore impressive that Morocco already included a question in its 2004 census. Thus census data on emigrants come from a short table at the end of the census schedule to record data resulting from a question inquiring of the household respondent whether any member of the household 12 months before the census is no longer present in the household, and whether this is due to out-migration abroad or death. Nevertheless, there are a number of problems with this question, including: (a) combining emigration and deaths, both relatively rare events and at least one an unpleasant event; (b) asking about emigration but not inquiring at the same time about the far more common out-migration to *internal* destinations in Morocco, which is also important in development; (c) data on recent deaths from such direct questions are almost invariably significantly underreported in censuses and even household surveys; (d) asking also details about each person who left or died doubtless led to more non-response (or even the enumerator not wanting to ask the question at all); (e) errors often exist in recalling whether the event was within the 12 month cut-off period or not (which can lead to under- or over-estimates); and last but not least, (f) locating the question at the very end of a relatively long census schedule surely also contributed to non-response, if not to the enumerator neglecting to ask the question. Given these considerations, it is not surprising that there appeared to be considerable underreporting of emigration in the census. This is also the view of current DS officials. Nevertheless, there is no source of data available other than the census that covers all geographic areas of Morocco.

At the request of the Consultant, DS thus prepared a tabulation on households with at least one emigrant in the 12 months preceding the census, from September 2003 to September 2004, for the 16 regions, 61 provinces and 1,497 communes of Morocco (tabulation may be requested from DS, and is considered implicitly to be Appendix A to this report). As time available did not permit analysing these data at the level of the

commune, data were studied only at the level of the province, the number of provinces being sufficient in Morocco to indicate the level and spatial variation in emigrant prevalence rates (which will be compared with ENDPR). Thus the total number of households with at least one emigrant in the census was 33,602, or only 0.59% of the census total of 5,665,264 households. The highest 10 and lowest 10 provinces were then identified, to compare with the ENDPR. The range of values for prevalence of households with emigrants for the highest 10 was 0.87-1.44, while for the lowest 10, it was 0.06-0.25 (with the other 41 provinces in the interim range from 0.26 to 0.86). The maximum value was 24 times the minimum value, while the range from the mid-point of the lowest 10 to the midpoint of the highest 10 was about 8 times (from .15 to 1.16). The highest three census values for provinces were, in descending order, Beni Mellal in the region Tadla Azilal, El Kelfa des Sraghna in Marrakech, and Khouribga in Chaouia Ouardigha (1.44% to 1.2%)—all in the center of Morocco. The lowest values (under 0.2%) were Es-Semara and Assa-Zag in Guelmim Es-semara in the southern desert, Essouira in Marrakech, and Moulay Yacoub in Fes Boulemane in the eastern mountains.

Special tabulations were similarly prepared for the ENDPR, which covered 60 of the 61 provinces, and collected data for the same thing—persons reported by households as having left to live abroad within 12 months following the baseline interview. Thismulti-round survey with repeated visits to the same households should provide more reliable data—based on the actual enumerations of persons living in the same interviewed household three times over a short period of one year--compared to census data based on the memory of someone (a proxy respondent) about when someone else left the household (a minor additional source of differences could be due to the different reference year). The proportion of households in ENDPR reporting at least one person leaving in the 12-month period was 0.72, compared to 0.59 in the census, suggesting the census was not as poor as expected overall (.59/.72 = 82%). The top and bottom 10 provinces were again determined, along with the top three, which were Jerada in l'Oriental region, Al Hoceima in Taza Al Hoceima Taounate, and Puajda-Angad, all three in the northeast. At the same time, with the far smaller sample size of 105,000 households vs. 5.66 million, four provinces had 0 emigrants in 2009-2010 and two more

values of .10 or less, all but one of the six in the provinces of the three regions in the southern dessert. Thus even though there appears to be a rough correlation at the level of the region, with most of the regions that are above (and below) the median in the census also above (and below) the median in the ENDPR, this correlation cannot be very high at the level of the province since the top and bottom ones do not coincide at all and are sometime opposite. The conclusion is that there appears little evidence in support of the proposition that the two sources of data have a high correlation in identifying provinces with high and low prevalence of migrants.

