



# RISQ

Representativity Indicators  
for Survey Quality



KATHOLIEKE UNIVERSITEIT  
**LEUVEN**

**ESRA** European Survey Research Association

# R-indicators and data monitoring

*Koen Beullens – [koen.beullens@soc.kuleuven.be](mailto:koen.beullens@soc.kuleuven.be)*

*Geert Loosveldt – [geert.loosveldt@soc.kuleuven.be](mailto:geert.loosveldt@soc.kuleuven.be)*

*Warsaw, July 1, 2009*





# RISQ

Representativity Indicators  
for Survey Quality



## outline

- Measuring representativeness: R-indicators
- Monitoring representativeness during the course of the fieldwork: application of TQM
- Example: ESS3 - Belgium



# RISQ

Representativity Indicators  
for Survey Quality



## R-indicator

- RISQ – 7th framework programme
- [www.r-indicator.eu](http://www.r-indicator.eu)
- Response rate is weak indicator for survey quality
- Also focus on sample composition (representativity)



## R-indicator

- Representative response:

$$\rho_i = P(r_i = 1 \mid s_i = 1) = \bar{\rho}, \quad \forall i$$

- Basic idea: variance of response propensities
- Use auxiliary variables (age, gender, ...) to estimate response propensities



## Strong & weak representativity

- Strong representativity:
  - Auxiliary variable  $x$  is capable of fully explaining response
- Weak representativity:
  - Response set is representative with respect to available variables  $x_1, x_2, \dots$



## Estimation of propensities

- Use generalized linear model:

$$\rho_i = g^{-1}(x_i' \beta)$$

- Logit, probit or identity link
- Only for categorical data





## From propensities to R-indicator

- Variance of propensities:

$$\hat{S}(\hat{\rho}_X) = \sqrt{\frac{1}{N-1} \sum_s d_i (\hat{\rho}_X(x_i) - \hat{\rho}_X)^2}$$

- R-indicator:

$$\hat{R}(\hat{\rho}_X) = 1 - 2\hat{S}(\hat{\rho}_X)$$

- Value between 0 and 1



## Nonresponse bias

- R-indicator can directly be related to nonresponse bias:

$$Bias(y_{dr}) \leq \frac{1 - R(\rho)}{4\bar{\rho}}$$

- R-indicator can be interpreted as contrast component of nonresponse bias





## Pros & cons

- Pros:
  - Univariate measure
  - Convenient for fieldwork monitoring
  - Variance decomposition to obtain partial R-indicators (attributable to auxiliary variables & strata)
- Cons:
  - Dependence on auxiliary variables
  - May be too complex for non-statisticians



