



Reaching hard-to-reach migrant populations using different methodologies: the Italian experience



Livia Elisa Ortensi

University of Milan – Bicocca, Italy

European Centre for Social Welfare Policy and Research, November 7th, 2018

Outline of the presentation

- Reaching hard to reach population: theoretical background
- The Center Sampling (CS) Method:
- The experience of the Regional Observatory on Integration and Multiethnicity of Lombardy (2001-2018)
- Surveying Female Genital Mutilation in Europe (2014-2017)
- Collecting data among street-based sex-workers (2006-2014)



Reaching hard to reach population theoretical background



Household-based survey designs are notoriously ineffective in studying hard-to-reach groups (e.g. irregular migrants, Roma...)

Two main issues:

- How to contact them:
In most demographic studies based on household surveys such marginalized population segments are lost by definition or, at best, are grossly underrepresented
- How to interact and collect information from them:
An additional set of problems concerns the willingness of individuals from marginalized, hard-to-reach groups to provide information to survey interviewers, especially about the type of sensitive matters that define their marginality: it can be argued that the same social characteristics and constraints that hinder access to these individuals may also impair their willingness or ability to answer survey questions (Agadjanian & Zotoya, 2012)

Reaching hard to reach populations: theoretical background



Concrete problems are:

- Lack of sampling frames: the use of population lists for sampling is considered a gold standard for obtaining representative probabilistic samples of a given target population. However, quite commonly, adequate sampling are unavailable (Reichel and Morales, 2017).
 - registers of the resident population may be inaccurate or outdated (migrants are more mobile than natives), biased or simply not in place
 - The necessary information for identifying a specific sub-population may not be available from the existing registers
- Difficulties in contacting & talking with people:
 - Gatekeepers
 - Mistrust, diffidence
 - Long and unusual working hours
 - Lack of language skills (both of interviewers or of interviewees)

The Italian region of Lombardy: a long experience of local surveys dedicated to migrants

A bit of history....

The first attempts to carry out sample surveys date back to the beginning of the '90s

The mission was:

- to increase the knowledge of the phenomenon of migration in Italy

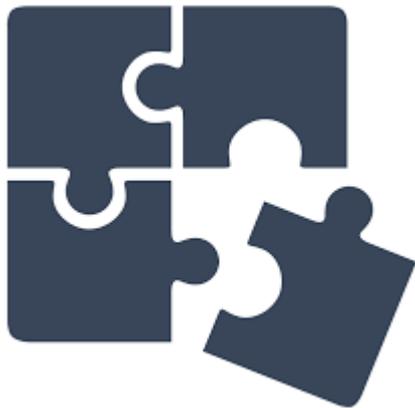
&

- to produce data based on a robust methodology



The context where the activity of what would later become the Regional Observatory for Integration and Multiethnicity (ORIM) was characterized by:

- scarcity of official data
- poor knowledge of what was, at that time, a new phenomenon for Italy
- high presence of undocumented migrants: even in case of good data quality (?) we were missing a piece of the puzzle



For these reasons early surveys aimed primarily at:

- Estimating the magnitude of the phenomenon, especially at assessing the proportion of undocumented migrants
- Exploring and describing its baseline characteristics: main sending countries, gender balance, age, integration in the job market...
- Analyzing reasons for migration & describe the first phases of migration



Starting point: challenges in surveying immigrants

- Unavailability of adequate sampling frames. Relevant particularly if the reference universe includes all the immigrants (irrespective on their juridical status).

The real problem becomes: how to select (at random, as requested by probabilistic samples) and to contact the sample units?



By CS method we can imagine that:

- the universe of foreign citizens is made up of a list of H statistical units
- each one by necessity keeps a set of contacts with some non-residential locations or gathering places located in the area called “centres”
- after extensive ethnographic research we know the main centres and we can gather them by typologies

Once a sufficiently wide and heterogeneous set of ‘centres’ is identified, the universe of foreign citizens, whose nominative list is not available, can be formally described by the following table:

List of units (unknown)

-----List of centres (known)-----

Sequence	Names	List of centres possibly attended						
		Centre 1	Centre 2	Centre 3	Centrek-1	Centre k
1	a	1	0	0	0	1
2	b	0	0	1	0	0
...
i	1	0	1	0
...
H-1	w	0	1	1	0	0
H	z	1	1	0	1	1
		Tot. H(1)	Tot. H(2)	Tot. H(3)	Tot. H(k-1)	Tot. (k)

In each column the value is 1 if the subject usually visit that centre, and 0 otherwise (we can also consider “how much time” is spent in each centre. In this case the attendance can be formally expressed by a value $0 \leq X \leq 1$). It follows that the total of a given column identifies the number of individuals (among the H constituting the universe of reference) visiting that centre.

This means that, instead of selecting n sample units by rows (i.e. n names from the unknown list) we can: *a)* select n columns/centres (known) and then *b)* choose randomly h individuals among those regularly visiting the selected centres.

According to this assumption the preliminary step is to identify all (or a sufficiently large set of) the centres located in the chosen territory and visited by the migrants

The challenge of this method is to undertake adequately the ethnographic study to find and list the universe of relevant locations that cover the target population.