Moving on to the data on *return migration*, first, the two data sources have a different time reference, the census providing data on the five year period 1999-2004 and the ENDPR on the one year period 2009-2010. From the census, data from cross tabulating information on (a) citizenship from question 4 and (b) place of residence in 1999 (at the time of the death of King Mohammed V) about five years before the census of September 2004 identifies households which had one or more current members living abroad in 1999 (532,507 households out of the total of 5.67 million census households). When the small number of households (3,564) with immigrants of non-Moroccan citizenship is excluded (but not excluding households with *both* Moroccan and non-Moroccan immigrants), the number of census households left is 530,069. This number, 9.34% per cent of total households, is the number of census households in Morocco that contained at least one Moroccan return migrant in the five year reference period. This percentage varied from 5.4 in Fahs Anjra in the extreme north region of Tanger Tetouan to 18.3 in Oued Ed Dahab in the extreme southern region of Oued Ed Dahab Lagouira. Indeed, nine of the 10 provinces with the highest percentage of households with a recent return migrant from abroad were in the three predominantly desert regions of the south (Rabat being the only exception), while at the same time, seven of the 10 with the lowest percentage of return migrant households are in the three extreme north-northeast regions of Tanger-Tetouan, Taza-Al Hoceima-Taounate, and Oriental.

Data from the 2009-2010 ENDPR were similarly tabulated to identify households that have an immigrant coming from abroad to live with them during the 12 month time period, finding only 167 households with at least

one immigrant, with the majority of provinces reporting zero, one or two households with an immigrant, and only four reporting 10 or more households.[1] Given that the ENDPR covered 60 provinces, it is evident that many will also have zero emigrants in a one year time period (in fact, 20), along with a dozen more reporting only one or two cases. As a result, the numbers are too small to yield meaningful estimates of proportions of households with an emigrant. Nevertheless, this survey and the census are the only two nationally representative sources of data on return migrants than can be compared. What we observe is, first, that the same three southern regions which had provinces reporting the highest percentage of households with a Moroccan return migrant in the 2004 census did not have a single province among the top 18 in the ENDPR. At the same time, the northern three regions reporting the highest return migration rates in the 2004 census (Tanger-Tetouan, Taza-Al Hoceima-Taounate, and l' Oriental) have a diverse mixture, with some provinces among the highest 18 in return migration prevalence along with 6 with zero values. Meanwhile, those in the top 18 in the ENDPR are also distributed broadly across the country, albeit concentrated in the provinces in the central urban regions of Casablanca, Rabat, and Marrakech. In summary, the conclusion is that the correlation between provinces reporting provinces with relatively high (and relatively low) prevalence of households with return migrants in the census and those reporting relatively high (and low) prevalence in the ENDPR is low—indeed, too low to rely on using either (or both) to classify provinces into strata based on the prevalence of households with return migrants.

The overall conclusion is thus that it is not advisable to use either of the two principal existing data sources to stratify provinces in Morocco according to the prevalence of households with either (recent) emigrants or Moroccan return migrants. While this does not prove that there is not a high positive correlation between the corresponding percentages at the level of the *commune*, it does cast serious doubt, and provides a strong case for *not* conducting the Maroc-HIMS before the 2014 census, and instead waiting to

---

[1] Data were not available in the tabulations on citizenship of return migrants, but judging from the data in the 2004 census, return migration of non-citizens is trivial so the missing data will not significantly affect the conclusions here.

use data from that planned census to create a sampling frame for conducting the survey in 2015.

5. Towards the 2015 Maroc-HIMS Sample: Preliminary Considerations

The first requirement for using the 2014 census to develop a sampling frame to select the sample for the Maroc-HIMS is to ensure that questions will be included in the census that permit the identification of households sending out recent emigrants and as well as of households receiving return migrants, since these are the two main groups of international migrants of interest for the MED-HIMS project that can be identified in a population census. Therefore it is vital to include appropriate questions in the 2014 census to identify these international migrants.

The census is expected to be carried out in September, 2014, to match the month of previous censuses in 2004, 1994, etc. The first and fundamental need then is to ensure that the new census includes questions sufficient to identify households with (recent) emigrants and return migrants. It is commonly assumed that five years is a reasonable period for respondents to recall migration events--that 10 years leads to more errors in recall while one year leads to numbers of events that may be too small to produce reliable data for small areas (see, e.g., recommendations of the UN Statistical Office, at UNstat.org, in press). This assumption underlies the recommendations for the census schedule below for Morocco, together with the goal of not overburdening the census with too many questions.

