

Key Issues of Survey and Sample Design for Surveys of International Migration

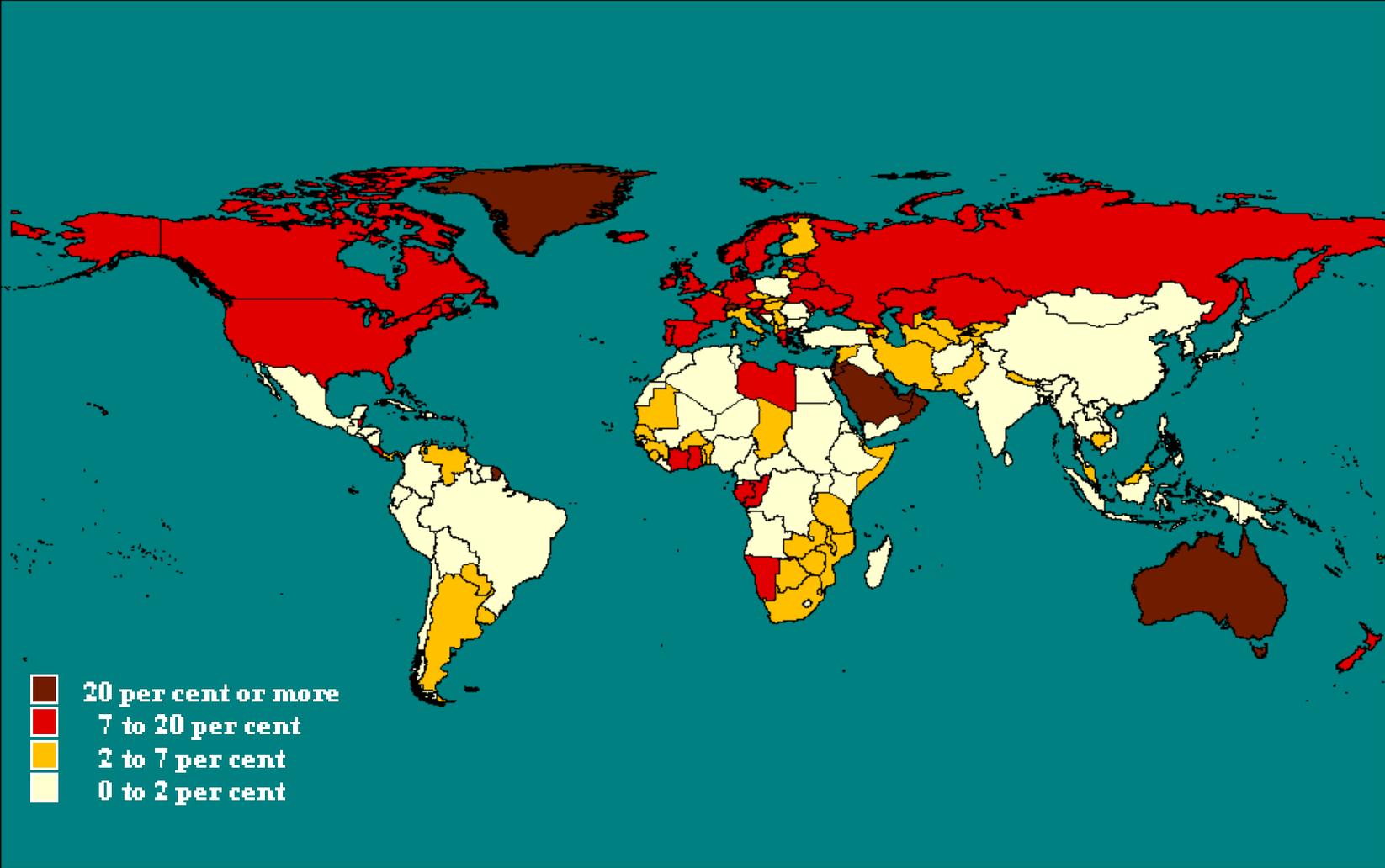
Richard E. Bilsborrow
University of North Carolina
Chapel Hill, NC
Richard_bilsborrow@unc.edu