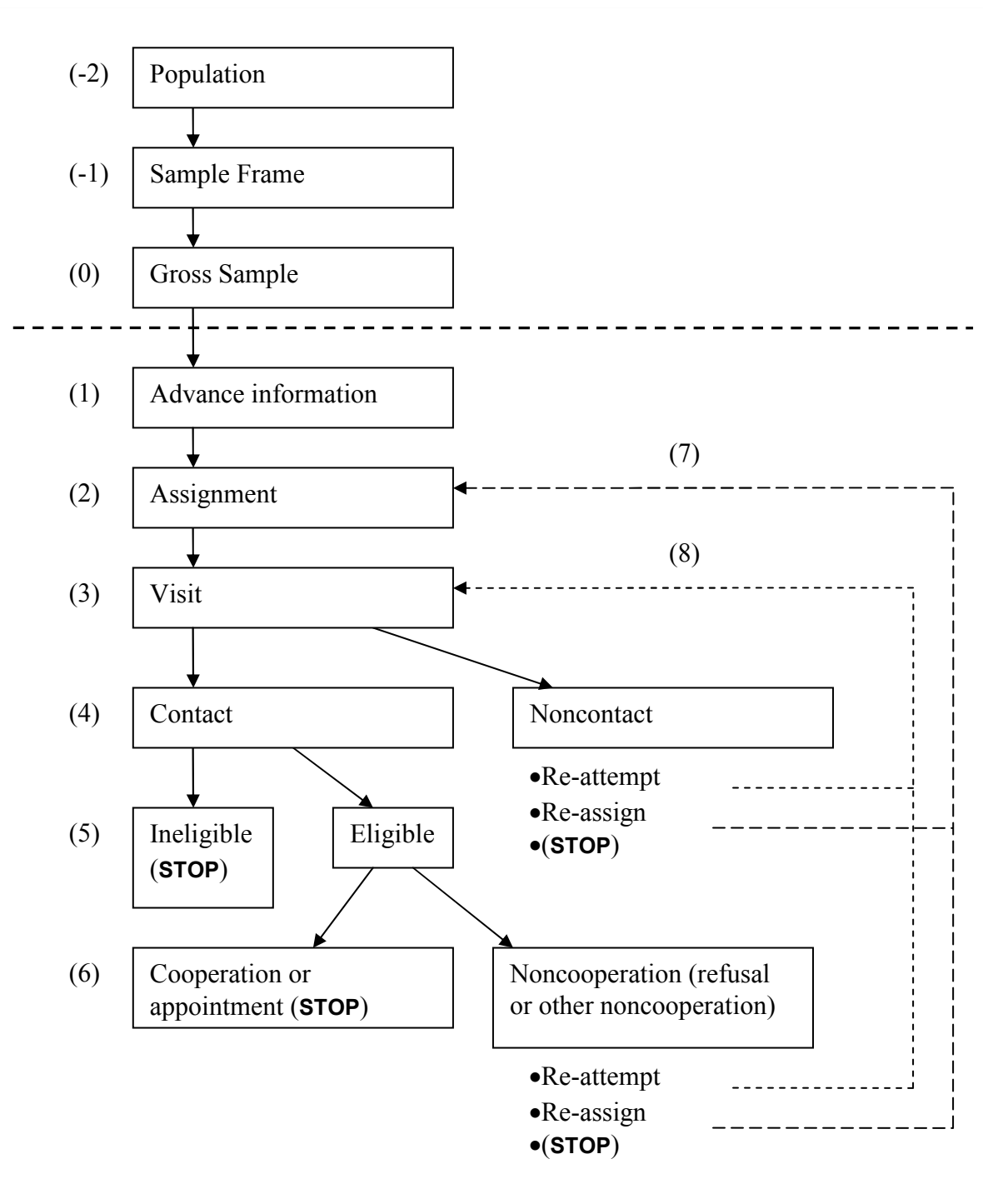
# RISQ

Representativity Indicators  
for Survey Quality



## R-indicators & monitoring

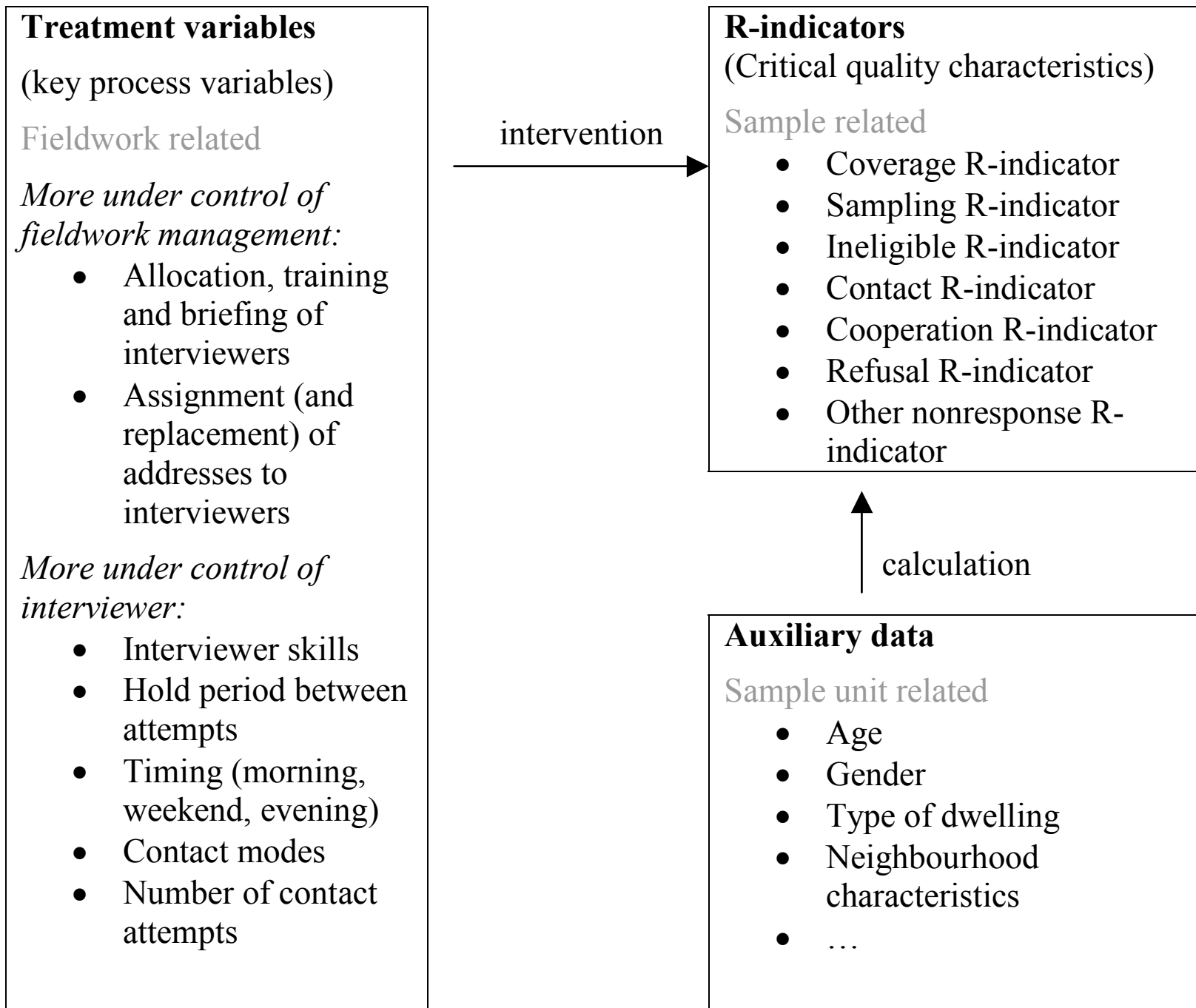
- Final quality depends on process quality (TQM)
- Importance of process data or paradata
  - Relatively new
  - Quality of paradata
  - Integrate paradata in process flow