It is important to note at the outset that the three questions used to capture immigration in 2004 (21-23), including the return migration of Moroccan citizens, are excellent, and should be repeated in 2014. They ask, for each current member of the household, the place of previous residence, duration of residence in the current household, and the place of residence at the time of death of King Mohammed V and the crowning of King Mohammed VI in 1999, five years before the census. From special cross-tabulations on question 4 (citizenship) and question 23, it is possible to identify return migrants in the five years prior to the census, as discussed above. In the case of the 2014 census, in the absence of an easily

referenced date in the past (as was possible with the 2004 census), the question should be modified to: "Where was X living in September 2009, or about five years ago?" These questions can be used to identify households which contain a current member who was born in Morocco but was living in another country five years before the census, viz., a Moroccan return migrant.[2] Tabulations can be then made on the percentage of households which contain at least one return migrant in each region, province and commune, as described for the 2004 census of population in section 4 above.

With respect to emigration, on the other hand, based on the discussion in section 3 above, it is desirable to drop the long and complex question and table at the end of the 2004 census schedule which asked about deaths and emigrants in the 12 months prior to the census combined and replace it with a simpler question, perhaps right after question 23 on immigrants, as follows:

"Has any member of this household left to live abroad since September 2009 and continues to live abroad?"

This single short question should be sufficient to obtain the data needed to draw the sample for the Maroc-HIMS in 2015, as it identifies households with and without emigrants in the previous five years who are still away (recent current emigrants). This can be used to tabulate the prevalence or proportion of households with emigrants for all administrative areas of the country, which can then be used to create strata (see below). However, it would surely be of considerable interest for DS/CERED and Morocco to have selected additional information about the emigrants, to compare their characteristics with emigrants from other/earlier data sources, especially the 2004 census. If this is possible and the census schedule can be designed to accommodate additional data, questions recommended are as follows, with the responses to be recorded in a mini table. Thus:

If Yes, record the following for each such emigrant (allowing space for up to four persons):

---

[2] It is not a significant issue that the person may have returned and re-emigrated in the interim. Of course, persons who were living in the household 5 years ago, then left and returned during the five-year period, are missed by such a question, but they are likely to be small in number and hence not a major concern.

Name	Sex	Age when left	Education when left[3] (Level/Years completed at level)

This should be sufficient to collect key data (name being only for the enumerator to use and not processed) without imposing a burden on respondents, enumerators or those entering/coding data: Thus it is not recommended to obtain additional details *in a census* on month or even year of departure, date of birth, marital status, whether accompanied by other persons, work status, or occupation at time of departure, some of which had been obtained in 2004, likely contributing to the high non-response.

A single-round census (or survey) cannot collect even remotely reliable data on the *emigration of whole households*[4], .

6. Towards the 2015 Maroc-HIMS Sample: Preliminary Considerations for Sampling Households from the Anticipated 2014 Census

It is assumed that the 2014 census will include the appropriate questions, as described above, to measure both recent emigration and return migration, and that the resulting data are processed and could be used to prepare tabulations to estimate the prevalence of (a) households with recent emigrants, and (b) households with recent return migrants (whether based on a one year or five year reference period). It is further assumed that this will be during the months following the September 2014 census without waiting for the creation of a new updated Master Sample. The data need not be perfectly clean to provide an adequate frame, and indeed, following the discussion below*, it is probably only necessary to process the single question on whether the household has an emigrant in the previous five years to design the sample*. In any case, it is crucial that census training not

---

[3] This could be pre-coded in half a dozen categories, such as none, 1-4 primary, 5-6 primary, 7-9 junior secondary school, incomplete secondary school, complete secondary, incomplete university, completed university or professional school.

[4] If mean household size were 5, then it can be shown that the 2009-2010 ENDPR results suggest that 83% of emigrants migrate as individuals (and hence can be captured by a census or single round survey), and 17% as entire households. DS officials reported that data from the earlier panel survey in 1986-88 found 74% migrated as individuals. In either case, most migrate as individuals and can potentially be captured by the 2014 census. An interesting methodological study would be to compare the main characteristics of those who migrate as individuals vs. households.

neglect or relegate training pertaining to the correct administration of the questions on emigration. This should include encouraging census enumerators to probe, in the event of a negative response to the census question about whether the household has an emigrant who left within the previous five years, "Are you *sure* that no member of this household has left within the past five years to work, study or live abroad?" Otherwise, it is too easy for respondents (and enumerators) to say no to speed up the enumeration.

According to the 2004 census, the number of households with one or more Moroccan return migrants in the past five years was about 1,187,000 while the number of emigrants in one year was 33,602. If the latter were undercounted by 50% and further multiplied by 4[5], this would lead to a crude estimate of emigration of individual migrants of about 269,000 over a five year period, for purposes of comparison. Then the ratio of (individual) emigrants to return migrants over a five year period could be roughly estimated to be 269/1187 = .23. Given the predominance of return migrants, the sample frame to develop strata on the prevalence of migrants should be based mainly and perhaps solely on the *prevalence of emigrants* (but see concluding section below).[6] Once the areas are accordingly selected using only data on the prevalence of emigrants (see below), it will likely not be difficult to find sufficient numbers of return migrants in the same sample areas.

