SAMPLING IN HOUSEHOLD AND PERSON SURVEYS

Tzahi Makovky
Deputy Senior Director,
ICBS Statistical Methodology Department

EU Twinning Project – Component D

Subjects

- Principles of sampling in ICBS
- Sampling in LFS & HES
- Sampling in Social Survey
- Future planning in Household surveys

Principles of sampling in ICBS

- Probability sampling
- Sampling frame – with minimal under-coverage and over-coverage; sampling unit properly defined ;variables necessary for sample design; variables to trace the sample
- Efficient sample design relating the main purposes of the survey
- Take into consideration constrains

Household and Person surveys

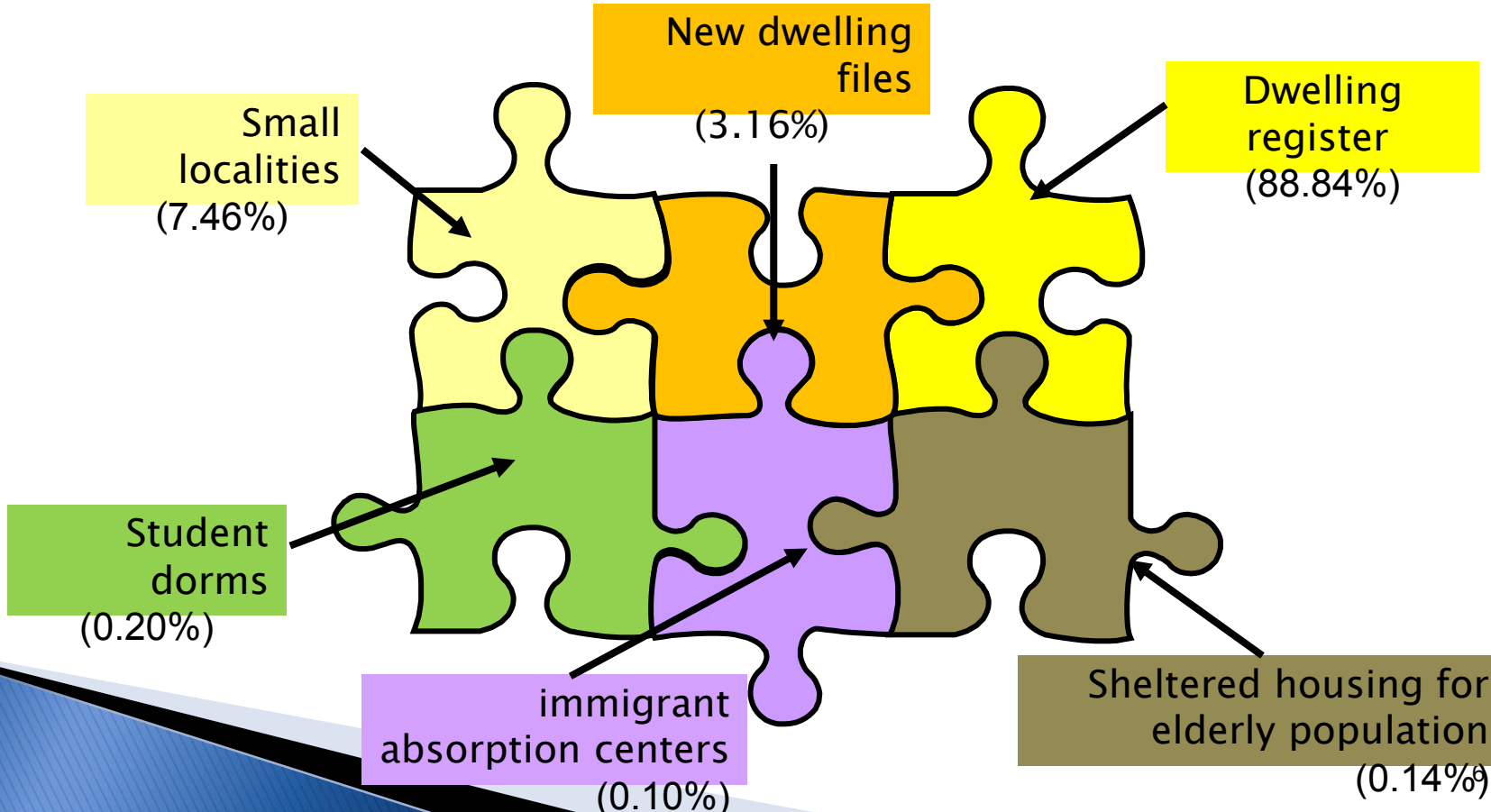
- In the past ICBS thought that he could place persons into dwelling by combining population register and dwelling register to a single register
- It was impossible to implement
- The conclusion was separate sampling frames for households and persons survey

