



The Information Manager

From Survey to Social Indicators: The Data Management System of EU-SILC in Austria

Nadja Lamei

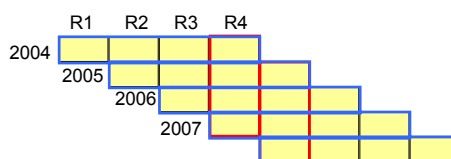
Methodological Workshop Helsinki

November 2006

www.statistik.at

EU-SILC in Austria

2003: cross-sectional survey
since 2004: integrated cross-sectional and longitudinal
survey with rotational design



minimum effective sample size
4.500 households per year
3.250 households between two waves

CAPI survey
fieldwork outsourced

A complex survey

Household context

Personal Register

Person1
Person2
Person3 (Child)

Household quest.

Housing, etc.

Personal quest. 2005

Person1
Income, Edu., Job ..
Person2
Income, Edu., Job ..

⇒ consistency is required!

Longitudinal component

Information from other waves

⇒ consistency also in the longitudinal component is required!

Personal quest. 2004

Person1
Income, Edu., Job ..
Person2
Income, Edu., Job ..

Changes in the survey

in the social transfers system
to the questionnaire

⇒ changes to variables must be met!

Nov 7th, 2006

Nadja Lamei, Statistics Austria

3

Data management...

... is the whole process from the survey to final data and indicators

during the field work

checks during the CAPI-interview
early data checks

after the field work

data editing, imputation, weighting
computation of target variables and indicators
validation

Nov 7th, 2006

Nadja Lamei, Statistics Austria

4

Structure

- Checking raw data and feeding back inconsistencies to field work
- Editing income-variables
- Editing non-income variables
- Imputations
- Weighting and non-response analysis
- Calculating Eurostat target variables
- Calculating Laeken indicators and other key indicators
- Creating tables for publication
- Creating a data set for external data users

Nov 7th, 2006

Nadja Lamei, Statistics Austria

5

Basic principles

Standardization

of all relevant tasks (checking, editing, imputation, analysis)

generalization vs. single case solutions

Transparency

program code and documentation

files saved at various stages

flags

Flexibility

amendments

changes

Nov 7th, 2006

Nadja Lamei, Statistics Austria

6

Example of how to generalize

Problem: No receipt of income from employment/self-employment for people with months worked in previous year >0:

- Check during field work
- Decisions on the basis of case listing

pid	pvrwr	pber	age	sex	aktiv	p001000	anzmens	penam	anzal	alm	int
1980002	0	0	39	2	12	3	0	.00	0	.00	4525
3860003	0	5	21	1	1	2	0	.00	1	1.00	40669
7420002	0	0	50	2	12	3	0	.00	0	.00	5780
28330002	0	0	64	2	12	3	0	.00	0	.00	6014
40000002	0	1	41	2	12	3	0	.00	0	.00	6014
52750002	0	1	32	2	12	3	0	.00	0	.00	5261
69640002	0	1	36	2	12	3	0	.00	0	.00	7456
69950001	0	1	45	2	12	3	0	.00	0	.00	6400
73290001	0	6	61	2	12	3	1	12.00	0	.00	7640
76310001	0	6	69	2	12	3	2	24.00	0	.00	10097
114720002	0	2	47	2	12	3	0	.00	0	.00	7571
121320001	0	0	49	1	12	1	0	.00	0	.00	5519