To consider how to use the forthcoming 2014 census to design the sample for the Maroc-HIMS, a quick review of administrative-political units used in Morocco is necessary. First, the census of population (among all possible

---

[5] It is multiplied by 4 instead of 5 to allow for both telescoping of data in retrospective memory-based reporting, as the reported number of persons who left within one, two, three, four and five years ago tends to decline over time even if the actual numbers were the same, to allow for undercount of emigrants who since died, and especially to allow for emigrants who have returned since emigrating.

[6] While this approach is the easiest and can be justified by assuming that return migrants must usually have left from interviewed households, since most (individual) emigrants return to their origin households when they come back, this is not true of all. Thus many who leave rural or small town households experience changes in tastes while living abroad, while others manage to save money, so instead they return to live in cities where they set up businesses or seek a wage or salaried job. This is most likely when they leave as single emigrants and change their marital status while away or on return. In this case, the person will appear in the census as an emigrant who left a different household (not the household where he/she is enumerated) but is a return migrant, in a new household, where enumerated in the census. If the person married someone in the other country, and returned with the spouse, the spouse is a return migrant if she had been born in Morocco and returned at age 15+.

sources of demographic data in developing countries) has the unique advantage of usually adequately covering the entire country and therefore being able to provide data, without sampling error, on all of Morocco's 16 regions, 61 provinces and 1,497 communes (1,298 rural, 199 urban), as of 2004.[7] The total census population in 2004 was about 5,665,000 households, 3.44 million in urban areas and 2.22 million in rural areas. This resulted in a mean number of households in urban areas of 17,286 per commune (varying from 215 to over 100,000) and 1,714 in rural communes (varying from 42 to perhaps 20,000). The population of each commune was further divided into Enumeration Areas (EAs) of around 160 households in 2004, meaning that there were on average 108 EAs in urban communes and 10.7, or one tenth as many in rural areas (based on mean household sizes of about 4.75 and 6.00, respectively, from the 2004 census).

In terms of anticipating how to select the sample in 2015, updated data on the number of EAs *per commune*, based on the new 2014 census, is recommended to be used for drawing the first stage sample, with EAs then selected in the second stage from sample communes (see below). However, it is interesting that while the current rate of population growth in Morocco is reported to be about 1%, there exist different views of current and projected fertility, which evidently has implications for population projections and therefore the number of people and households expected to be present at the time of the 2014 census. In current official projections, it is assumed that fertility in Morocco is continuing its downward trend, projected to decline from a total fertility rate of 2.5 in 2009 (from the ENDPR) to 2.1 by 2013y. However, DS officials informed the Consultant that the TFR was most recently estimated to still be 2.5 from the 2011 Morocco Demographic and Health Survey. Apart from this, the number of communes in Morocco has already changed compared to 2004 and so will the number of EAs per commune if the mean number of households per EA is to remain 160.

Following these preliminary considerations, the design of the sample for the 2015 Maroc-HIMS can be sketched below, based on how it could have been selected from the 2004 census if the data had been better.

---

[7] These numbers have already changed since 2004 and will evidently need to be updated to the 2014 census context.

Large areas of Morocco (as in most countries) have few recent emigrants (or return migrants), so it is desirable to select a geographically concentrated sample of areas in which migrants are concentrated to find as many households with international migrants as efficiently as possible, concentrating fieldwork to make it less costly to achieve any given sample size. Alternatively, a given budget will permit collecting data for a larger sample.

There exists a rule of thumb in determining whether it is worthwhile to form strata rather than to simply select PSUs randomly, that the mean values of the key selection variable should differ by at least a factor of four from the high to the low stratum. This is easily satisfied even at the province level in the case of Morocco. Thus, the highest 10 provinces had values from 0.87 to 1.44 (per cent of households with an emigrant in the previous year) while the lowest 10 varied from 0.06 to 0.25. The midpoints of the two (not identical to means but good enough for illustrative purposes here) were 1.15 and .15, and hence differ by a factor of nearly eight. Thus there is no doubt that a stratified sample of provinces could be drawn as PSUs in lieu of a random sample. But this would mean some provinces would be excluded completely at the outset (especially in the south) and still likely requires selecting a second stage sample of communes anyway.