The second step is to choose the individuals at random in this center

Let us assume that in the chosen territory there are k centres visited by the migrants and these centres are of different size.

- *the number of interviews in a certain centre depends on its size*
- *If the centre is considered to be small, a small number of interviewees will be chosen. On the contrary, the bigger the centre, the more migrants visit it, the more individuals will be interviewed.*

Later, the interviewees (chosen individuals) are asked to answer to a set of questions (questionnaire) concerning her/his structural characteristics, both individual and family ones, as for example: gender, age, marriage status, citizenship, legal status, education, religion, housing conditions, job position, remittances, etc.

They are also asked which of the k centres (indicated on a specific annex to the questionnaire) they normally visit.



OSSERVATORIO REGIONALE PER L'INTEGRAZIONE E LA MULTIETNICITA'.
Questionario di rilevazione. Anno 2016

A. Numero Questionario:.....|_|_|_|_|

B. Comune di rilevazione:.....Cod. Istat |_|_|_|_|_|

C. Intervistatore:.....Cod. |_|_|

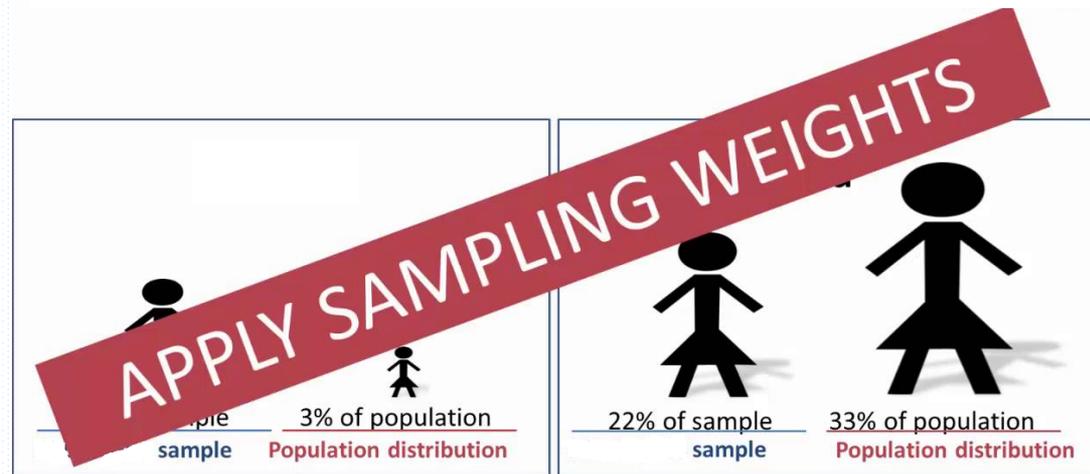
D. Luogo di rilevazione:.....Cod. |_|_|_|
(scrivere il nome/tipo di luogo in cui si sta svolgendo l'intervista e il rispettivo codice come da elenco in E)

Qcontrol1 <i>[INT.: inserire data: __/__/2015]</i>	Qcontrol2 <i>[INT.: inserire ora di inizio: __: __]</i>
--	---

E. Quali luoghi/centri sul territorio frequenta in questo periodo?

01. Centri che offrono servizi e assistenza (accoglienza, lavoro, sanità, centri di ascolto, mense, uffici pubblici...)[...]**01**
02. Centri di formazione (corsi di italiano, corsi di formazione professionale, CTP, scuole, Università...).....[...]**02**
03. Luoghi di culto (chiese, moschee, templi..).....[...]**03**
04. Negozi etnici (Kebab, macellerie islamiche, take-away, prodotti alimentari...).....[...]**04**
05. Luoghi di svago (cinema, discoteche, strutture sportive, bar, ristoranti..)[...]**05**
06. Centri commerciali.....[...]**06**
07. Ritrovi, luoghi di incontro all'aperto (stazioni, piazze, parchi, laghi..)[...]**07**
08. Mercati in genere (mercati comunali, mercato dei fiori, ortofrutticolo...).....[...]**08**
09. Luoghi di lavoro o di reclutamento forza lavoro (cantieri, laboratori tessili, ristoranti e alberghi, portinerie; campi agricoli e allevamenti...).....[...]**09**
10. Associazioni e centri culturali[...]**10**
11. Centri servizi (phone center, agenzie per il trasferimento di denaro, lavanderie automatiche..).....[...]**11**
12. Abitazione privata (feste private, ecc.)[...]**12**

The sample that we collect by CS technique is potentially biased so sample weights are associated with each sample unit.



The interviewees are given:

- a profile according to the centres they visit (all the individuals who visit the same centres are given the same profiles).

- an individual probability of inclusion in the sample has been determined as dependent:

- 1) directly on the number of selected centres the person really visits; and
- 2) inversely on the number of individuals from the population who visit that centre.

In other words:

the more centres any individual in the universe visits, the larger the inclusion probability of being interviewed will be.

Consequently, if drawn into the sample, he will be associated *ex-post* with a lower weight.

But, the *ex-post* weights also depend on the number of individuals who visit those centres.

The larger and more visited the centre is, the smaller the inclusion probability of the person is, and therefore the value of the weight for this individual is higher.

