

Designs for detection and adjustment of bias in the European Social Survey

Jaak Billiet

CeSO - K.U. Leuven

Member CCT (ESS)

Portuguese Social Attitudes. X Seminar of results'
discussion ESS 2006 – Liboa, November 27, 2008

Outline (1)

- Non-response bias in cross-national research: what is the problem?
- Approaches to the assessment of bias applied in ESS
 1. Bias as deviation between **obtained sample** and **population** (or '*Golden standard*' survey)
 2. Bias as difference between **cooperative** and **converted refusals** collected via refusal conversion
 3. Bias as difference in 'observable' data among **respondents** and **non-respondents** (collected in contact forms)
 4. (Bias as difference between **respondents** and **non-respondents** collected via post hoc nonresponse survey)

Outline (2)

for each approach of **detection** of bias
also approach to the **adjustment** of sample
in order to correct for NR bias

ad 1. post-stratification weighting: some results

ad 2, 3 & 4. weighting according to crucial variables or
using model based methods (*e.g. propensity
scores*) if meaningful (= if good quality of data)

- Conclusions: critical review of the approaches
(*not at the end but after each of the four approaches*)

I. The problem (1)

non-response = an important threat to the validity
of survey research:

*the failure to obtain responses or
measurements for all sample units*

Why a threat? Nonresponse *can* produce **bias** in
the results

What is nonresponse bias?

The context (2)

- Nonresponse bias is determined by **two** factors:
 - Nonresponse rate
 - Differences between respondents and non-respondents (e.g. differences in means and variances)
- In cross-cultural studies: these 2 factors, and thus non-response bias, may differ from one country to another
- Focus on the analysis of ***differences*** in non-response bias between country means (or proportions)
- Consequence: international comparisons cannot be made without adjustments for non-response bias

The problem (3)

- Consequence: international comparisons cannot be made without adjustments for non-response bias
- When bias in cross-nation comparison ?

