European Health Interview Survey in Latvia – Challenges and Opportunities

Mārtiņš Liberts

Central Statistical Bureau of Latvia

25 August 2015
European Health Interview Survey

- Reliable data on:
  - health status
  - health care
  - health determinants
- from all EU member states
- on regular bases
The First Wave of EHIS

- Gentlemen’s agreement
- 17 EU member states:
  - Austria, Belgium, Bulgaria, Cyprus, Czech republic, Estonia, France, Germany, Greece, Hungary, Latvia, Malta, Poland, Romania, Slovakia, Slovenia, Spain
- 2006–2009
The Second Wave of EHIS


- \textbf{every five years} statistics on health status, access and use of healthcare and health determinants


- European Health Interview Survey Manual
EHIS in Latvia

- The first wave was done in 2008
- The second wave was done in 2014
Population Frame

- The population frame was made as a list of individuals

- Data sources:
  - Statistical Dwelling Register – base for the frame (list of individuals)
  - Population statistics – for over-coverage reduction
  - National Health Service – indication if individual has used national health services
  - Population Census 2011 – additional contact information
Precision Requirements

- Precision requirements are defined with minimum effective sample size for estimating the population parameter “percentage of people severely limited in daily activities”
- It is 4555 for Latvia (defined for each country in the implementation regulation)
- Sample size should be at least 4555 if simple random sampling of individuals is used (full response and frame without over-coverage errors).
Precision Requirements

- Minimum effective sample size can be expressed in terms of the coefficient of variation (CV)

- Input:
  - Minimum effective sample size: \( n_e = 4555 \) (defined by the regulation)
  - Population size: \( N = 1711928 \) (residents of private households in age 15+) computed from the population frame
  - Parameter of interest: \( \hat{p} = 0.106 \) estimated from The European Health and Social Integration Survey (EHSIS) 2012
Precision Requirements

- Expected population variation:

\[ s^2 = \frac{N}{N - 1} \hat{p} (1 - \hat{p}) = 0.0950332 \]

- Corresponding variation, standard error and CV:

\[ \text{var} (\hat{p}) = \frac{1 - \frac{n_e}{N} s^2}{n_e} = 0.0000208 \]

\[ \text{se} (\hat{p}) = \sqrt{\text{var} (\hat{p})} = 0.00456 \]

\[ \text{cv} (\hat{p}) = \frac{\text{se} (\hat{p})}{\hat{p}} = 0.043 = 4.3\% \]
The contact phone number was known for 57% of individuals in the population frame.

The population frame was divided into two groups (master strata):
- CAPI – population part surveyed by personal interviews
- CATI – population part surveyed by phone
Sampling design

- Sample size: 11 340

- Different sampling designs were chosen:
  - CAPI – two stage sampling to balance the cost and precision (sample size 6252)
  - CATI – one stage sampling to increase precision (sample size 5088)
Sampling design – CAPI

- The 1st stage:
  - Stratification by the degree of urbanisation (four strata)
  - Systematic \( \pi \)ps sampling of census counting areas
  - Ordering by geographical location in each stratum
  - Sample size: 1042 PSUs

- The 2nd stage:
  - Sample size: six persons in each PSU
  - Systematic sampling of individuals in each sampled PSU
  - Ordering by NHS, gender, age, random number
Sampling design – CATI

- The 1st stage:
  - Stratification by NHS and seven age groups (14 strata)
  - Systematic sampling of individuals
  - Ordering by region and gender in each stratum
Expected precision

- Expected precision for the estimate of population parameter:
  - population size (population frame),
  - population variation (EHSIS data),
  - response rate (EHSIS data)
  - design effect (EHSIS data: $\text{deff} = 2$ for CAPI, $\text{deff} = 1$ for CATI)
  - sample allocation

\[
\text{Var} (\hat{p}) = \sum_{h=1}^{H} N_h^2 \frac{1 - \frac{n_h r_h}{N_h}}{n_h r_h} s_h^2 \text{deff} (\hat{p})
\]

- It was 3.6% (precision requirements were 4.3%)
Weighting

- There are two sources for auxiliary information:
  - Population frame – many auxiliary variables
  - Population statistics – more precise population counts
- The aim was to use both available sources of auxiliary information in weighting of EHIS
Weighting

- Auxiliary information – population counts by:
  - Gender
  - Age groups
  - Region
  - Education level
  - Usage of public health services
  - Household size
  - Economic activity (employed, unemployed, ...)
  - Degree of urbanisation (DEGURBA)
Two step weighting

- The 1st step:
  - The population frame was calibrated to the population statistics

- The 2nd step:
  - Additional calibration variables were introduced
  - Calibration totals were computed as weighted sums of calibration variables from the population frame
Two step weighting