It can be shown that by the adoption of these weights the sample that comes out by CS technique can be considered as representative of the whole universe and fully comparable to a hypothetical traditional simple random sample for which, in the contrary the (generally unknown) list of units is strictly required.



To calculate weights:

- Baio G., Blangiardo G.C., Blangiardo M., (2011) Centre sampling technique in foreign migration surveys: a methodological note, *Journal of Official Statistics*, vol.27, 3, 2011, pp.451-465
- A Pearl code is available (Blangiardo)
- An R code is available (Reichel)

To return to our history..

Surveys using CS method were implemented first in some cities and metropolitan areas of Lombardy



Milan (1991,1992, 1996), Monza (1992), Brescia (1993)

Then in provinces of Lombardy



Milan (1997-2000), Lodi (1999), Mantua (2000),
Lecco (2000), Varese (2000), Cremona (2000)

Target: individuals with a foreign background from main sending countries aged 14 and over

And finally in the entire region of Lombardy
(ORIM – Regional Observatory on Multiethnicity and Integration)



Year	2001	2002	2003	2004	2005	2006	2007	2008	
sample size	8,000	8,000	8,000	8,000	8,000	9,000	9,000	9,000	
Year	2009	2010	2011	2012	2013	2014	2015	2016	2017
sample size	9,000	8,000	8,000	7,000	3,500	3,500	3,500	3,500	1,900

Other surveys using CS:



Provinces outside Lombardy:

Alessandria (2008), Biella (2006), Cuneo (2007),
Venice (2007)

Specific National projects:



- **Effect of amnesties in southern Italy (2005):** Progetto Sud /30 provinces in Southern Italy and 10 provinces in Centre –North Italy & 30,000 interviews
- **Integration (2008/09):** Integrometro/ 33 local areas & 12,000 interviews
- **Job trajectories:** Per.La (2009) 18 local areas (provinces) & 13,000 interviews

International projects:

- **Localmultidem** - *Multicultural Democracy and Immigrants' Social Capital in Europe*. Project funded by the EU Commission within the 6th framework programme (data for 9 cities: Barcelona, Budapest, Geneva, London, Lyon, Madrid, Milan, Stockholm and Zurich).

<https://www.unige.ch/sciences-societe/incite/welcome-to-the-incite-website/data/1/>

- **ICS** - *Immigrant Citizens Survey* (data for 14 cities in Belgium, France, Germany, Hungary, Italy, Portugal, and Spain) <http://www.immigrantsurvey.org>



Subject and territorial reference	No. of sample units	Year	Research Institute
Local areas			
Foreign migrants living in Metropolitan area of Milan	500	1991 - 1992	IReR-OETAMM
Foreign migrants living in the municipality of Monza	200	1992	IReR-OETAMM
Foreign migrants living in the municipality of Brescia	300	1993	IReR-OETAMM
Foreign migrants living in Metropolitan area of Milan	1000	1996	ISMU
Foreign migrants living in the province of Milan	2,000 per year	1997-2000	ISMU
Foreign migrants living in the province of Lodi	500 per year	1999 & 2001	ISMU
Foreign migrants living in the province of Mantua	500 per year	2000 & 2001	ISMU
Foreign migrants living in the province of Lecco	500 per year	2000 & 2001	ISMU
Foreign migrants living in the province of Varese	500	2000	ISMU
Foreign migrants living in the province of Cremona	500	2000	ISMU
Foreign migrants living in the province of Biella	500	2006	Provincial Observatory
Foreign migrants living in the province of Cuneo	1,000	2007	Provincial Observatory
Foreign migrants living in the province of Venice	800	2007	Provincial Observatory
Foreign migrants living in the province of Alessandria	540	2008	
“Integration indicators” on migrants living in Italy Milan, Bologna, Ancon, Turin, Rome, Latina, Naples	3000	1993-1994	National Academic Research Group on foreign migrants living in Italy
Lombardy Region			
Foreign migrants living in Lombardy region	8,000 per year	2001-2005, 2010-2011	ISMU
	9,000 per year	2006-2009	ISMU
	7,000	2012	ISMU
	3,500/4,000 per year	2013-2016	ISMU
	1900	2017	ISMU
Specific subgroups			
Egyptians, Filipinos, Ecuadorians living in the municipality of Milan	900	2007	Eurostat and Department of Sociology of the University of Trento
Egyptians and Ghanaians living in 5 Italian municipality Milan, Rome, Caserta, Modena, Vicenza	1,000	1997	
“Insertion in the labour market, income and remittances sent by foreign immigrants in Italy” Migrants coming from the former Yugoslavia, Albania, Poland, Romania and Morocco and residing in Rome and in some provinces of the Veneto and of Campania regions	1,920	1997/1998	Univesità Napoli Federico II
National level			
30 provinces in Southern Italy and 10 provinces in Centre – Nord Italy “2003 regularization effects”.	30,000	2005	ISMU
“Integrometro Survey” National sample of 33 local areas (province, municipality)	12,000	2008/2009	ISMU
PER.LA Survey National sample of 18 local areas (provinces)	13,000	2009	ISMU, CENSIS and IPRS
International level			
“Immigrant Citizens Survey” for Italy, Portugal and Hungary	3,000	2011	King Boudain Foundation and Migration Policy Group