Sample design in Household surveys – General view

- Two stage sampling design – reflected by field constrains
- First stage – sampling of localities. The method is PPS. ≈130 biggest localities (≈95 in HES) are selected in certainty. ≈90% of the total sample are in those localities (≈85 in HES)
- Stage two – sampling dwellings within the sampled localities
- Finite sampling frame are fixed, but slightly change over the years

Sampling frames in second stage

- The best is to have only one sampling frame but to face the potential under-coverage of main frame we add Complementary frames



Comparison between New Dwellings file and the other sources

	New dwellings	Other sources
Demographic characteristics from LFS 2014		
With 1-2 persons in HH	21.40%	27.30%
With 3-4 persons in HH	44.00%	36.30%
Without children 0-14 in HH	43.50%	55.10%
With 2-3 children 0-14 in HH	32.10%	21.70%
Percentage of HH which the older person is less than 45 years	58.50%	37.90%

Comparison between New Dwellings file and the other sources (2)

	New dwellings	Other sources
Socio-economic characteristics from LFS 2014		
Percentage of not in labor force	24.90%	34.60%
percentage of employment	70.30%	61.50%
Education - without any Diploma until Junior High	16.40%	20.60%
Education - BA	20.70%	16.60%

Comparison between New Dwellings file and the other sources (3)

	New dwellings	Other sources
Socio-economic characteristics from HES 2013		
Gross income (NS)	17,873	16,558
Net income (NS)	15,251	13,809
Total consumption expenditure (NS)	18,285	14,215
Total expenditure on furniture and household equipment (NS)	2,251	662
Expenditure on health (%)	3.40%	5.60%

Dwelling Register

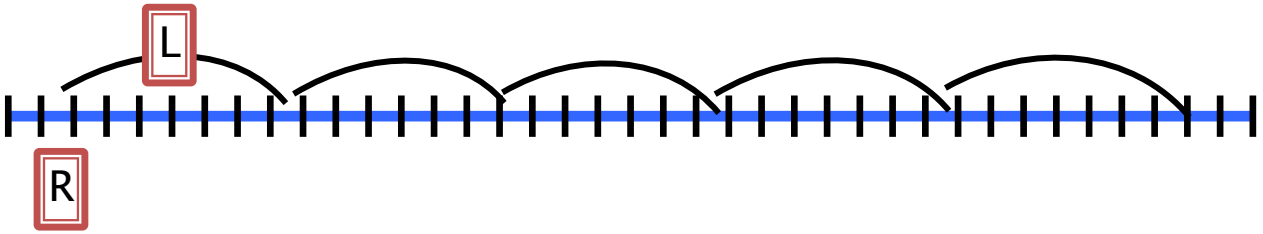
- Source: Municipal Tax Files obtained from several suppliers (companies) who are responsible for managing the municipal tax accounts of local municipal authorities
- Demography and Census Department are responsible for receiving files, checks, maintenance and managing of the Register
- Files are received annually in May–July in a fixed permanent structure (the quality of fields vary between localities)
- Files are received into unified framework and basic checks are conducted

Dwelling Register (2)

- Statistical Methodology Department is responsible for two aspects of accepting and checking files into the register: (a) Determining types of properties for residential. Non-residential records are omitted from the frame (b) determine if number of residential records is reasonable
- Geographic variables are the only key variable for sample design
- Files send to GIS for Automatic Geocoding (frame level) to get Statistical Area
- Imputation of Statistical Area for records which failed in the Automatic Geo-coding (by nearest neighbor method)

Sampling from Dwelling Register

- Drawing a Sample is system developed by IT team
- In each sampled locality, records are sorted by Geographic variables, such as statistical area and a systematic sample is drawn with a random start (R) and length (L) equal to reciprocal of sampling probability.