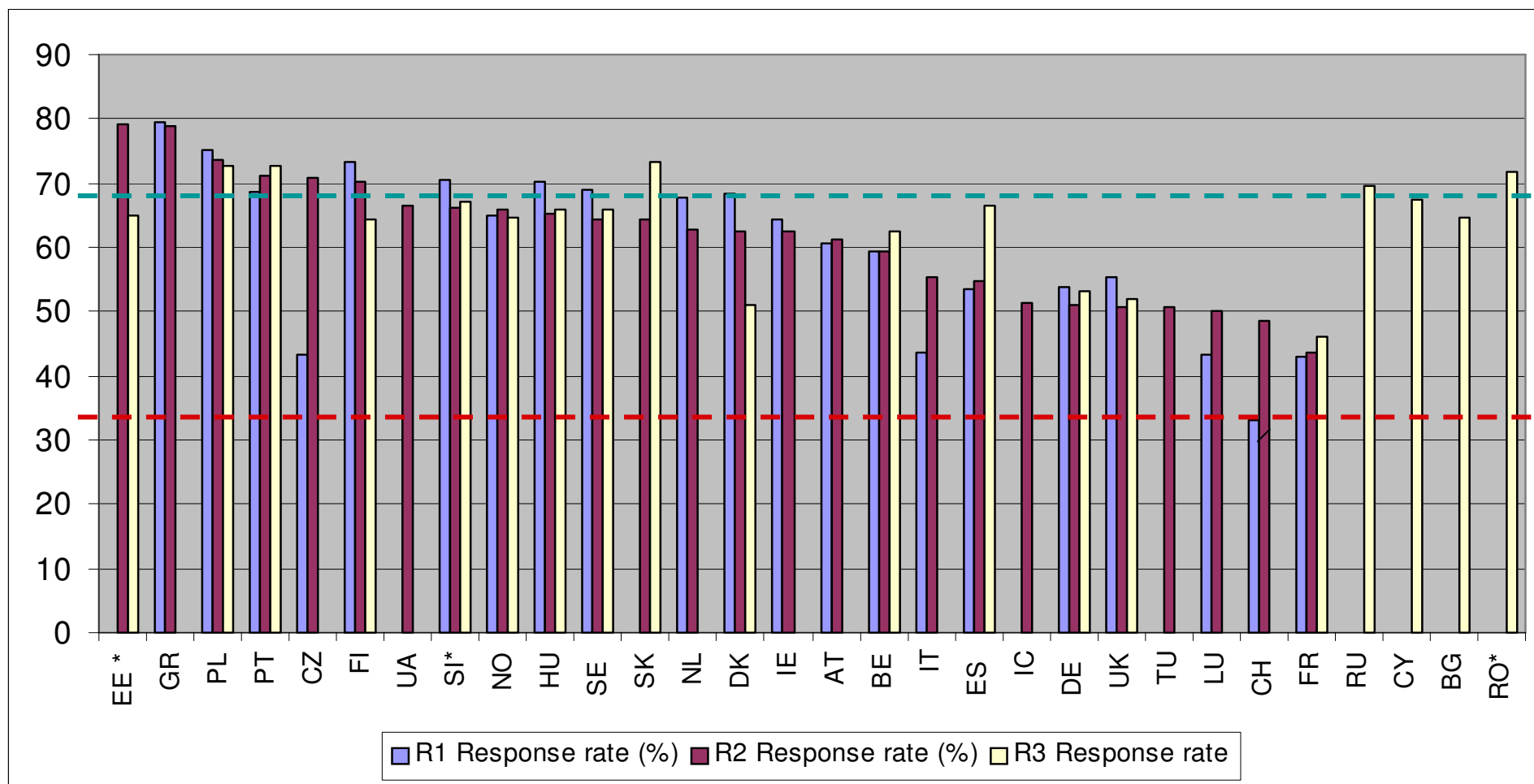
| SITUATION | BIAS |
|--|-------|
| no bias in separate countries | NO |
| equal bias in separate countries and non response rates are equal | NO |
| equal bias in separate countries and non response rates very different | YES ? |
| different bias in separate countries | YES |

The problem (4)

- In order to obtain comparable response rates (equivalence in rates) the ESS norm was set to 70%
- As a result of the strategies to obtain this aim:
 - response rates higher than in comparable surveys
 - larger average response rates in R3 (65% - 60.4% in R1))
and much lower SD (R1 = 10.2; R3 = 7.1)
 - but... still large differences between countries

(see: figure for previous rounds)
- **Consequently:** NR bias likely in ESS rounds:
estimation and adjustment for differences in non-response bias required

Figure 1: obtained response rates in ESS R1-R3 (based on CF)



* For these countries based on NTS in R3

II. Approaches to NR bias applied in ESS

- 1. Bias as deviation between obtained sample and population: method = post-stratification**

Post-stratification (2)

- **Methodology used:** (applied in R1-R2 reports of V. Vehovar, 2007)

Step 1: compare sample distribution against population distribution on 3 PS variables (gender, age, education)

Differences between $W1$ and $W2$ inform about size of (MAR) bias, magnitude of PS weights and VIF in each country

$W1$ = observed sample only weighted for design weights

$W2$ = final weighted sample weighed for dweights x ps weights = representation of population distribution

Post-stratification (3)

Estimate *Variance Inflation Factor (VIF)* after PF weighting in each country

$$VIF = 1 + CV^2_w \quad \text{minimum} = 1.0$$

CV^2_w = ratio between the elementary variance of the weight var w and the square of the arithmetic mean of weight variable w

