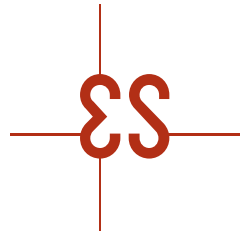


Use of administrative data for official statistics in Statistics Estonia

Kaja Sõstra, PhD
Statistics Estonia

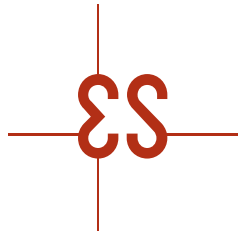
4th Baltic-Nordic Conference on Survey Statistics
Helsinki, August 24-28, 2015





Overview

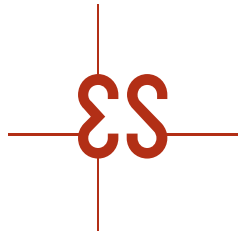
- Growing use of administrative data in Statistics Estonia
- Towards register based population and housing census
- Electronic annual bookkeeping report data for business statistics
- Development of a methodological and statistical basis for a new indicator of services production
- Presentation is based on the research work of my co-workers: Kristi Lehto, Vassili Levenko, Kairiin Kütt, Ethel Maasing, Ebu Tamm, Annika Krutto



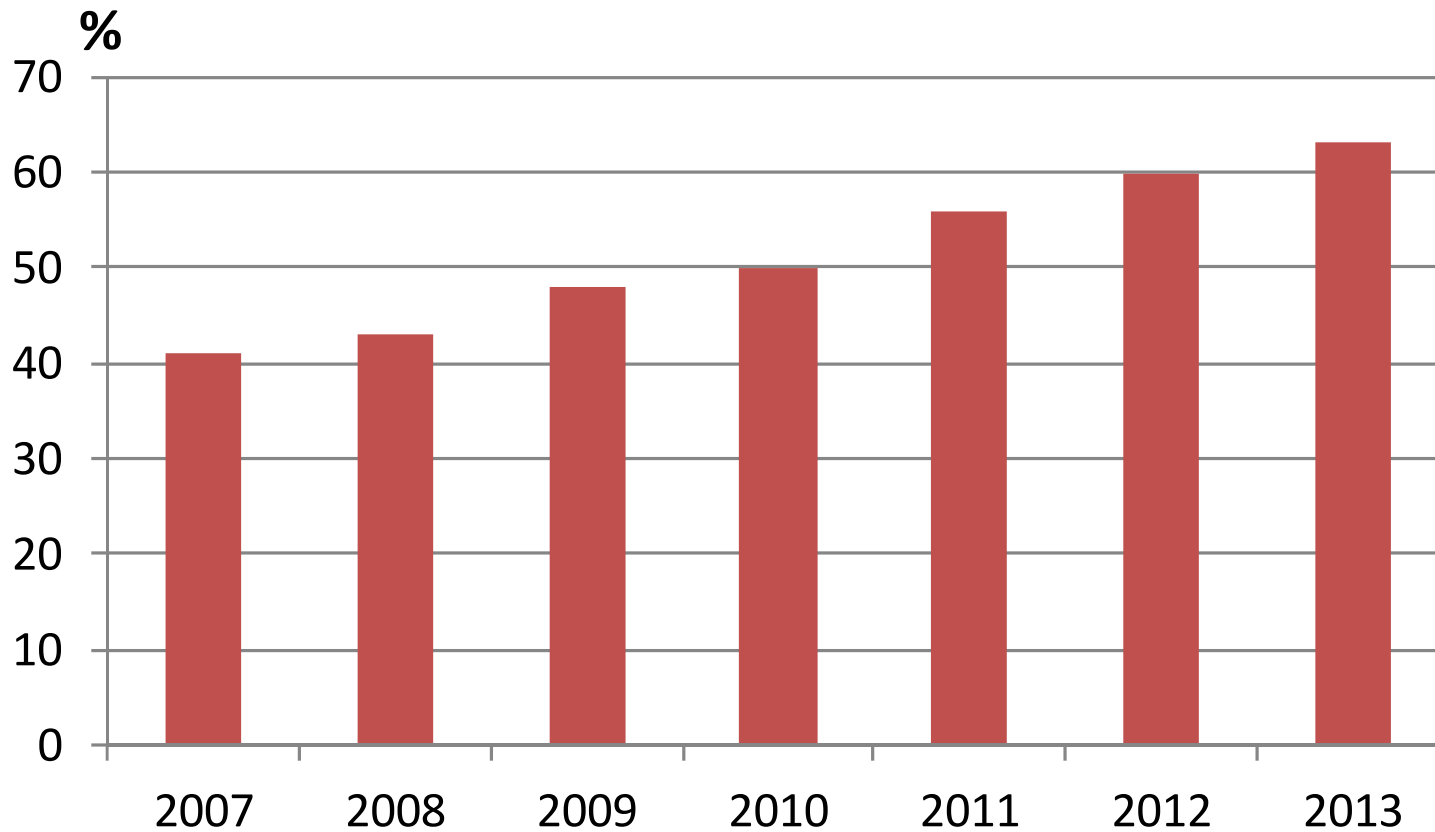
Official Statistics Act

The Official Statistics Act states the following:

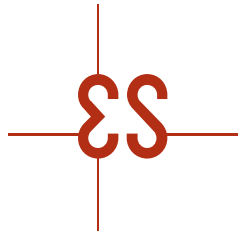
- § 29 (1) Upon the production of official statistics, a producer of official statistics shall primarily use data collected in administrative records /.../ if such data allow the production of official statistics complying with the quality criteria of official statistics.
- § 29 (2) A producer of official statistics has the right to make proposals for amending the composition of data and the classifications used in the administrative records and databases /.../



Use of administrative data in SE

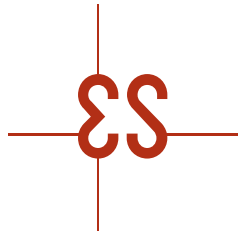


Proportion of statistical surveys that is totally or partly based on administrative data



Towards register based census

- Before the last census, in 2011, preparations were started for the next population and housing census
- An aim is set that the next census, held in 2020-2021 census round, will be register-based
- Instead of interviewing people, data that already exists in national databases and registers will be used.

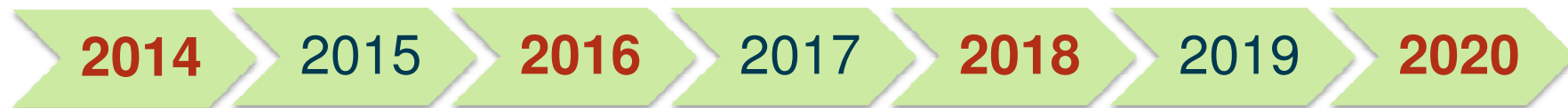