Presented at Conference on Comparative
and Multi-sited Approaches to International
Migration
INED, Paris
December 12-14, 2012

Outline of talk

- 1. Rarity of international migrants
- 2. Why we need specialized surveys
- 3. Issues of survey design
- 4. Need to use specialized sampling techniques
- 5. Some examples: not always possible to get what you want
- 6. A note on analysis which has data collection implications

LIFETIME MIGRANTS AS PERCENTAGE OF POPULATION, 2005



Importance of international migration in the world, demographically

- UN (2009) estimated 214 million persons will live in a country other than that of their birth in 2010, 3.1% of the world population
- This compares with 75 million, 2.9% in 1960, so not a huge increase in percentage (not in this sense an “Age of Migration”)
- But the distribution has changed, now being much more concentrated in developed countries (9.6% of their population) than developing countries (1.4%)

Above is *lifetime* data and data on *stocks* of migrants measured as foreign-born: *flow data* show better how rare IM is in general: Mean annual in-migration **rate** during 2000-2005 was only 0.22% in MDCs and -0.05% in LDCs

International migration as a rare phenomenon

- Developed countries with highest inflows in 2000-2005 were Spain and Ireland at 0.97 and 0.98% per annum (under 1%)
- Others included Canada 0.7, Australia 0.5, USA 0.4, Germany 0.3, UK 0.2, Netherlands 0.2, France 0.1
- Countries with highest annual outflows: Armenia -0.7; Albania -0.6; Guatemala -0.5; Mexico, Ecuador, Iran -0.4
- There were few other demographically significant countries in the world (with over 5 million population) that had annual outflows over -0.1% (e.g., China, India -0.03%)
- During 2005-2010, only 3 countries in the world (with over 1 million population) had a net annual immigration rate as high as 1%, and only two a net emigration rate over 1% (Zimbabwe and Georgia).

Why do people migrate?

- Economic theories date from Adam Smith who saw workers moving from lower to higher wage areas
- Human capital model of individual migration of Sjaastad, modified by Todaro
- Explains age-education-skill selectivity of migrants
- Household model approach of Mincer, etc., followed
- New Economics of Labor Migration, sees migration as a *household* decision to diversity income sources, and thru remittances, gain liquidity and ensure against risk
- Non-economic factors also affect migration, including family ties, marriage, community of origin
- For international migration, policies of countries of origin and especially of destination add a new dimension

Censuses vs. Surveys as Sources of data on international migrants

What do we want, vs. what is
feasible and practical to collect?

Consider what is the main interest in data on international migrants-- is it to...?

- *Measure/count stocks or flows* of international migrants?
--of immigrants, *emigrants*, return migrants, or.....?
--over what time period (past 1, 2, 5, 10 ... years)?
- *Characterize migrants*: age, sex, education, work experience, economic assets owned, etc.?
- Collect data on *remittances* sent/received/both?
- Study *determinants and/or consequences* of international migration?
- If latter, a census (or any other source, such as a continuous population register, border/admission statistics, registers of foreigners, etc.) cannot collect the kind of data needed.

My own work has focused on developing countries of net emigration

- Data on individuals who have left (emigrated) from households can be obtained from household members remaining behind (proxy respondents)
- But limitations in depth and type of data that can be obtained from proxy respondents
- In addition, data on *whole* households who emigrated is usually not available, and normally obtainable only through a survey in the country/ies of *destination*
- This is a major limitation of a survey carried out only in a country of origin (also of census)

Migration is a socio-economic experiment

- Migrants, by moving, are subject to a “treatment” (usually self-administered!)
- To assess the effects of the “treatment”, one needs to compare migrants with an appropriate group of “not treated persons”

Key issue in surveys of migration: use of *appropriate comparison groups*

- To study either the determinants or consequences of migration, need data on *both* migrants and non-migrants
- To study determinants, need data on both the migrants and non-migrants who constitute the rest of the population “at risk of migration” in the country of origin – latter constitute the appropriate comparison group
- But when whole households migrate, need data on those households in the destination country; for them, appropriate comparison group is households in origin country that did not migrate

The determinants of migration—
what are the appropriate
comparison groups

Two options: Two surveys or one at origin

BEST OPTION 1

Survey at origin



PLUS

Survey at destination

OPTION 2

Survey at origin:
using proxy
respondents for
migrants

Proviso:

Migrants
have to
leave
someone
behind

Appropriate time reference is needed: when did they leave as well as why?