ORIM surveys – Information collected

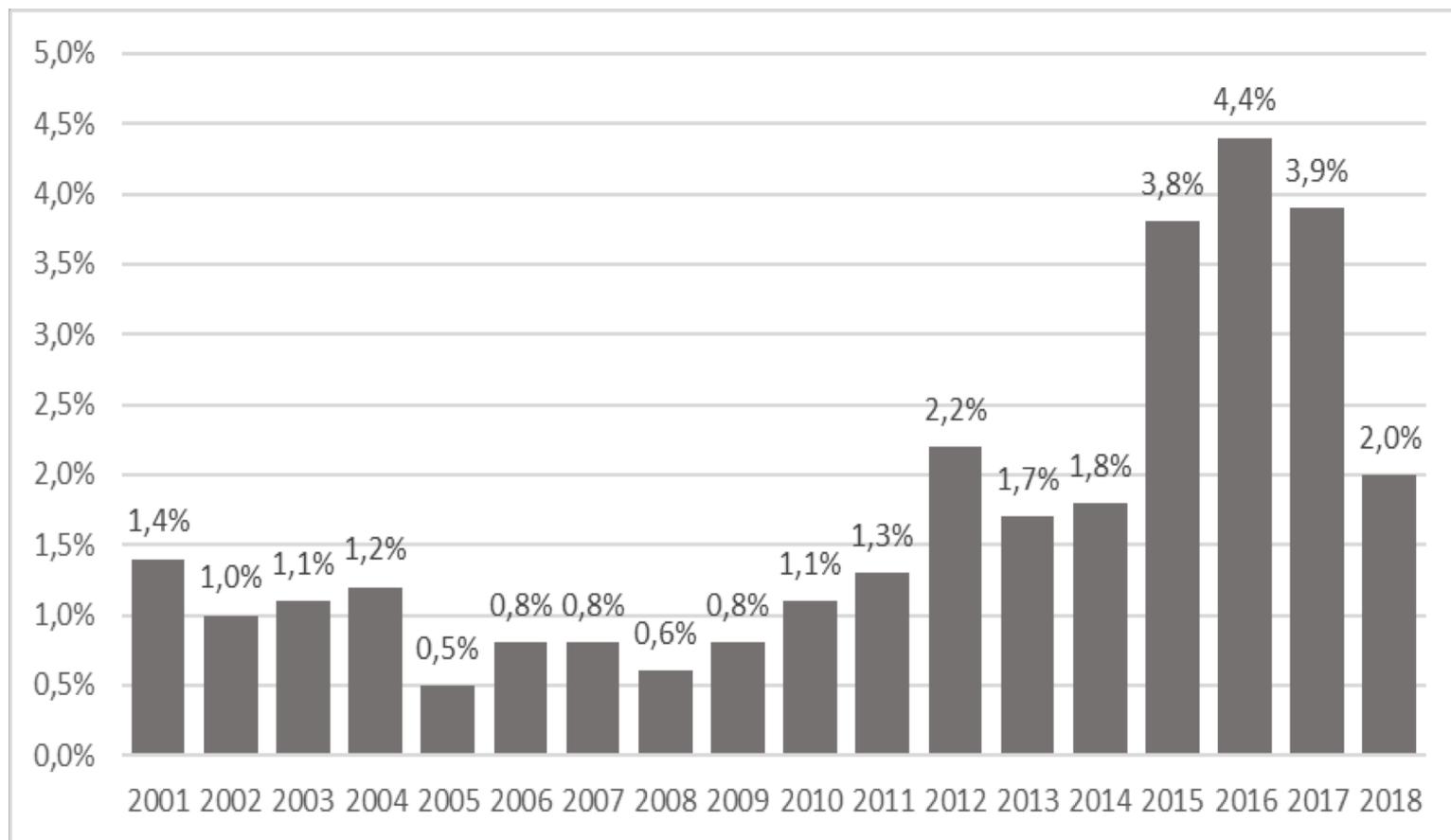
A set of baseline information is collected every year: e.g. gender, age, year of arrival, citizenship, legal status, marriage status, number of children, education, job position, wage...

Then other topics are covered in different years according to issues of interest:

- At the beginning: origin, work, reason for migration, participation in amnesties
- As migration became a well established phenomenon: health, remittances, opinions, integration proxies, naturalization
- In more recent years: intention about onward/return migration, ways of facing the crisis , poverty indicators

CS sensitivity to small subpopulation dynamics (1)

Asylum Seekers and Refugees/holders of humanitarian protection permits in Lombardy 2001-2018.