Register based census timetable

2010-2013 methodology project - Statistics Estonia partnered with the Estonian Institute for Population Studies of Tallinn University and the consulting firm Ernst&Young Baltic AS

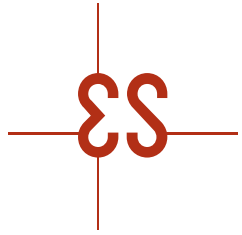
**Pilot census
(31.12.13)**

Test census



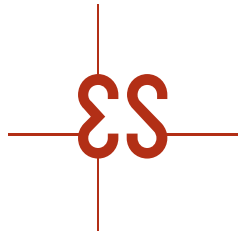
Test census

**Census
2020**



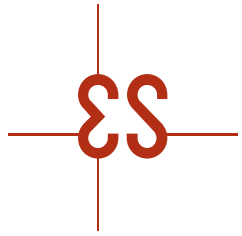
Methodology project

- A total of 20 databases and the data collected in them were analysed
 - meta-analysis
 - definitions and availability
 - detailed analysis
 - construction of census variables (if feasible)
 - data quality assessment (record-linkage to various statistical surveys and 2011 census)



Main registers

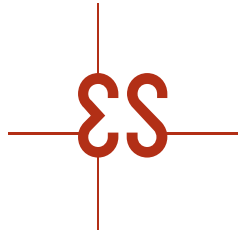
Abbr	Name of register
RR	Population Register
ETR	Register of Residence and Work Permits
EHIS	Estonian Education Information System
EMTA	Register of Taxable Persons
EHR	Estonian Register of Buildings
KR	Land Register
ARIREG	Commercial Register
RKOARR	State Register of State and Local Government Institutions
STAR	Social Services and Benefits Registry



Methodology project

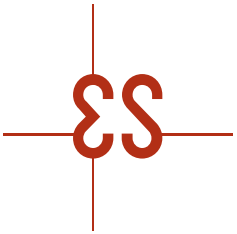
Assessment: core characteristics

Compliance to census standard	Number of characteristics	Non-derived characteristics
A. Complete compliance	4 (14%)	Gender, age, country of birth, country of citizenship
B. Partial compliance	4 (14%)	Current activity status, status in employment, type of living quarters, floor space/number of rooms
C. Limited compliance	17 (61%)	Education, legal marital status, place of usual residence, previous place of usual residence one year prior to the census, ever resided abroad and year of arrival in the country, relationships between household members, tenure status of households, 10 housing characteristics
D. Not available	3 (11%)	Occupation, location of place of work (local activity unit), industry (local activity unit)