Also need to know situation of non-migrants in the past: at the average time when migrants left

Necessary information:

Situation when migrants left

To study the determinants, the situation at the time of the survey is irrelevant

In sum, to study migration's determinants, need data as follows:

- **In country of origin: survey households with and without recent emigrants**
- **In country of destination: survey recent immigrants originating in the particular origin country of interest**
- **Compare: situation of migrants at time of departure from origin country with situation of non-migrants at the average time of departure of migrants**

The consequences of migration

Contrary to standard practice--

- Comparing immigrants with non-migrants at destination does NOT provide evidence about the *consequences* of migration
- Instead, such studies allow the assessment of the adaptation or *assimilation* of migrants to the host society, but not about how migrants have benefited or not from migration

To study the consequences of migration

- To get comparable data on immigrants at destination and non-migrants at origin: contemporaneous surveys are needed in both origin and destination countries
- Alternative: a survey at origin asking “proxy respondents” about how emigrants are doing

Appropriate time reference when the question is: how have they fared?



Current situation of non-migrants

Current situation of migrants

To study consequences, the situation at the *time of the survey* should be the focus

Summary so far

- Ideally, surveys should be carried out in *several countries of origin and/or destination* to study either the causes or consequences of international migration
- Combining data on non-migrants from countries of origin A-C with data on migrants from A-C to a country of destination Z makes possible better analyses of the factors determining migration to Z
- Combining data on non-migrants from A with data on migrants from A to several countries of destination X-Z allows a fuller analysis of the consequences of migration for migrants who leave A

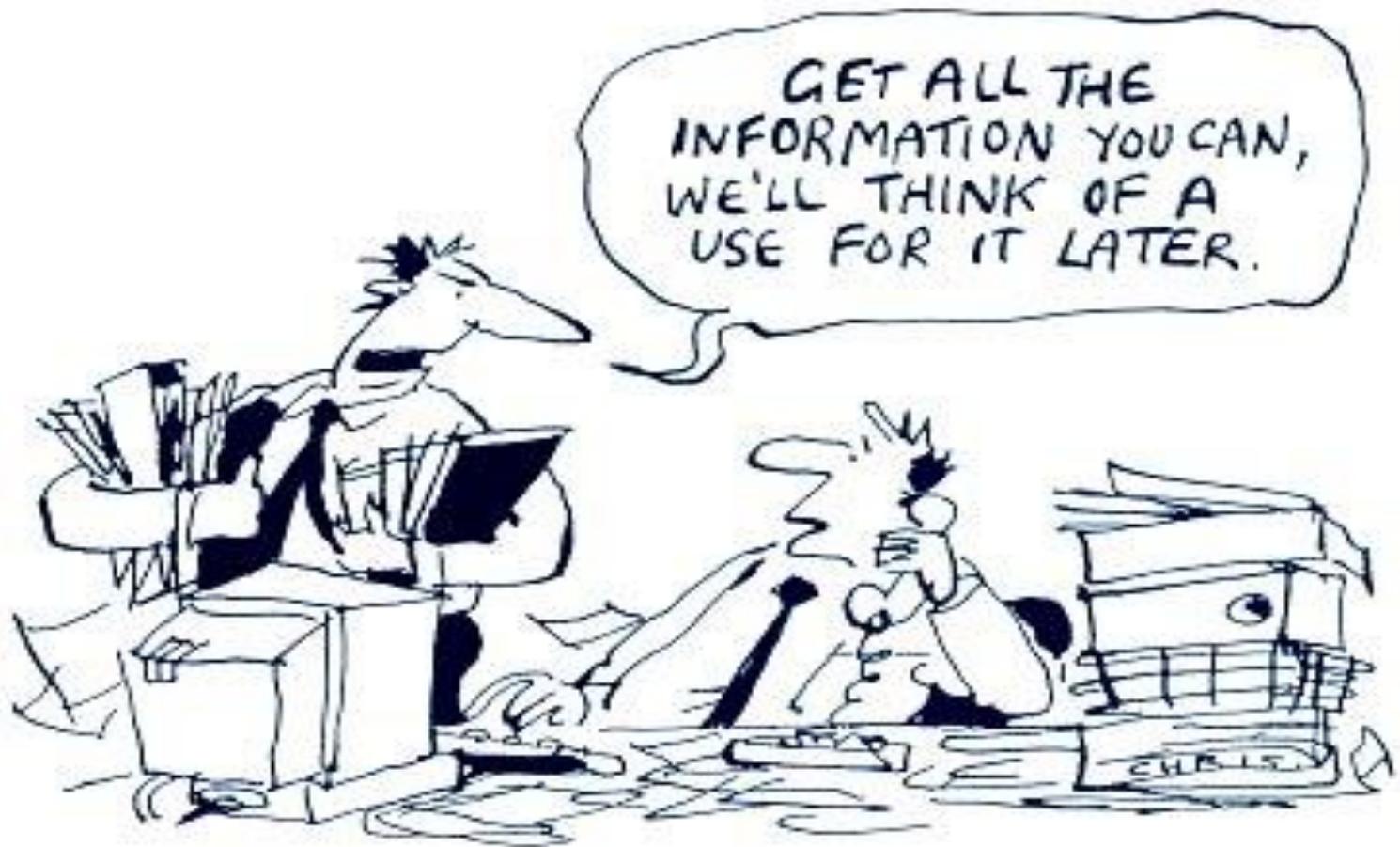
Implications for questionnaire design: big ignored methodological issue