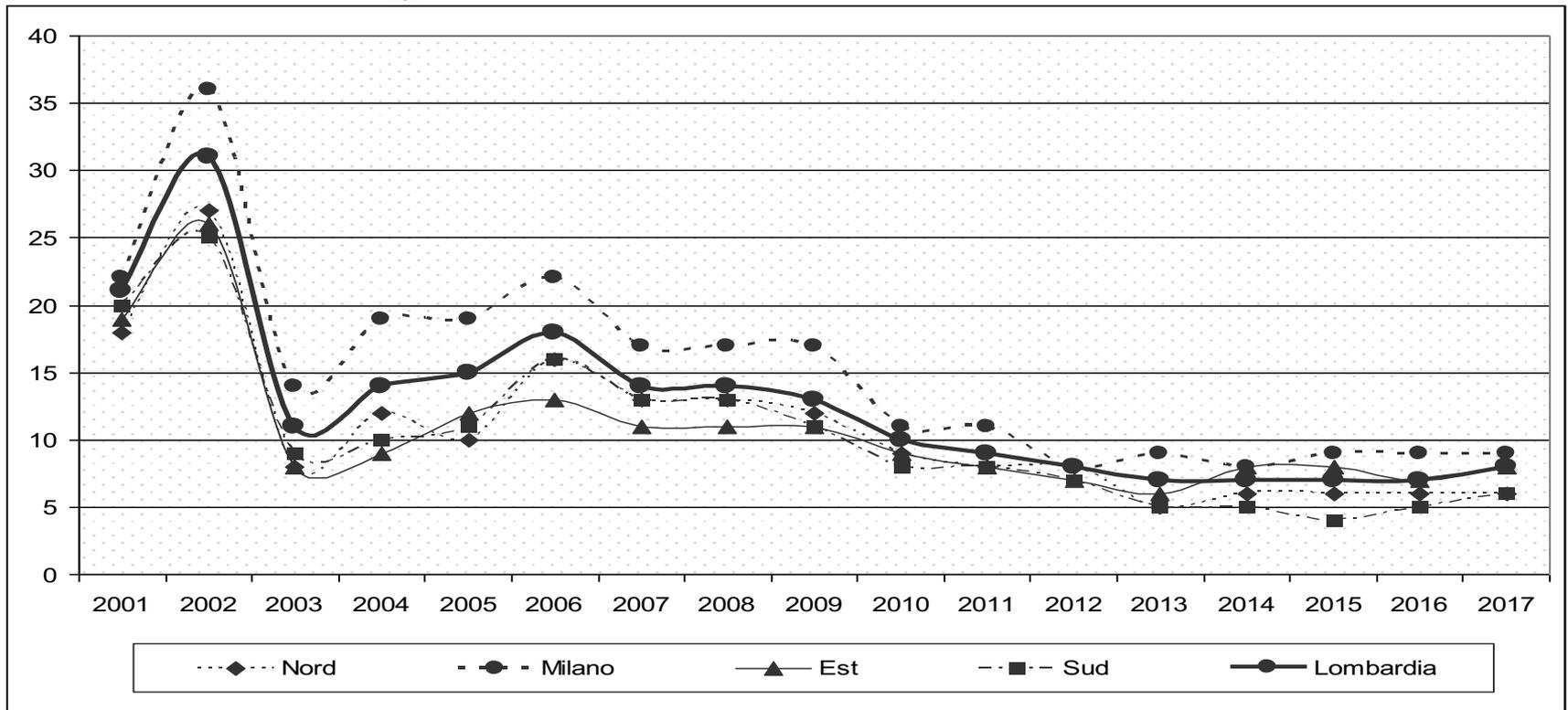
CS sensitivity to small subpopulation dynamics (2)

Events affecting the proportion of undocumented migrants



Year	Event	People involved
2002	Bossi-Fini law	647,000
2006	Quota system	170,000
2006	Quota system/2	350,000
2007	EU enlargement	444,000
2009	Security law	300,000
2012	Legislative Decree	99,000

Proportion of undocumented migrants in Lombardy, by areas





Towards a better estimation of prevalence of female genital mutilation in the European Union 2014-2017 (FGM-Prev; funded under the Daphne programme) PI Els Leye

An (even) more difficult challenge compared to ORIM:

- Only selected nationalities (quota)
- Only female interviewees
- A very difficult topic to approach (Female Genital Mutilation is forbidden in Europe)
- Fieldwork in two countries: initially it was planned France & Italy, later Belgium and Italy

Methodology:

- Initially strong emphasis on robust statistical methodologies: the first attempt was to test a combination of TLS (Time location sampling also known as time-space sampling) and RDS (Respondent Driven Sampling)

Partial failure:

- In the initial phases (pretest) of the survey it was soon evident that the two methods were too rigid and in the end it was impossible to apply them in our fieldwork



Location needed to be selected according to the aim of the study and to the target population (e.g. hospitals with cultural mediators of selected nationalities, communities meeting points...)

Time location sampling (TLS)

- It requires the casual extraction of the time & place where an interviewer should visit the venue
- It requires a very detailed list of venues
- It implies a considerable amount of time in extracting and replacing venues that are not accessible or are not accessible in the hour selected

Respondent-driven sampling (RDS), combines “snowball sampling” with a mathematical model that weighs the sample to compensate for the fact that the sample was collected in a non-random way.