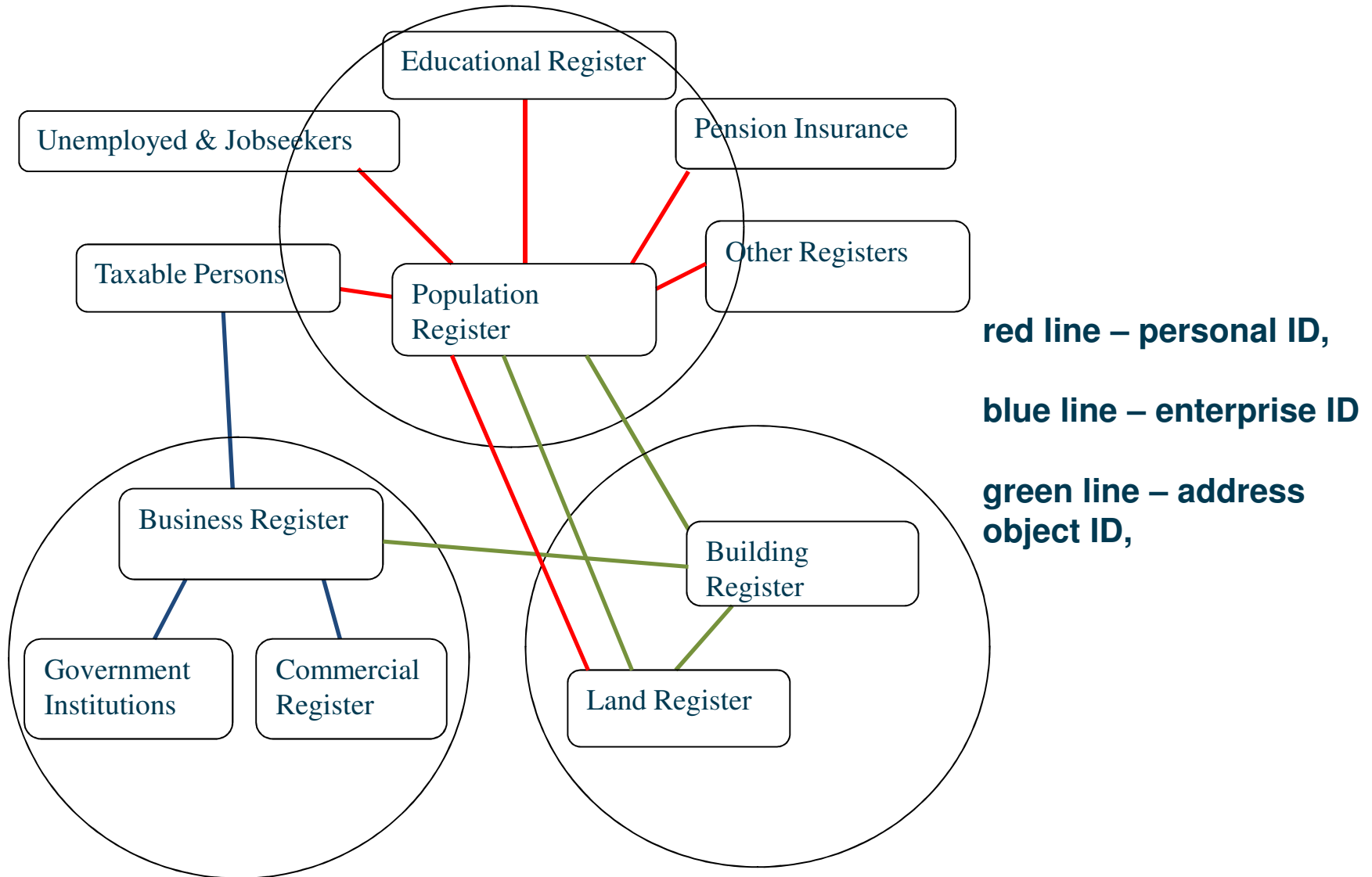


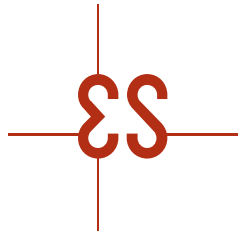
Unique identifications

- For linking registers we need unique identifiers
 - personal ID
 - enterprise ID
 - address object ID