Therefore, the preliminary recommendation of the Consultant is to *select communes as the PSUs*, and then Enumeration Areas (EAs) from sample communes as the second stage units. Using communes will make it possible to select areas more efficiently on a national scale than beginning with provinces, and pinpoint high prevalence PSUs even if they are located in provinces with low mean prevalence levels. Note, however, that this does not guarantee than all 61 provinces will be represented. That would require a *different* sample design, in which some minimum number of communes would be selected per province even from the low prevalence provinces as well as limit the number in the high prevalence provinces. Such a design is certainly possible but more complex and not discussed further here, where the approach is intended to select a *nationally representative sample* as efficiently as possible.

If province values range from 0.06 to 1.44, commune values for mean prevalence of households with emigrants must vary far more widely, from 0 to perhaps 10 per cent or more. In the selection of communes as PSUs, it shouldl be possible to form at least four strata which could have limit points and frequencies (hypothetical) as indicated in Table 1 (below). The stratum limit points reflect the expectation, customary in migration data, that there is a strong clustering of areas (communes) with no recent emigrant households or very few, with the highest frequency at zero (the mode), tapering off, exponentially, towards the maximum value, taken here to be 9.9%. The latter value could be considerably higher in a few communes with small populations, though this is limited by the fact that those communes with small populations tend to be rural ones which in most countries tend to have a lower overall prevalence of households with emigrants. The N_h values are assumed frequencies, perhaps not too implausible for communes in the 2014 census.[8]

The table shows how inefficient proportionate sampling would be in terms of distributing the field work evenly across strata, and also provides a simple alternative way to distribute the sample across the four strata in a way which results in heavy oversampling from stratum 1 and undersampling from stratum 4 (compute n_h/N_h values).[9] Other ways of oversampling are also available.[10]

---

[8] After this report was finished, the Consultant observed data which showed that only 199 of the 1497 communes were urban in 2004, but covered 60% of the population. In order to make it possible to select more EAs from the far more populous urban communes (than the 8 EAs proposed for each, if in stratum 1), where there are likely more migrant households, the Consultant would recommend modifying the sampling plan presented in the text. Thus urban communes should be listed in the stratum to which their prevalence level corresponds *multiple times* in accordance with their population size. Instead of listing only 199 urban communes, there would be a total number reflecting the urban share of the population of Morocco. Since there were 1298 rural communes representing 40% of the population, the number of urban communes with a similar mean population size as the average of the rural communes would need to be about 2000, or 10 times the actual administrative number of 199. Hence, on average, urban communes would be listed on average ten times, but scaled up or down according to their relative population size. This is not difficult to do and could be readily implemented once the new 2014 census has been carried out.

[9] The table illustrates a distribution for a total sample size of 108 communes. Depending on the budget, this could be higher.

[10] A number of other ways of oversampling can be considered, including so-called "optimum sampling" based on the standard deviations of the mean of the four strata, as well as other methods of disproportionate sampling. The Consultant would in fact suggest selecting 42, 32, 22, and 12 communes from the four strata assumed in Table 1, since this increases the likelihood of finding EAs with households with emigrants. The oversampling can be readily adjusted for using weights (next section).

Table 1. Illustration of use of communes classified in 4 strata to draw the sample for the Morocco-HIMS in 2015 (hypothetical values)

Stratum No. and Stratum limits (%)	N_h	Mean Proportion International Migrants	(1)x(2)	Standard Deviation of Mean	N_h with	
					Proportionate Allocation	Optional Allocation A
	(1)	(2)	(3)	(4)	(5)	(6)
1. 5.00 – 9.99	100	.070	7.0	.260	7	27
2. 3.00 – 4.99	400	.036	14.4	.190	27	27
3. 0.50 – 2.99	500	.010	5.0	.100	34	27
4. 0.00 – 0.49	600	.004	2.4	.064	40	27
Total	1500				108	108

The second stage of the sample is to select EAs from the 108 PSU sample communes. The number of EAs per rural commune varies from 1 to around 150 while that for urban communes varies from 2 to about 600, based on 2004 population sizes per commune and an assumed typical EA size of 160, as used in the 2004 census (all of these numbers, including mean EA size, may change by the time of the 2014 census). Separate samples could be drawn for urban and rural communes, but table 1 assumes this is not necessary, so we are dealing with numbers of EAs that could vary from one to about 600 per PSU. It is logical to continue oversampling as above, within sample communes, as follows. First, one computes the percentage of households with one or more emigrants for each EA in each sample commune. Then one (over)samples EAs in proportion to the per cent of households in the 2014 census with an emigrant. A straight-forward procedure for doing this is to first list all EAs in the sample commune in some logical (e.g., geographical, alphabetical, etc.) order, recording their per cent of households with an emigrant. Second, these per cents are cumulated, creating a listing of rising numbers up to the sum of all the percentages for all the EAs in the commune for the last one in the list. Then based on the number of EAs to be selected from each commune (see below), the sampling

interval in the commune is determined. This plus a random start makes it easy to use systematic sampling to select the desired number of EAs per commune.