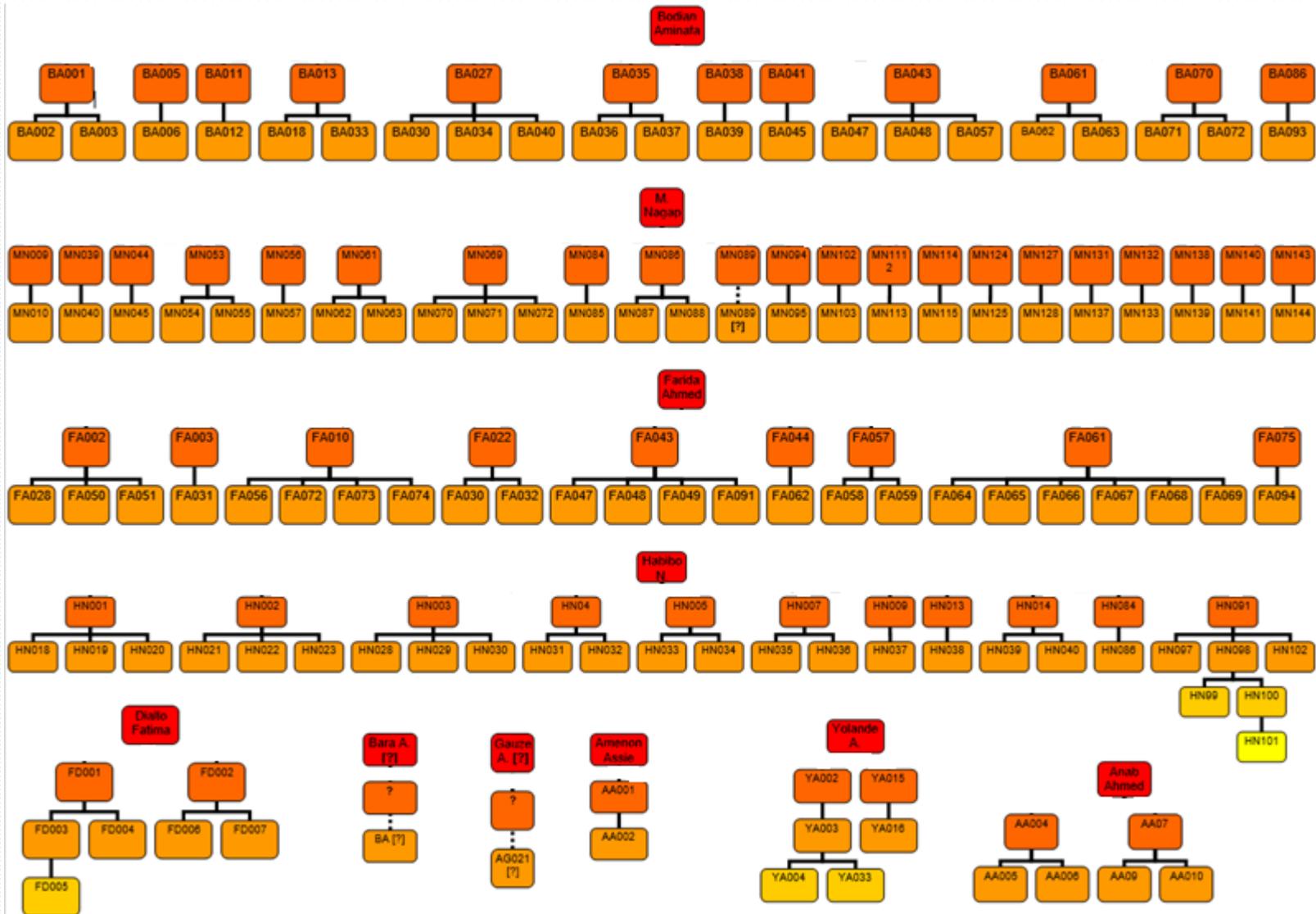
- It requires specific questions in the questionnaire and the use of coupons to link respondents that proved very difficult to handle for the interviewers

In Italy the survey was successful (1,400 interviews): what worked was **a combination of facility based sampling and snowball sampling**:



- As for the center sampling we collected a list of places where we could meet women but we let the interviewer free of suggesting and choosing some of them autonomously
- After a women was interviewed she was asked if she knew some other women that could be interviewed
- The number of interviews by nationality was determined according in two stages, according to technical consideration on the prevalence that was to be estimated at the nationality level.
- The idea was to compensate the loss of statistical robustness by trying to intercept different networks of women (different ethnicities, and networks)

Snowball links (8%)



Methodological support to the fieldwork

Dynamic allocation of the number of interviews:

	<i>Interview 1 step</i>	<i>reached</i>	<i>prevalence on- the-field according to the first step result</i>	<i>Minimum to be reached with second step</i>	<i>Lowest number reached in the first step</i>	<i>Additional minimum interviews second step</i>	<i>Final</i>	<i>Minimum reached</i>
Egypt	50	yes	43,9%	95	no	45	306	yes
Nigeria	72	yes	70,3%	80	no	9	167	yes
Senegal	71	yes	35,5%	88	no	16	142	yes
Cote d'Ivoire	91	yes	13,6%	45	yes		115	yes
Burkina Faso	70	yes	74,7%	73	no	2	244	yes
Eritrea	54	yes	52,7%	96	no	42	134	yes
Ethiopia	73	no		0	no		20	no
Somalia	50	yes	67,5%	84	no	34	131	yes
Cameroon	50	yes	17,5%	55	no	5	65	yes
Ghana	50	yes	9,0%	31	yes		54	yes
Total	632			648			1378	