Number of cases read: 12 Number of cases listed: 12

- Definition of scenarios:
Verification of plausibility for different groups
e.g. apprentices, unpaid family workers, voluntary workers etc.
- Formulating general rules and programming different solutions

```

***Keine Einkommensbezüge bei aktiven Personen (ohne mithelfende), die auch jetzt aktiv sind.

* sollten ein Erwerbseinkommen haben, außer 12 Monate eine Pension oder eine
Arbeitslosenleistung, Krankengeld oder Wochengeld bezogen oder mithelfend.
miss val p075012 p080002 (to thru 0).
compute anz_monate=0.
compute anz_monate=sum(alm, penam, p075012, p080002).
miss val p075012 p080002 (.).

temp.
sel if aktiv>0 and anzerwerb=0 and p001000<=2 and anz_monate<12 and p014000<=81 and
p014000<=82.
list cases /pid age sex aktiv p001000 p038003 p038005 p038004 alm penam p075012 p080002
anz_monate int.

* Personen, die im Bezugsjahr 20 und jünger waren, sich als erwerbstätig einstufen, werden auf
ja beim Unselbständigeneinkommen gesetzt.
* Leihlingsproblem.
do if age<=20 and problem=1.

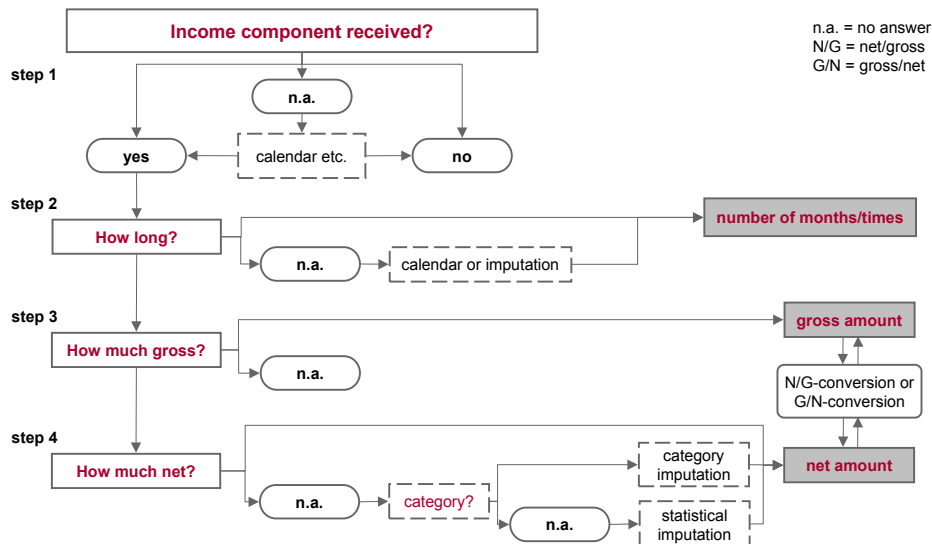
```

Nov 7th, 2006

Nadja Lamei, Statistics Austria

7

Editing procedures for income data



Nov 7th, 2006

Nadja Lamei, Statistics Austria

8

Technical implementation

Programming

in SPSS code

main-control-syntax with sub-tasks in single syntax files

common programming rules

automation and standardization with macros

Variable names

variable name convention helps to standardize

Flags

meta-information for income variables useful for standardization and documentation

Nov 7th, 2006

Nadja Lamei, Statistics Austria

9

Variable name convention

Meaning of the digits

digit 1: identification of the data file (D, R, H, K, oder P)
digits 2-4: number of the question corresponding to the questionnaire
digits 5-6: number of the sub-item
digit 7: information on variable type
0=no income variable
1-5=income variable:
1=component received yes/no
2=frequency of receipt
3=gross amount
4=net amount
5=category
digit 8: to be filled optionally, e.g. with "F" for flag variable

Example

digit	1	2	3	4	5	6	7	8
variable	P	0	5	7	0	2	4	F

P..... The variable comes from the **personal questionnaire**.
P057.... It is question **number 57** from the personal questionnaire.
P05702 It is the **second item** of question 57.
P057024 It is the question for the **net amount**.
P057024F It is the **flag variable** of variable p057024.

Nov 7th, 2006

Nadja Lamei, Statistics Austria

10

Flags for income variables

- 2 not applicable
- 1 no answer and not (yet) imputed
- 1 value according to survey
- 2 value from category imputation
- 3 value from net-gross or gross-net conversion
- 4 value logically deduced
- 5 value statistically imputed with longitudinal method
- 6 value statistically imputed with cross-sectional method
- 7 value from survey was corrected
- value computed from a monthly income (this code applies only to variables of yearly income)

Example: Income distribution by flag for survivors' pensions (EU-SILC 2005)

	1	2	3	5	6	
	survey	category	gross-net	imp long	imp cross	total distr.
Valid n	519	10	12	9	15	566
Valid n in %	92	2	2	2	3	100
Minimum	50	17	99	301	126	8
Maximum	4.000	960	1.221	1.933	1.056	4.000
Median	600	440	549	906	568	600
Mean	623	527	591	1.141	617	627
Stddev	352	340	331	546	285	359

Nov 7th, 2006

Nadja Lamei, Statistics Austria

11

Finally...

...calculation of target variables, Laeken indicators and Gender Pay Gap

Analysis and quality assessment

- Comparison to external data and previous SILC waves
- Eurostat data checks
- Tables

Reports and publications

- Quality reporting
- Publication of results, news items, articles
- Data set for external users

Nov 7th, 2006

Nadja Lamei, Statistics Austria

12

Challenges

- Longitudinal checks and analysis, longitudinal target variables!
- Improve documentation
- Integration of external comparisons and quality control
- Editing of analysis variables and providing them for external users
- Integration of questionnaire