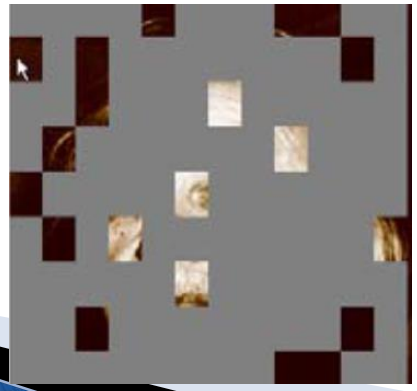


- This method guarantees that the sample spread all over the locality.
- Systematic sampling is highly efficient especially against cluster sampling

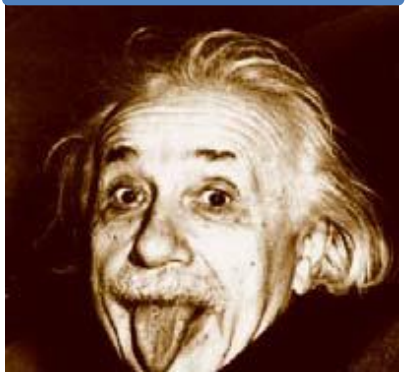
Illustration: comparison between cluster and systematic sampling

Population

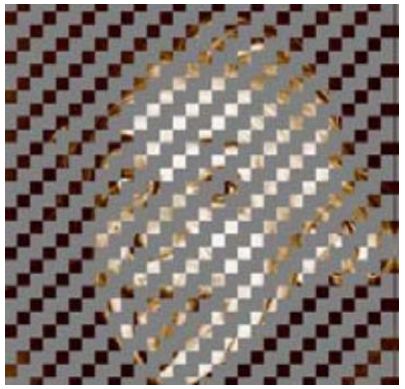
Clustering Sampling



Population



Systematic Sampling



Coordination between LFS And HES

- For a certain year the Dwelling Register is the main sampling frame for both LFS and HES
- LFS is handle first and then HES. Each one of them with the appropriate parameters (Random Start and Length). We check in advance if there may a dwelling in both LFS and HES sample and if so choosing a new random start

Coordinating with samples of the past

- Another issue is to avoid, as far as possible, to sample dwellings that participate in the last 3 years. It is derived from the commitment of ICBS to the public
- if a dwelling is already in selected to LFS and HES of the last 3 years, substitution is made with neighboring record in the sampling frame
- Better way to perform coordination between the two surveys and with the past – later in the presentation