So far we have only focused on **methodologies** trying to understand how to work without sampling frames

Possibilities are:

- to use aggregation centres
- To use networks of interviewees and interviewers

However another big issue on the field is

- How to speak with interviewees avoiding their refuse to be interviewed
- How to be accepted in centres

In both ORIM and FGM-Prev what proved to be crucial was:

- A field coordinator with a **specific knowledge** and solid experience of migration surveys
- **Building trust:** people who are responsible of centres might not let our interviewers in (gatekeepers). They need to know how we will use data, they need to trust our institution (formal centres) or trust the interviewer (informal centres)
- A consolidated network of trained **interviewers from immigrant communities** (insiders). Most of them are cultural mediators or work with migrants (help desks, NGOs).
 - Possibility to interview in native languages
 - Possibility to establish a relation based on trust with the interviewee (we also ask sensitive questions)
 - The possibility to have access to informal venues (parties, women meetings, weddings, informal aggregation centers) and avoid gate-keeping, thanks to the reputation/citizenship of the interviewers (insiders)

In Belgium the survey was not successful

- Restricted time on the field & no previous experience with immigrants surveys
- Not enough time to recruit the interviewers, or to contact the people responsible of the centres and win their trust (many centres refused to collaborate): this phase requires a lot of time and efforts

Earning the access to centres and building an outstanding team of interviewers is key

Data collection on street based sex workers

- Very visible population but stigmatized
- Difficulty in getting in touch with sex workers/earning trust

Solution:

- Here we relied entirely on NGOs (insiders, not part of the community but having the trust of the target population)
- The idea: using data routinely collected by different NGOs (no additional expenses for the research)
 - Inviting all NGOs working on the territory & offering a service of support on the road (“Street Units”)
 - Proposing them to harmonize their data collection tool in order to build a common database (usually they all collect some information of some kind)
 - Using **capture-recapture methodology** to produce unbiased estimations of population size
 - We produce data & analyses for the NGOs as a compensation for their work

Capture Recapture Method

Single Recapture

Notation:

- X - initial sample size captured and marked
- y - second sample size recaptured independently
- x - number of sample in the recaptured one that are marked
- τ - total population size

Question: How do we estimate the total population size?

Since the proportion of the marked subjects in the recaptured sample is likely to be about the same as the first sample in the whole population:

$$\frac{x}{y} \cong \frac{X}{\tau} \qquad \hat{\tau} = \frac{y}{x} \cdot X$$

An estimate of the variance of $\hat{\tau}$ is:

$$\hat{V}ar(\hat{\tau}) = \frac{Xy(X-x)(y-x)}{x^3}$$

An approximate $100(1 - \alpha)$ % confidence interval is:

$$\hat{\tau} \pm z\sqrt{\hat{V}ar(\hat{\tau})}$$

Chao, A. (2015). Capture-Recapture for Human Populations. *Wiley StatsRef: Statistics Reference Online*. 1–16. DOI:

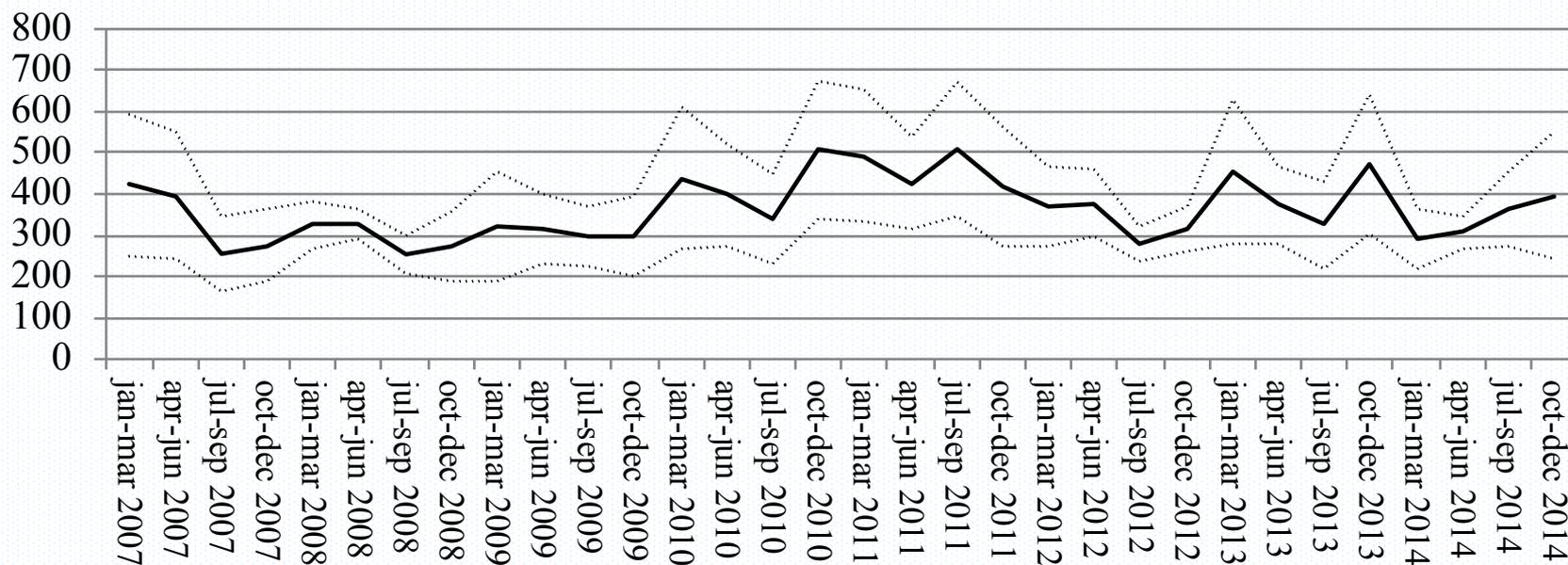
Rivest, L-P., Baillargeon, S. (2007). Rcapture: Loglinear models for Capture-Recapture in R. *Journal of Statistical Software*. 19(5), 1-31.