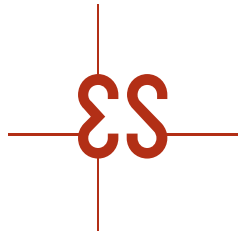
Registers and connections





Problems and solutions

- Missing variables – occupation, place of work
 - Plan A – adding variables to the register of employees, data already exist for immigrant workers
 - Plan B – use of existing information and surveys
 - Register of public employees
 - Register of teachers, medical workers
 - Labour Force Survey
 - Structural Earnings Survey



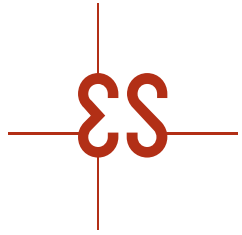
Problems and solutions

- Overcoverage in population register due to unregistered emigration (about 3%)
- Solutions:
 - Project of residency determination
 - Master's thesis "Permanent Residency Status Determination in Register-Based Census", Ethel Maasing, 2015
 - Continuation work with residency index by Ethel Maasing and Ene-Margit Tiit



Problems and solutions

- Large discrepancy between registered and real place of residence (about 20%) due to different benefits (free public transportation, kindergarten, school, social benefits)
- Result: increase of the number of single-parent households compared to census results
- Possible solutions:
 - Cooperation between public institutions
 - Big data?