But how many should be selected per EA? Presumably it would again make sense to select more from communes in stratum 1 relative to stratum 2, more in stratum 2 than stratum 3, etc. Let us suppose we select 8 EAs from each sample commune in stratum 1 (if there are fewer than 8, all are selected automatically), 6 from each commune in stratum 2, 4 from each in stratum 3, and 2 from each in stratum 4. This yields a total of 540 EAs in the national sample. Using the procedure above, these could be selected with systematic sampling, i.e., in proportion to their 2014 census percentage of households with an emigrant. As an example, suppose the mean number of households with emigrants per selected EA in stratum 1 is 20, and the means in strata 2, 3 and 4 are, respectively, 10, 6 and 4,[11] then the total number of possible households with an emigrant would be 27 x (8x20 + 6x10 +4x6 + 2x4) = 6,904. If 4 non-migrant households are selected from each sample EA in all strata, then the total number of non-migrant households selected is 2160. While the number of households with return migrants is not known a priori using this approach, it is likely to be larger than the number of households with emigrants even though only the latter is used to select PSUs, because of the considerably higher prevalence of households with return migrants than emigrants in the country (see section 4 above). If so, then it will be desirable to randomly select some fraction (such as a half or a quarter) of the 2014 census households found to have an apparent return migrant to visit for interview rather than all of those househlds. The goal is to select a fraction that will result in an *a priori* sample of at least 4000 households with one or more return migrants.

The above are all rough estimates and are only presented to indicate the main principles to use in selecting the sample for the 2015 Maroc-HIMS based on the actual 2014 census results, with households classified according to emigration status from the census data. When the sample households selected from sample EAs at the last stage are visited to

---

[11] These are the means following the use of systematic sampling, as the means across all EAs in all communes in the strata are much lower, and indicated by the values in Table 1. The means in the text are higher precisely due to the use of oversampling of EAs in the second stage within sample communes.

conduct the detailed follow-up MED-HIMS household and individual interviews, some will be found to *not* have an appropriate adult migrant: Thus the person reported in the census as having emigrated from the household did so outside the reference period, or was not aged 15+ at the time of emigration; or the census household refuses to provide the data for the emigrant; or the return migrant refuses to be interviewed; or the census household cannot be located. Thus the actual number of households with appropriate migrants who also respond will be smaller than expected, perhaps by as much as 10-20%, which should be tested for in the pretest. This is also why the sample size drawn should be larger than the sample desired. It will not be possible to have precise data on emigrants or return migrants from the census since it is not practicable to ask the additional questions in the census needed to identify with certainty if the emigrant or return migrant qualifies for the individual interview. If not, the household is not interviewed either, but recorded in a supervisor sheet in a column along with the non-responses (see below).

At the same time, some responding households will have more than one emigrant, or return migrant, or even have both. All households selected in the sample and visited which are found to have an appropriate emigrant, return migrant, or non-migrant should be administered the MED-HIMS household and appropriate individual questionnaire(s).[12] In each migrant household, the migrant (if qualified on basis of age at migration and migration in the reference interval) should be interviewed individually (by proxy, if an emigrant), along with one non-migrant adult selected randomly. In each non-migrant household, one non-migrant adult should be selected randomly for interview, to complete the household interview. There should never be replacement of a selected migrant or non-migrant household by another household, regardless of the reason for not conducting the interview. Normally there should also not be replacement of a randomly selected *individual* by another person in the household; rather a time should be sought when the person could be expected to be available and a return visit

---

[12] This includes the forced migrants encountered from applying the MED-HIMS questionnaire 1. The number is expected to be only 200-300 in the total sample, representative of their share in the total population of Morocco.

made. This and other details of oversampling in migration surveys are discussed in Bilsborrow et al. (1997).