## Different variables

- Quality variables (R-indicators)
  - Measured at different subprocesses (e.g. contact representativity, eligible representativity)
- Auxiliary variables
  - Available for all sample units
  - Are constant during fieldwork
- Treatment variables
  - Under control of specific survey agent





## Three kinds of propensities

- Raw response propensities
- Equal selection propensities
  - Introduce weights to correct for systematic selection of particular profiles
- Equal selection and treatment propensities
  - Introduce treatment variable to control for systematic assignment of particular profiles to particular treatment
  - Don't use treatment variables for propensities.



## Three kinds of propensities

- Discrete time hazard model
  - Visits on micro-level
  - Individuals on macro-level

$$\begin{aligned} g(h(\rho_{ij})) = & [\alpha_1 V_{1ij} + \alpha_2 V_{2ij} + \dots + \alpha_J V_{Jij}] \\ & + [\beta_1 A_{1i} + \beta_2 A_{2i} + \dots + \beta_P A_{Pi}] \\ & + [\beta_{T1} T_{1i} + \beta_{T2} T_{2i} + \dots + \beta_{TP} T_{Pi}] \end{aligned}$$



person <sub><i>i</i></sub>	visit <sub><i>j</i></sub>	gender	mode	response	(re)visit
1	1	M	F2F	0	1
1	2	M	TEL	1	1
1	3	M			
1	4	M			
2	1	F	F2F	1	1
2	2	F			
2	3	F			
2	4	F			
3	1	M	F2F	0	1
3	2	M	F2F	0	1
3	3	M			0
3	4	M			0
4	1	F	F2F	0	1
4	2	F	TEL	1	1
4	3	F			
4	4	F			
5	1	M	F2F	0	1
5	2	M	F2F	0	1
5	3	M	TEL	0	1
5	4	M	F2F	1	1



# RISQ

Representativity Indicators  
for Survey Quality



## Example: ESS3 - Belgium

- N=3249
- Face-to-face
- Auxiliary variables:
  - Age, gender
  - Belgian regions (Flanders, Brussels, Wallonia)
  - Population density, average income, % foreigners
  - Type of dwelling, neighbourhood conditions



# RISQ

Representativity Indicators  
for Survey Quality



## Example: ESS3 - Belgium

- Treatment variables
  - Interviewer skills (contact skills, persuasive skills)
  - Mode of contact
  - Elapsed time between visits
  - Daily period (morning, afternoon, evening)
  - Day of the week
  - New interviewer
  - Number of attempts
- Quality of process data!