- Auxiliary vector $\mathbf{x}_i = (\mathbf{x}_i^A, \mathbf{x}_i^B)'$

- Frame totals:
  - $X^A = \sum_U \mathbf{x}_i^A$
  - $X^B = \sum_U \mathbf{x}_i^B$

- Additional totals $\tilde{X}^A$ are available

- Totals $\tilde{X}^A$ are more precise if compared to $X^A$
Frame calibration

- Frame calibration to known population totals $\tilde{X}^A$

- Input for calibration:
  - Design weights: $d_i^F = 1$
  - Calibration variables: $x_i^A$
  - Totals: $\tilde{X}^A$

- Result of frame calibration:
  - Frame calibration weights $g_i^F$
  - Properties: $\sum_U x_i^A d_i^F g_i^F = \sum_U x_i^A g_i^F = \tilde{X}^A$

- $\tilde{X}^B = \sum_U x_i^B g_i^F$
Calibration of respondents

- Respondents can be linked to the population frame through the ID number

- Input for calibration:
  - Design weights with non-response correction: \( d_i^R = \frac{1}{\pi_i p_i^R} \)
  - Calibration variables: \( x_i = (x_i^A, x_i^B)' \)
  - Totals: \( \tilde{X} = (\tilde{X}^A, \tilde{X}^B)' = \sum_U x_i g_i^F \)

- Result of calibration:
  - Calibration weights \( g_i^R \)
  - Properties: \( \sum_{sR} x_i d_i^R g_i^R = \tilde{X} \)
Design Effect

- Two components:
  - Sampling effect:
    \[
    \text{deff}_{\text{sam}} (\hat{\Theta}) = \frac{\text{Var} (\hat{\Theta} | \text{current, HT})}{\text{Var} (\hat{\Theta} | \text{SRS, HT})}
    \]
  - Sampling effect:
    \[
    \text{deff}_{\text{est}} (\hat{\Theta}) = \frac{\text{Var} (\hat{\Theta} | \text{current, cal})}{\text{Var} (\hat{\Theta} | \text{current, HT})}
    \]
Design Effect

Design effect:

\[
\text{deff } (\hat{\Theta}) = \text{deff } (\hat{\Theta})_{\text{sam}} \cdot \text{deff } (\hat{\Theta})_{\text{est}} = \frac{\text{Var } (\hat{\Theta} | \text{current, HT})}{\text{Var } (\hat{\Theta} | \text{SRS, HT})} \cdot \frac{\text{Var } (\hat{\Theta} | \text{current, cal})}{\text{Var } (\hat{\Theta} | \text{current, HT})} = \frac{\text{Var } (\hat{\Theta} | \text{current, cal})}{\text{Var } (\hat{\Theta} | \text{SRS, HT})}
\]
Results

- The main population parameter “percentage of people severely limited in daily activities” was estimated as 0.104 (it was 0.106 in EHSIS 2012)
- Coefficient of variation for the estimate is 3.0 % (calibration effect is taken into account)
- The design effect is 0.717 where:
  - The sampling effect is 0.817
  - The estimation effect is 0.877
Results

Total sample size is 11340:

- Over-coverage cases: 378 (3%)
- Eligible cases: 10962 (97%)
  - Non-response cases: 3885 (35%)
  - Response cases: 7077 (65%)
    - Item non-response cases: 8
    - Item response cases: 7069

Effective sample size is 9858 (requirement is 4555)

\[ n_{eff} = \frac{n_{resp}}{deff(\Theta)} \]
Conclusions

- Combination of one stage and two stage sampling designs:
  - Improvement of precision – the sampling effect is 0.817
  - Increase in interview travelling costs (820 cases were transferred from CATI to CAPI)
Conclusions

- Combination of one stage and two stage sampling designs:
  - Improvement of precision – the sampling effect is 0.817
  - Increase in interview travelling costs (820 cases were transferred from CATI to CAPI)

- Two step weighting:
  - All available auxiliary information (individual level and population level) has been used
  - The estimation effect is 0.877
Conclusions

- Combination of one stage and two stage sampling designs:
  - Improvement of precision – the sampling effect is 0.817
  - Increase in interview travelling costs (820 cases were transferred from CATI to CAPI)

- Two step weighting:
  - All available auxiliary information (individual level and population level) has been used
  - The estimation effect is 0.877

- The precision requirements are fulfilled (effective sample size is 9858)
Thank you!
## Data Collection Modes

<table>
<thead>
<tr>
<th>Mode</th>
<th>NA</th>
<th>CAPI</th>
<th>CATI</th>
<th>CAWI</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAPI</td>
<td>46</td>
<td>5679</td>
<td>163</td>
<td>406</td>
</tr>
<tr>
<td>CATI</td>
<td>0</td>
<td>820</td>
<td>3743</td>
<td>483</td>
</tr>
</tbody>
</table>