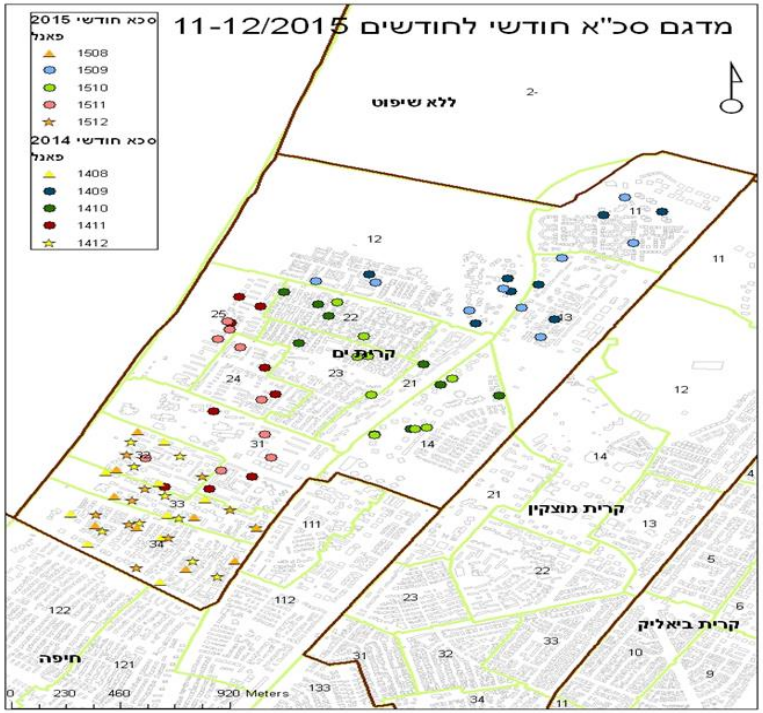
Panels in LFS

- The scheme is 4-8-4 – similar to the American structure
- Each panel is spread over 16 months
- Two consecutive months has 6 overlapping panels
- There is also overlap between 2 consecutive years

Panels in LFS (2)

panel	year 2015												yaer 2016			
	1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4
1310	8															
1311	7	8														
1312	6	7	8													
1401	5	6	7	8												
1402		5	6	7	8											
1403			5	6	7	8										
1404				5	6	7	8									
1405					5	6	7	8								
1406						5	6	7	8							
1407							5	6	7	8						
1408								5	6	7	8					
1409									5	6	7	8				
1410	4									5	6	7	8			
1411	3	4									5	6	7	8		
1412	2	3	4									5	6	7	8	
1501	1	2	3	4									5	6	7	8
1502		1	2	3	4									5	6	7
1503			1	2	3	4									5	6
1504				1	2	3	4									5
1505					1	2	3	4								
1506						1	2	3	4							
1507							1	2	3	4						
1508								1	2	3	4					
1509									1	2	3	4				
1510										1	2	3	4			
1511											1	2	3	4		
1512												1	2	3	4	

Panels in LFS (3)



- Each panel is spread over a quarter of locality
- Since workloads are built from the sample in panel, they are spread out over a relatively small area

Sample allocation over time

- Time is important: (a) The main estimates in LFS are monthly (b) The main characteristics are dependent on data collected over time
- The sample is balanced over time (size and geographic variables)
- Workloads are allocated into week of investigation
- In LFS interviewer required to finish the workload in a week or at the latest in the following week

Sample allocation over time (2)

- In HES the length of investigation is 4 weeks and (Unfortunately) there are too many changes from program provided by the Methodology Department
- The main challenge is in LFS where each month 2 panels rotate in (wave 1 – new panel and wave 5 – after 8 month rest) instead of 2 panels (end wave 4 and wave 8)

Sample allocation over time (3)

- Wave 5 has already a plan. Most of the workloads of wave 1 get automatically the weeks of parallel workloads of the panel rotate out from the same (or nearest) area. part of workloads needed a manually interference
- In HES the balancing is also for socio-economic level of the workloads (according to the last Census)

Social Survey – sample design

- Annually Survey
- Survey population – persons aged 20+
- 85% of the sample is a one stage stratified design
- The sampling frame is based on Population Register
- Main advantage – very rich frame for sampling design. In the other hand geographic variables are poor
- Sample size – 9,600 persons

Social Survey – sample design (2)

- 86 design groups (strata) defined by population group X age group X gender
- The allocation between strata is proportional to size
- In each strata records are sorted by geographic variables
- To illustrate the difference between the design of Social Survey and LFS & HES – population of the city Jerusalem is spread over 70 different strata

Social Survey – sample design (3)

- Original workloads are created from the sample. Each one with ≈30 persons design for 4 weeks of investigation
- For each original workload we determine the quarter, taking into consideration the variability of number of interviewers over the year
- Time is less important in Social Survey than LFS and HES