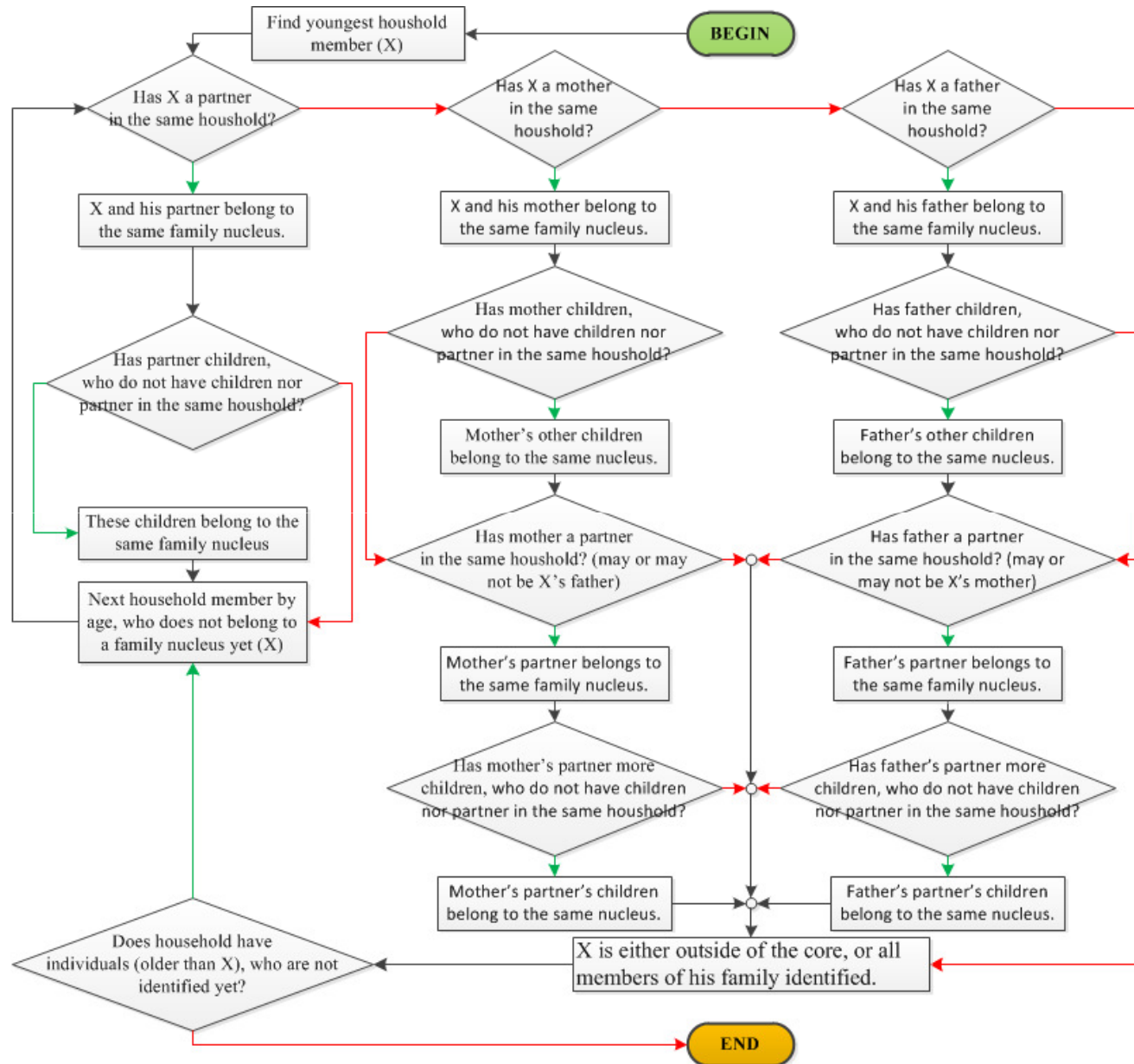


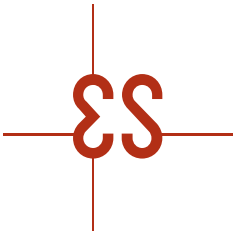
Problems and solutions

- Partly information about household composition
- Solution
 - Algorithm based on information available in population register
 - Address code
 - Mother's ID
 - Father's ID
 - Spouse's ID
 - Master's thesis "Households and families in register-based census" Kairiin Kütt, 2014

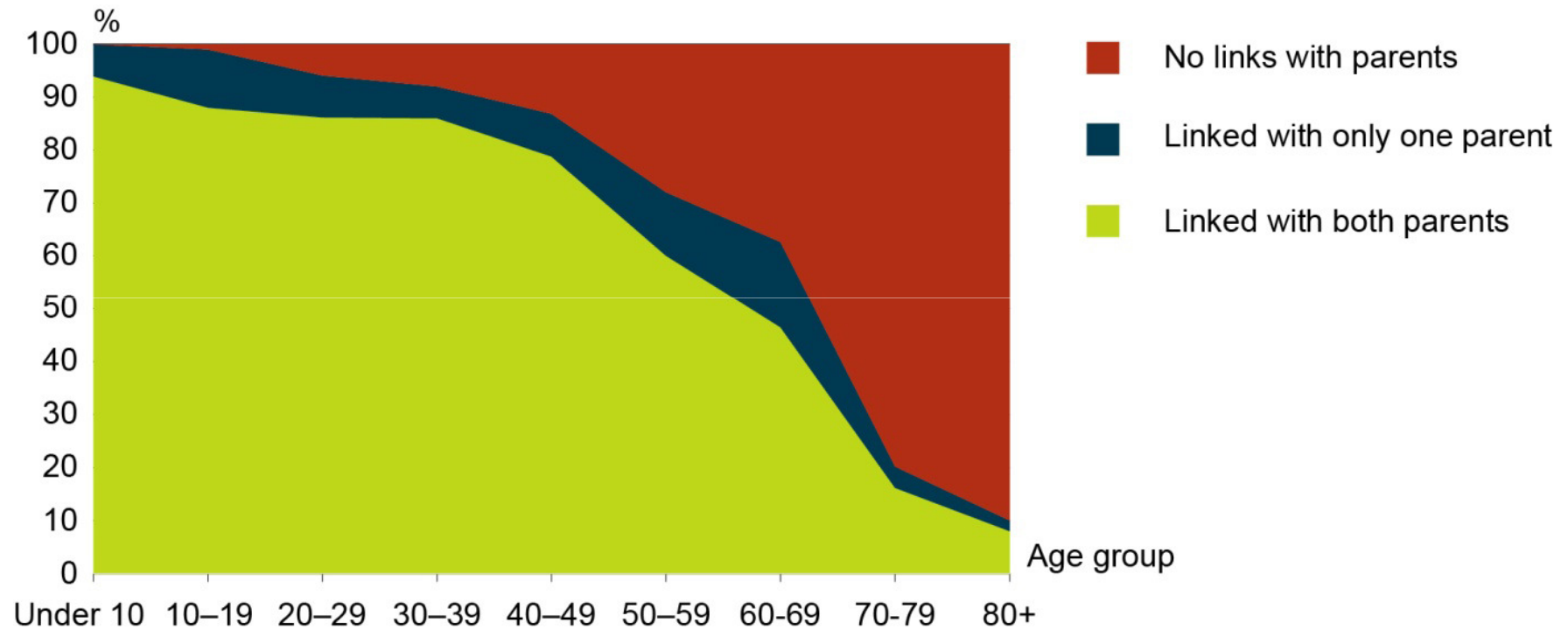


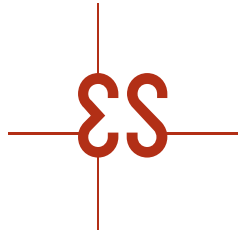
Distribution of household members into families