- Each NGO goes in the streets with a camper/van one or twice a week offering support (coffee, condoms, medical assistance, information on laws and projects supporting those who want to leave the sex work)
- Each NGO covers a part of the city (no overlapping)
- Each women is given an ID (the first given name is usually collected) and each time the women is met some information are collected (multiple “captures” for each ID)
- If the sex worker has met another street unit the information is collected, trying to link different part of his/her story
- Information collected: gender, year of birth, nationality, year of arrival, number of children, previous work in another territory, use of services offered by the Street Unit & an evaluation given by an operator about the sex worker exploitation status (i.e. is he/she exploited by someone or he/she works autonomously?)



Output

- An estimation of the number of street based sex workers in a territory continuously updated
- Basic information about the characteristics of the population
- Evaluation of flows, turnover, % of self declared minors, etc..



Results (2)

Year	Number of female street-based sex workers met by the NGOs	Estimate of the number of female street-based sex workers who were ever present in the city of Milan	Mean number of women estimated for each 3-month period	Number of female street-based sex workers for every 10,000 inhabitants	Number of female street-based sex workers for every 10,000 female inhabitants aged 15–49	Total number of female sex workers (including indoor-based)*
2007	366	563 (498–628)	338	4.3	19.3	938
2008	453	566 (522–610)	295	4.4	19.6	943
2009	382	572 (508–636)	307	4.4	19.8	953
2010	414	585 (535–635)	420	4.6	20.0	975
2011	446	609 (562–656)	461	4.9	21.5	1,015
2012	522	668 (624–712)	337	5.3	24.5	1,113
2013	416	560 (516–604)	408	4.3	20.2	933
2014	428	574 (524–624)	340	4.3	20.0	957

Limitations

- Different NGOs are not always willing to work together and share data, working on cooperation and mutual trust is important
- NGOs' workers and volunteers are often not interested in collecting high quality data, they see this activity as a useless burden: they have to be trained and sensitized about data usefulness
- Tablet or laptop should be available along with a common interface
- Data have to be processed very carefully by researchers looking for mistakes and overlaps
- Some part of cities where street workers work might not be covered, patterns change fast

Conclusions & take-home messages

- Surveying hard to reach population is possible
- As all surveys using “non classical statistical samples” loss of statistic robustness might occur. Researchers should be aware of limitations of their data and results
- Field preparation and choice of interviewers are key points: especially if we are exploring a new issue a lot of time has to be allocated to this two aspects
- The use of insiders or people of trust is important
- Sometimes raw data on hard to reach population exists but are not used

References:

- Baio G., Blangiardo G.C., Blangiardo M., (2011) Centre sampling technique in foreign migration surveys: a methodological note, *Journal of Official Statistics*, vol.27, 3, 2011, pp.451-465
- Blangiardo, G. C. (2008): The centre sampling technique in surveys on foreign migrants. The balance of a multi-year experience, *Working paper 12, prepared for the UNECE/Eurostat Work Session on Migration Statistics*.
- Reichel D., Morales L., Surveying immigrants without sampling frames – evaluating the success of alternative field methods *Comparative Migration Studies* 2017 5:1
- Serrano Sanguilinda I., Barbiano di Belgiojoso E., González Ferrer A., Rimoldi S.M.L., Blangiardo G.C., Surveying immigrants in Southern Europe: Spanish and Italian strategies in comparative perspective *Comparative Migration Studies* 2017 5:17
- Leye E, De Schrijver L, Van Baelen L, Andro A, Lesclingand M, Ortensi LE, Farina P (2017) Estimating FGM prevalence in Europe. Findings of a pilot study. Research report. ICRH: Ghent.
- Chao, A. (2015). Capture-Recapture for Human Populations. *Wiley StatsRef: Statistics Reference Online*. 1–16. DOI:
- Rivest, L-P., Baillargeon, S. (2007). Rcapture: Loglinear models for Capture-Recapture in R. *Journal of Statistical Software*. 19(5), 1-31.