- My way—collect detailed data on *last migration*: circumstances at time of migration
- Leads to detailed data on recent move, more reliable, but only one move
- Alternative: use truncated migration history to identify all moves of a sample since some time in past (e.g., 5 or 10 years ago), use events calendar perhaps, obtains less detailed data but more moves, higher statistical power
- Big question is which tells us more about why people migrate??

Need for multi-level modeling

- Of the determinants *or* consequences of migration....
- To control for the effects of context, which is needed to isolate the effects of individual/household factors, as well as provide information useful for policy.
- Hence it is usually desirable to administer a community-level survey linked to the household survey to collect contextual data.
- Sometimes other contextual data are already available or can be collected at reasonable cost.

A challenge: Deciding how the information will be used



Issues in sample design: sampling rare elements

Sampling Migrants in specialized surveys of international migration

- Absolute need for probability sampling
- Need to first define migrants of interest and survey purpose (e.g., study *determinants or consequences or both*); the purpose identifies what the appropriate comparison groups are (and country/ies)
 - E.g., in the NIDI surveys, the focus was on emigrants who left within the previous 10 years, requiring identifying households with one or more emigrants.
- Second, based on budgetary resources, define the survey domain and target sample size and geographic distribution
- Third, create a sampling frame (from the previous census?) that permits designing a sample that takes into account that households with recent emigrants (or immigrants) are *rare* elements, and concentrated in particular locations

1. Use of disproportionate sampling

- In the country of Origin, goal is to sample (select) households with emigrants and those without (and possibly a third group--those with return migrants).
- From the latest census or other source, form strata based on the expected prevalence of international migrants.
- Oversample areas or Primary Sampling Units (PSUs) from strata with higher proportions of households with emigrants at *each* sampling stage: This means selecting, e.g., provinces or other PSUs at the 1st stage using disproportionate sampling, then at the 2nd stage doing the same for selecting districts, etc., and finally at the last stage selecting Ultimate Area Units (UAUs), such as census sectors or (urban) blocks.
- Even highly disproportionate sampling fractions can be used, since that can be adjusted for in the analysis using weights.