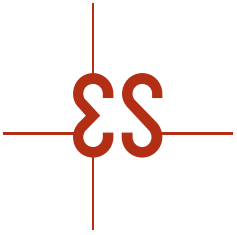
Information about parents in population register by age group, 1 January 2015



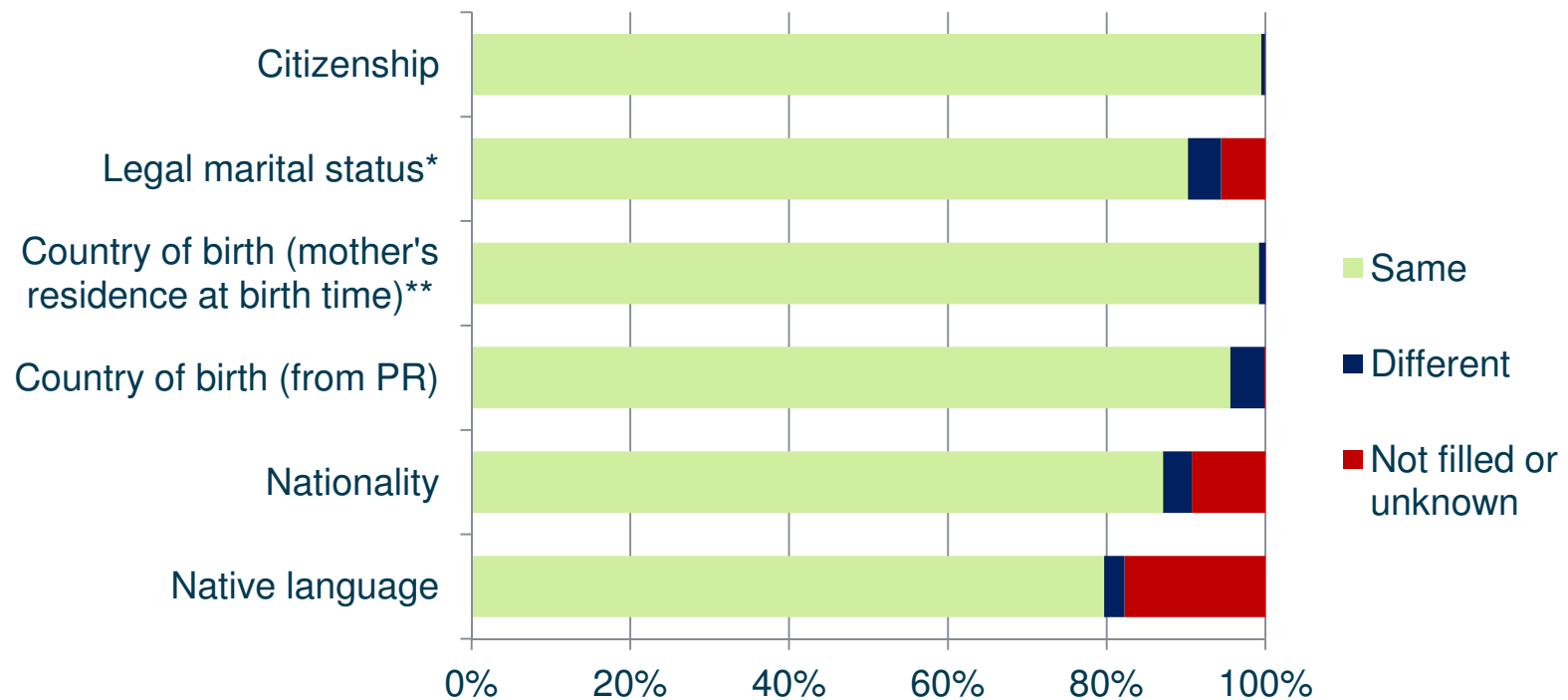


Problems and solutions

- Other quality problems
 - Undercoverage in building register
 - Missing IDs of foreign students in educational register
 - Missing values, erroneous values
- Solutions
 - Cooperation with register owners
 - Regular feedback about quality problems
 - Developing quality framework