If the mean number of emigrants per emigrant household were 1.2, then the total numbers of *individual* emigrant interviews, assuming each is interviewed, is 8,285 (but in practice less for reasons indicated), while if one non-migrant is interviewed in each migrant household along with each non-migrant household, the total number of *non-migrants interviewed* would be 6,904 + 2,160 + 4,000 (if 4,000 is the number of return migrant households) = 13,064, which is the *total number of households* as well. This is why a larger sample of households with non-migrants is not proposed (throughout the MED-HIMS project countries). If 100% of all households had eligible respondents who were successfully interviewed, the *total number of individuals* in the final sample interviewed would be 6,904 x 1.2 + 4,000 + 13,064 = 21,349.

Tabulations to select from the 2014 census are based on those DS made from the 2004 census, on both emigrants and return migrants (see text above and Appendices A and B). This should be done at all the main geographic levels, for regions, provinces and communes, urban and rural, and also for all EAs (about 40,000 in 2004, though the only EA data needed for the second stage sample would be from the communes selected in the first stage sampling process.

A final matter is to clarify the need to develop and use a supervisor summary listing sheet (see example for Egypt above) for each of the (e.g.) 540 sample EAs selected in the sample, to keep track of the number of households with each type of migrant and non-migrant (according to the 2014) census, the number selected in the sample, and the number successfully interviewed, meaning that both the household interview and all the desired individual interviews were completed satisfactorily. A row will be needed to indicate the total number of households selected and expected to be visited with emigrants (in the EA), the number who did not qualify, refused, or were not found, and the number successfully interviewed. This is needed to determine the EA-specific weight for emigrant households in each EA. For return migrants, the apparent number observed in the census is tabulated for the EA, perhaps half are randomly selected for visit, then of those the number successfully

completed is recorded, to determine the weight for return migrant households successfully interviewed in the EA. The procedure is similar for non-migrant households and forced migrant households found from the administration of MQ1 to all of the sampled emigrant, return migrant, and non-migrant households. What is described in this paragraph is a limited form of two-phase sampling (see Bilsborrow et al., 1997).

Even though census listings will be used to identify sample migrant and non-migrant households to be visited, a household listing sheet will still also be needed to record at the top the province, commune and EA names and codes, date, name of interviewer compiling the list, and supervisor. While the exact nature of the listing operation remains to be established for this survey, the usual procedure is to use one line per household to record the data across the sheet for each household in the sample EA from the census, identifying the name of the household head, whether it is an emigrant, return migrant, forced migrant, or non-migrant household, whether selected in the sample, and then whether successfully interviewed. DS staff in Rabat will draw the entire sample based on census data for each sample EA in each sample commune *prior to the fieldwork*; thus no sampling will need to be done in the field by supervisors. Note that no (e.g., non-cooperating) sample household not interviewed successfully should ever be replaced by another household by a supervisor or interviewer in the field, e.g., by one nearby or the next one on a listing sheet, as that invariably introduces biases.

7. Sampling Weights

The final sample comprises four sets or levels of sample weights: (1) four weights to adjust for the inverse of the probability of selection of PSUs (communes) from the four strata depending on the stratum they are in; (2) weights to adjust for the proportion of the population of the sample commune represented by the EAs selected into the sample using oversampling; (3) weights to adjust for the probability of selection of households of each type from lists for sample EAs compiled from the 2014 census (100% of those with emigrants or forced migrants found from administering MQ1; perhaps half of return migrant households (to be

determined), and 4 non-migrant households from the X in the EA); and finally (4) individual-level weights at the last stage for differentially selecting and completing individual-level interviews in sample households.

For (1), the weights are the inverse of the probability of selection of the commune from each stratum. Actually, the weights should be adjusted to be the inverse of the total census population in those communes selected in the sample from the stratum divided by the total census population in all communes in the stratum. The same argument applies to (2) weighting the data for sample EAs selected from each sample commune. For (3), separate weights are needed for each type of migrant and non-migrant household selected in each EA, following the discussion in section 6 above, resulting in four EA-specific weights for the three types of migrant households and for non-migrant households. As these four weights are specific for each EA, careful records need to be maintained using the listing sheets and supervisor control sheets, mentioned above.

Each household will thus have attached to it three weights multiplied together for weighting the data so that the weighted data of sample households represents the total population of Moroccan households in the 2014 census.

8. Precautions and Possible Adjustments

First, the overriding issue is whether the 2015 census can significantly reduce the undercoverage of emigrants (especially), and to a lesser degree, of return migrants, compared to the previous 2004 census.