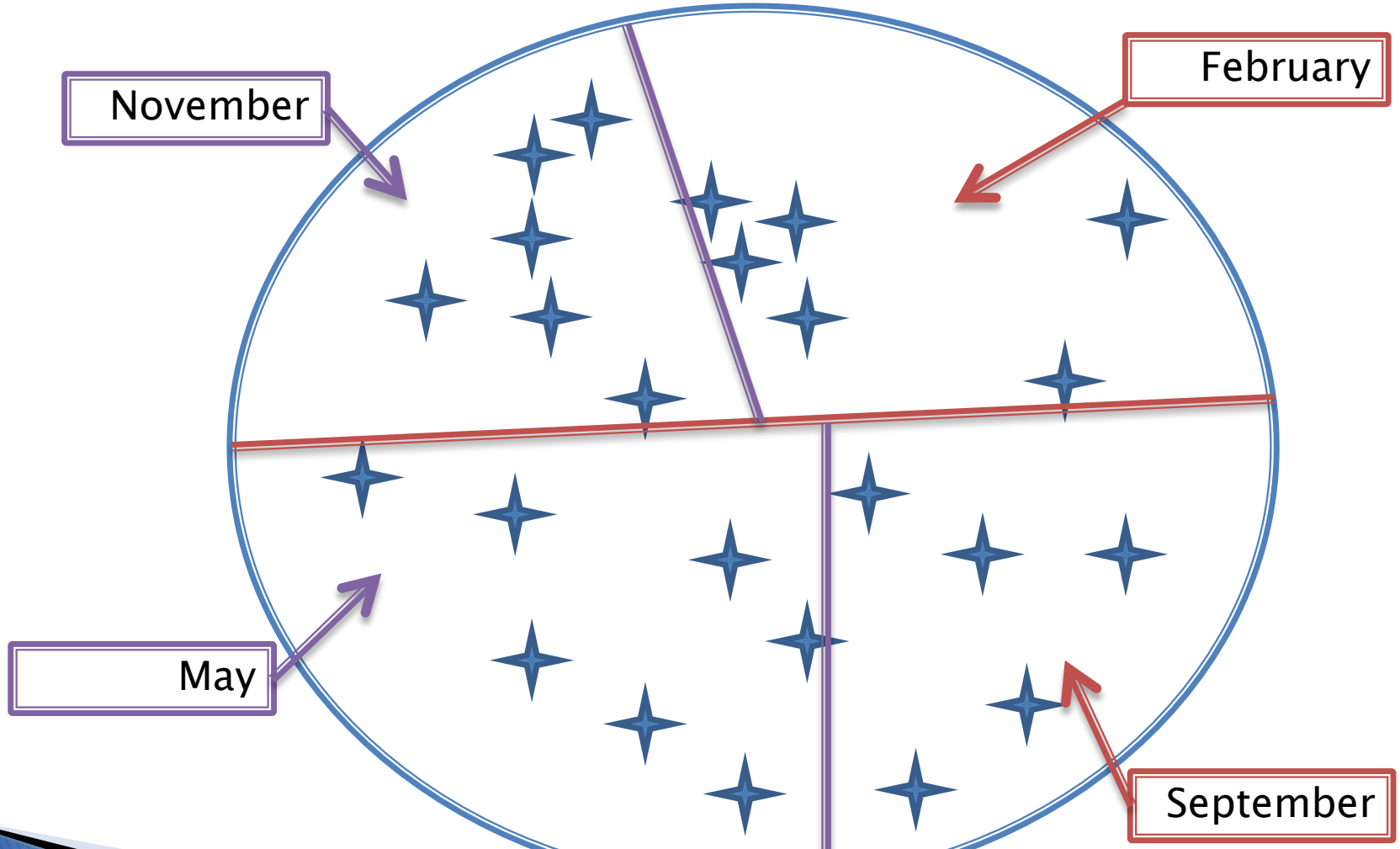
Coordinating with samples of the past

- In each strata we have the sample size allocated. From frame we omit those selected in last 3 years from persons frames. And then selected the sample according to sample size
- In this method those selected in the last 3 years have 0 chance to be selected again and that is better approach than swapping

Future planning in HH surveys

- The plan is for LFS and HES (HH surveys)
- The plan is to design and select each year "master sample". Foe example If LFS has probability of 1/100 and HES 1/200, a master sample will be selected in same method described earlier but with probability 3:200 (sum of 1/100 and 1/200). Then the master sample divided systematically to LFS and HES.
- This approach prevent the chance of dwelling to be selected to both surveys

Future planning in HH surveys – combining field work (HES and LFS)



Thank You