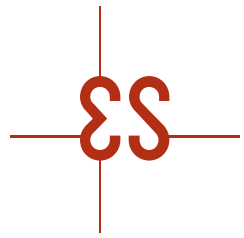


PHC 2011 and PR comparison



* only 15+ years old

** only those whose mother's residence at birth time is filled



Roadmap to a register-based census

- Improving the compatibility of registers, more systematic data quality checks, refining classifications and better documentation of register data
- Adding missing characteristics to the registers
- Improving the accuracy of data on permanent residence, data on dwelling
- Statistics Estonia co-operating with registers, research institutions and experts



Annual Bookkeeping Report electronically

- Annual bookkeeping report (AR) is the primary administrative source for annual business statistics
- All Estonian enterprises are obliged to submit their annual bookkeeping reports to the Commercial Register 6 months after the end of accounting year
- Starting from January 1, 2010 annual bookkeeping reports can be submitted only electronically using XBRL (eXtensible Business Reporting Language) standard
- Since 2012 SE uses AR data online for prefilling statistical questionnaires of annual business statistics
- Since 2014 the cut-off sample is used for some activities replaced by AR data



Estimation of missing variables (1)

- Main problems:
 - annual bookkeeping report does not cover all topics needed for business statistics
 - some parts of report are voluntary
- Different estimation methods were tested before implementing cut-off sample
 - Project “Integrating annual bookkeeping reports into statistical production system” financed by Eurostat
 - Bachelor’s thesis “Imputation of inventories in Estonian Commercial Register” (Cliona Georgia Dalberg, 2013)



Estimation of missing variables (2)

- Population should be divided into three parts:
 - completely enumerated U_C ,
 - sampled U_S
 - excluded from the survey U_E
- Population total estimate of variable y

$$\hat{t}_y = t_y^C + \hat{t}_y^S + \hat{t}_y^E$$

- If the total for below-threshold units is relatively large, it needs to be estimated using the available information



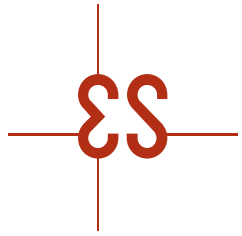
Estimation of missing variables (3)

- Model was created to estimate cost components missing in annual bookkeeping report

$$\hat{t}_{yj}^E = \hat{t}_{yj}^S / \hat{t}_y^S \times t_y^E$$

Where $\hat{t}_{yj}^S / \hat{t}_y^S$ is the proportion of cost component j in sampled part and t_y^E is the total cost for below-threshold enterprises from annual bookkeeping report

- Assumption - the distribution of total costs between components is similar for below-threshold enterprises and for enterprises in the next size class



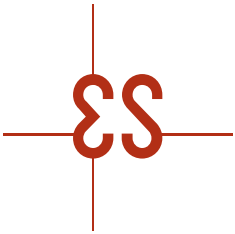
Development of methodology for a new indicator of services production

- The objective is to develop of collection and estimation methods for the production of monthly service turnover and volume indicator with a minimum increase in the reporting burden on enterprises
- Available data
 - quarterly sample survey of enterprises (STS)
 - monthly VAT data of all taxable enterprises
- VAT data differs from surveyed turnover due to the dissimilarity of definitions, value added tax groups, small businesses and credit notes