Second, will there continue to be a much larger number of households with return migrants compared to households with emigrants, even if only data on emigrants are used to create the PSUs, as suggested above? If so, it leads to the question of whether households with return migrants should be undersampled relative to those with emigrants, as suggested above; but if not, it leads to a different question--whether data on *both* the prevalence of households with emigrants and the prevalence of households with return migrants should be used to develop the strata to select the sample of communes.

Finally, much remains to be seen with the new data forthcoming from the 2014 census. It is only the new data, hopefully with less undercoverage, that one can use develop a full updated and appropriate sample design. The discussion here is only intended to outline a plausible approach; the details remain to be developed once the actual data have been collected, and the desirable tabulations from the 2014 census prepared.

Appendices

Tables with the following titles were prepared by DS for the Consultant, and are available from DS.

A. Census of 2004: Number of households with at least one emigrant between 2003 and 2004, by region, province and commune.

B. Census of 2004: Number of households with at least one immigrant based on residence in 1999 and citizenship, by region, province and commune

C. Longitudinal Household Survey of 2009-2010: emigration of individuals and households, and immigration in 12-month period

References (apart from country data sources indicated in the text)

References Cited

Bilsborrow, Richard E., Graeme Hugo, A. Oberai, and Hania Zlotnik, 1997. International Migration Statistics: Guidelines for Improving Data Collection Systems. Geneva, Switz.: The International Labour Office.

Kish, Leslie. 1965. *Survey Sampling*.   New York: John Wiley & Sons, Inc..

United Nations, in press. *Recommendations on Statistics of International Migration*. New York: Statistical Division, Department of Economic and Social Affairs, United Nations.

MED-HIMS is a joint programme of regionally coordinated migration surveys currently supported by the European Commission (EC), the World Bank, the UNFPA, the UNHCR, the IOM, the ILO and the League of Arab States (LAS). It has its roots in the EC's MEDSTAT programme, and has been developed at the request of and together with the NSOs of seven of the MEDSTAT countries (Algeria, Egypt, Jordan, Lebanon, Morocco, Palestine and Tunisia).

MED-HIMS Coordination Committee (as of October 2013):

- **European Commission: EuropeAid** (Unit F4 - Regional Programme Neighbours South), **Eurostat** (Unit A6 – Statistical co-operation)
- **The World Bank** (Development Data Group)
- **UNFPA** (Arab States Regional Office)
- **UNHCR** (Division of Programme Support and Management)
- **IOM** (Regional Office for Middle East and North Africa)
- **ILO** (Arab States Regional Office)
- **The League of Arab States** (Migration and Arab Expatriates Department)

MED-HIMS participating national agencies:

- **National Statistical Office (ONS)** , Algeria
- **Central Agency for Public Mobilization and Statistics (CAPMAS),** Egypt
- **Department of Statistics (DoS),** Jordan
- **Central Administration of Statistics (CAS),** Lebanon
- **High Planning Commission, Directorate for Statistics (HCP/DS),** Morocco
- **Palestinian Central Bureau of Statistics (PCBS),** Palestine
- **National Statistical Institute (INS),** Tunisia

For more information:

Contacts for the MED-HIMS Programme and its model manuals

- Samir FARID, MED-HIMS Chief Technical Adviser, samirfarid@hotmail.com
- Giambattista CANTISANI, Key Expert for MEDSTAT III Migration, gcantisani@ymail.com

Contacts for national surveys / National Coordinators for MEDSTAT III Migration

- Amel LAKEHAL, Director of Population and Employment Statistics, ONS, Algeria, lakehal@ons.dz
- Abdallah ABDEL RAZEK, Specialist of Studies & Researches, CAPMAS, Egypt, pres_capmas@capmas.gov.eg
- Mohammad AL-ASSAF, Director of Population & Social Statistics, DoS, Jordan, assaf@dos.gov.jo
- Lara BADRE, Sociologist, CAS, Lebanon, lara_badre@yahoo.com
- Bouchra BOUZIANI, Head of Social Indicators Service, HCP/DS, Morocco, bbouziani@statistic.gov.ma
- Mohammed DURAIDI, Director of Technical Operation, PCBS, Palestine, mduraidi@pcbs.gov.ps
- Nadia TOUIHRI, Head of Social Statistics Service, INS, Tunisia, touihrinadia.ins@mdci.gov.tn

Other means

- MED-HIMS page at Eurostat Website: http://epp.eurostat.ec.europa.eu/portal/page/portal/european_neighbourhood_policy/enp_south/med_hims
- MED-HIMS Newsletter (issued every 6-9 months)
- Progress reports, papers and other documents for international meetings