2. Use two-phase sampling in last stage

- Once the final Ultimate Area Units (UAUs) have been selected, in each sample UAU, first conduct a listing or screening operation, to list all occupied households to identify those with and without emigrants
- Create separate lists for each type of element or household of interest, e.g., households with one or more former members who emigrated and did not return in the previous 12 years, those without such a person (including those who might have had a member leave more than 12 years ago), and those with return migrants
- Sample from each list separately, taking higher proportions from the lists of households with migrants and return migrants
- In phase 2, conduct interviews of sample households from both lists

- Some quick summaries of examples from my personal experience

Survey of Colombian migrants in Ecuador in 2006

- Funded by UNHCR
- Used 2002 Ecuador population census as sampling frame, prevalence of Colombians in Ecuador in census sectors in 5 main provinces (in northern Ecuador)
- But found far fewer Colombians than expected in 4 of 6 provinces—Why?
- Added snowball component, it was also not successful---why?

MED-HIMS Surveys

- We proposed to use the specialized sampling techniques appropriate for rare populations, as described above whenever possible
- In Egypt, the first country where the survey has begun, there was no adequate source of data on the prevalence of international migrants to create strata to use oversampling, so a simple, self-weighting PPES sample was used
- On the other hand, in Jordan, there were two sources of data, which, albeit also deficient, could be examined to see if they could be used.

What to do if have 2 different, dubious sources of data? Example of new sample in Jordan for 2013

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		

Example of oversampling PSUs: Jordan, 2013

Stratum	Number in stratum, N_h	Mean proportion international migrants	Proportionate allocation	(1) X (2)	Disproportionate Allocation A	Disproportionate Allocation B
	(1)	(2)	(3)	(4)	(5)	(6)
High	14	0.032	6	13	10	14
Medium	25	0.016	9	11	10	14
Low	50	0.004	15	6	10	2
Total	89		30	30	30	30

Some multi-country survey projects on international migration

- Netherlands Interdisciplinary Demographic Institute (NIDI)
- Latin America Migration Project of Massey and Princeton University, et al
- MAFE
- MEDSTAT-MEDHIMS in eight countries of Middle East-North Africa
- World Bank: Sub-Saharan Africa in 2009-10, and hopefully CIS States starting 2013

NiDi 7-country “Push-Pulls Project” on Causes of Migration to the EU, 1997-98

- 5 countries of origin, Turkey, Egypt, Morocco, Senegal and Ghana
- Two of destination, Italy and Spain
- Defined migrant households in origin country as those with member who left to live abroad without returning in the previous 10 years
- In Italy and Spain, defined migrant household as containing someone who had come from one of two specific origin countries in past 10 years
- Collected data for appropriate comparison groups in countries of origin & destination

NIDI-Eurostat Survey Project, 1997-1998

- Common methodology not only in defining migrants but in design of samples, questionnaires
- Sample sizes range from 1100-2000 in origin countries, but required screening 10 X these numbers
- Migrants still rare elements, so used special sampling methods, including stratification, oversampling PSUs, and using 2-phase sampling in last stage
- Valuable data sets not fully analyzed even now, a gold mine!
- E.g., Ghanian non-migrants in Ghana and Ghanian migrants in Italy

On a steep learning curve now?

- Conference on Hard to Reach (H2R) Populations in Nov. in US—new methods especially RDS and venue-based sampling
- So what we can learn in this conference not only substantively but in terms of migration survey methodology from the MAFE experience, and hopefully in the future from the CIS and other projects.

Why is the state of knowledge weak regarding international migration?

- 1. the complexity of the phenomenon (including definitional problems, lack of an accepted clear/universal theoretical framework (each social science focuses on variables of special interest to it));
- 2. it involves two countries;
- 3. international migrants are rare elements, so specialized sampling methods are desirable;
- 3. 1-3 complicate data collection and increase costs.
- 4. As a result, there are few good data sets, which ideally should involve coordinated data collection in multiple countries of origin and/or destination.
- 5. Hence there are few good detailed quantitative studies of the determinants or consequences of IM. Back to title.

A world migration survey?

- We almost had one once—
- But on internal migration and development.
- Now there is much more interest in (concern about?) international migration,
- So
- Why not a world survey on international migration?