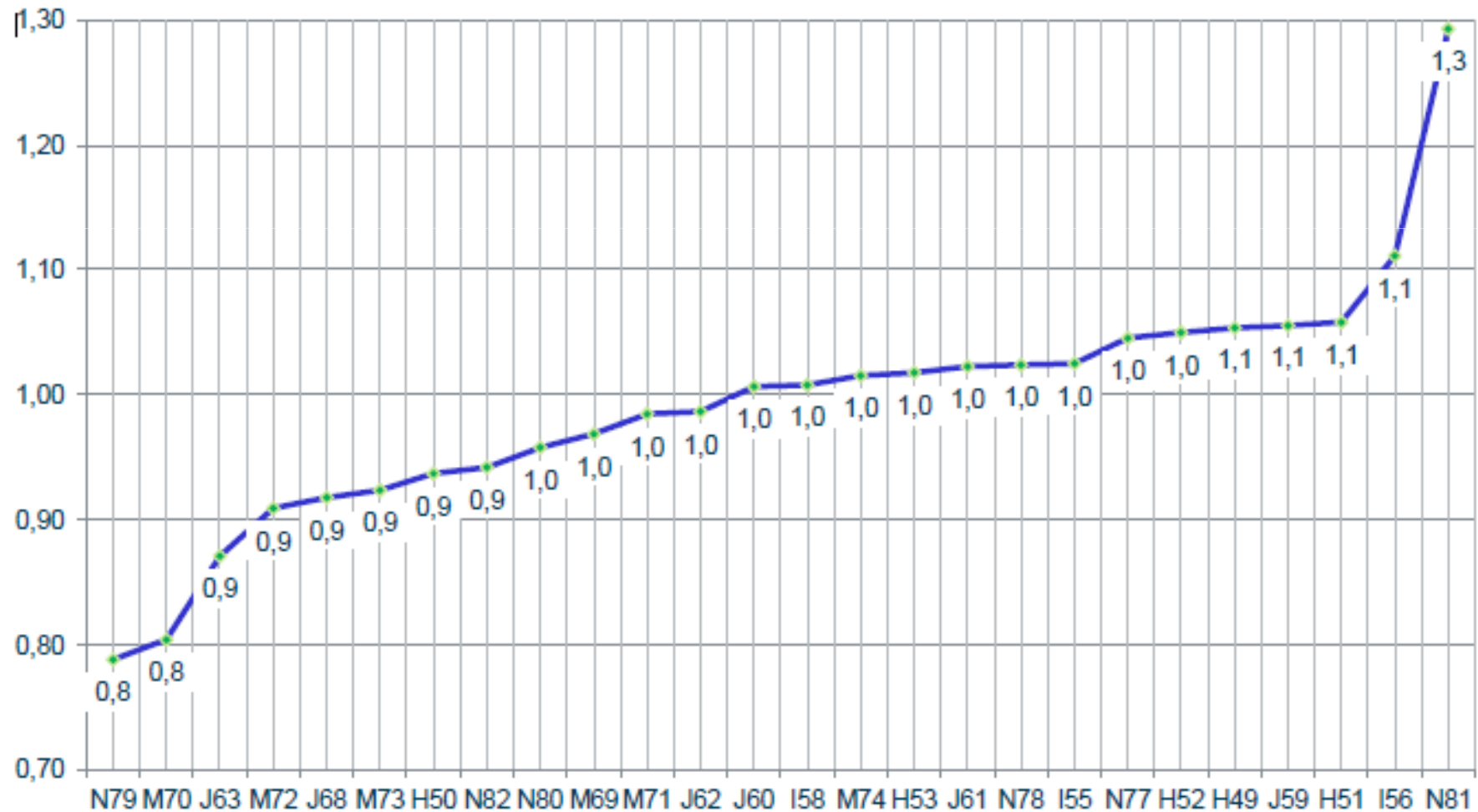
Idea of modelling monthly turnover

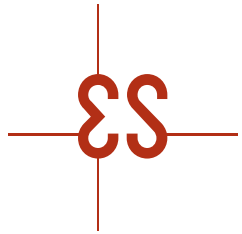
- Data: 2005-2014 quarterly STS survey data; $n=39$;
- Variable: Quarterly proportion of weighted STS turnover to total VAT data on NACE 2-digit levels;
- Model: ARIMA (ARMA) \Rightarrow estimations of proportions on NACE 2-digit levels for next quarter
- Application: the estimated proportions will be applied to monthly VAT data \Rightarrow monthly STS turnover
- Monthly STS data (estimated values) will be corrected after the quarterly data (real values) is available



Results of prediction for 2014 IVq

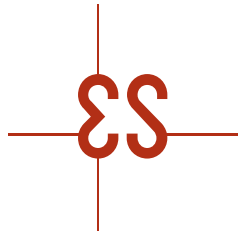
(ARIMA predictions for STS data)/(real STS data) for 2014 IVQ (from 2005 IIQ to 2014 IIIQ, n=38)





Conclusions

- The use of administrative data for producing official statistics has advantages and problems compared to traditional surveys
- The comprehensive methodological work is needed for solving problems: missing variables and values, differences in definitions, coverage problems etc
- Nevertheless Statistics Estonia continues the work with better use of administrative data



References

- Matteus, D. (2013) Roadmap to register-based census. Quarterly Bulletin of Statistics Estonia 4/2013
<http://www.stat.ee/dokumendid/75154>
- Official Statistics Act of Estonia (2010)
<https://www.riigiteataja.ee/en/eli/ee/506012015002/consolide/current>
- Tamm, E., Põldsaar, M., Nestor, R. (2015) Use of annual reports and statistical models in the production of statistics. Quarterly Bulletin of Statistics Estonia 2/2015.
http://www.stat.ee/publication-download-pdf?publication_id=39431

Thank you!