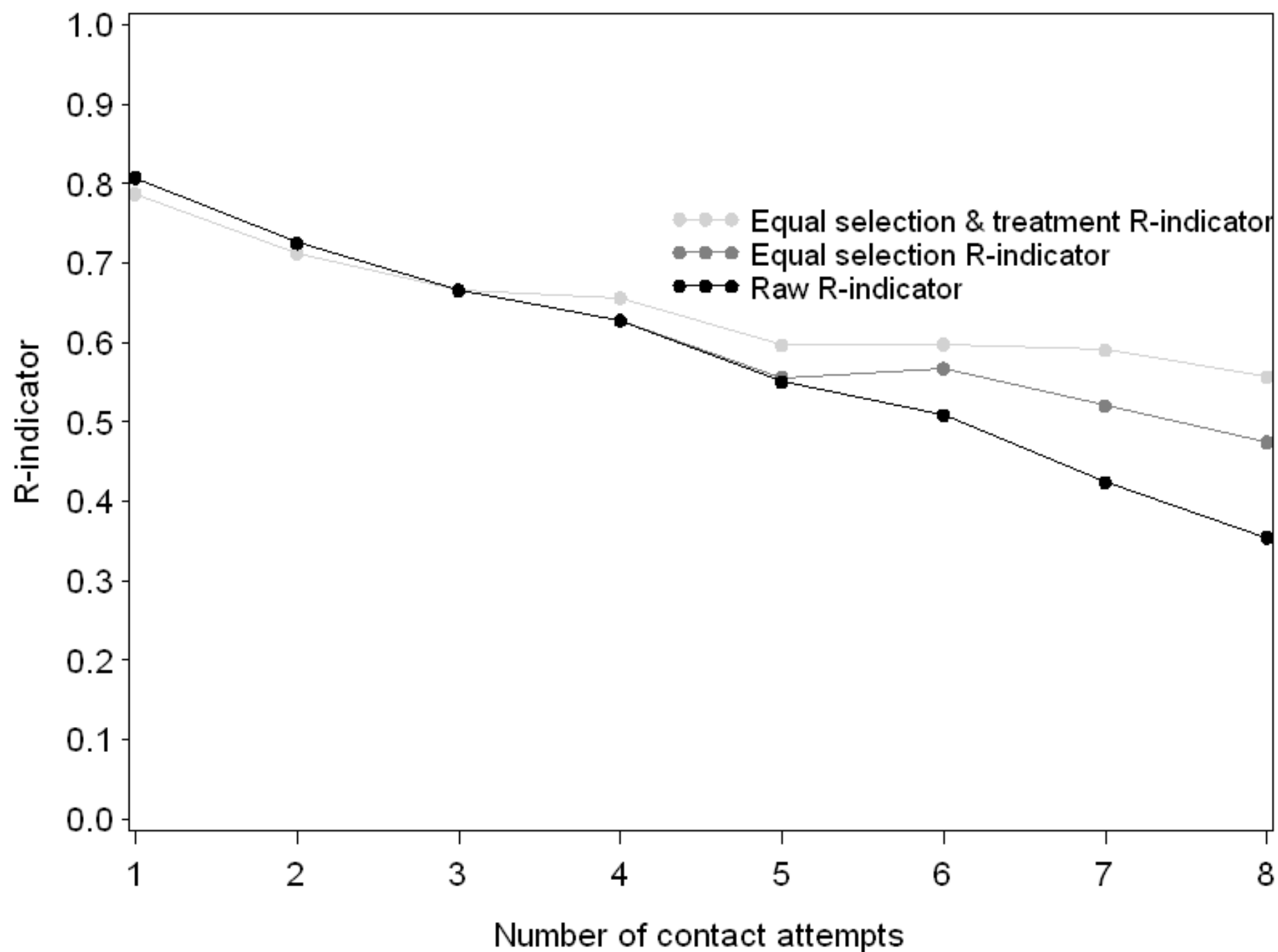
# RISQ

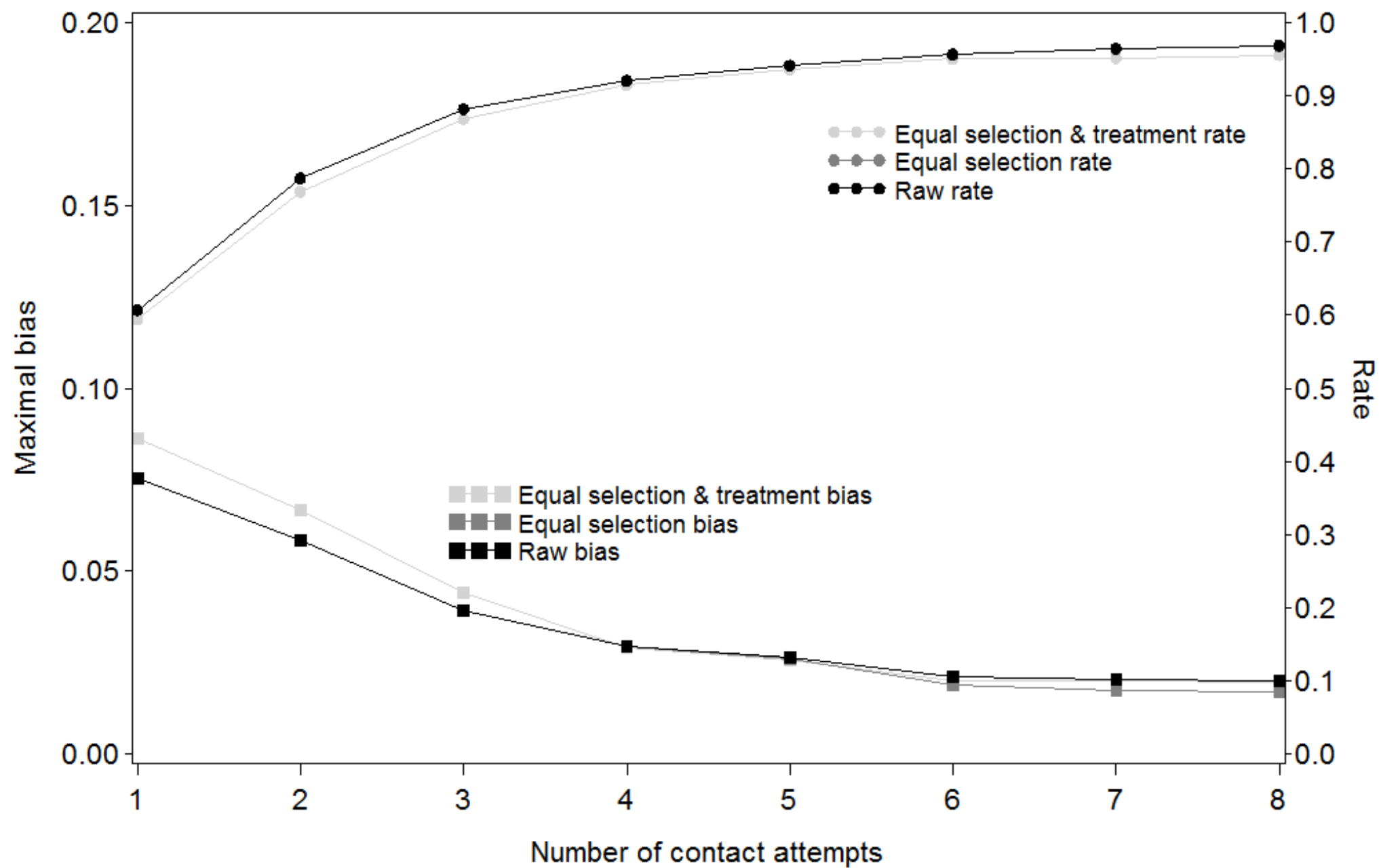
Representativity Indicators  
for Survey Quality



## Example: ESS3 - Belgium

- Contact representativity
  - As a function of number of contact attempt







## Example: ESS3 - Belgium

- Noncontact become more atypical as contact rates increases
  - Reinforces by systematic selection
  - Also reinforced by systematic treatment
- Fieldwork seems to have prioritised the most promising cases





### Partial indicators – equal treatment

<i>Age</i>		<i>Non-Belgians in area</i>	
Age <20	0,04	<2%	0,12
Age 21-40	-0,01	2-5%	0,08
Age 41-60	-0,03	5-15%	-0,05
Age >60	0,03	>15%	-0,15
<i>Gender</i>		<i>Annual Income in area</i>	
Female	0,05	<12.000 €	-0,08
Male	-0,06	12.000-14.000 €	-0,01
		14.000-16.000 €	0,12
		>16.000 €	-0,09
<i>Region</i>			
Flanders	0,11		
Brussels	-0,15	<i>Dwelling</i>	
Wallonia	-0,06	No apartment	0,15
		Apartment	-0,31
<i>Population density</i>			
≤200 inh./km <sup>2</sup>	0,06	<i>Neighbourhood quality</i>	
201-400 inh./km <sup>2</sup>	0,07	Poor	-0,18
401-700 inh./km <sup>2</sup>	0,10	Good	0,09
701-2500 inh./km <sup>2</sup>	-0,08	Excellent	0,04
>2501 inh./km <sup>2</sup>	-0,16		



## Conclusions

- R-indicator useful instrument for fieldwork monitoring
  - Variance function
  - Decomposition into partial R-indicators
- Fieldwork monitoring focuses on treatment variable
  - Can be used for simulation
- Still rather complex activity
  - Need massive amount of data
  - Paradata quality is an issue