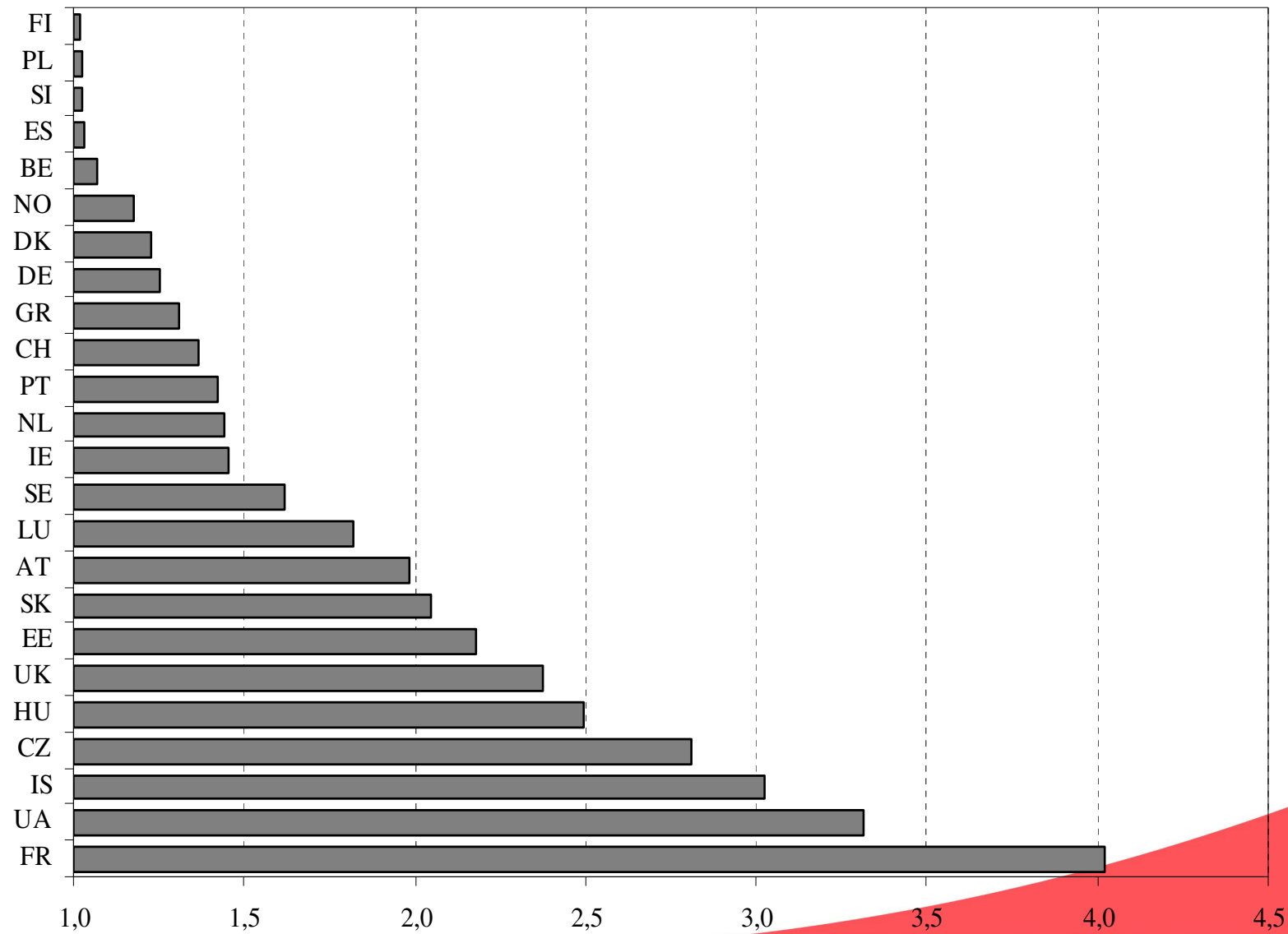
estimates the effect on sampling variance of weighted variables y_w in comparison with variance in unweighted sample y^*

$$Var(\bar{y}_w) = Var(\bar{y}) \times VIF$$

very large post-stratification weights & *VIF* in FR, UA, IS, CZ, UK, HU, EE, SK, AT

* Somewhat more complex because there is also VIF because of design weight (adjustment of non *EPSEM* samples)

Figure 2. Variance inflation factors (final weights) ESS R2 (*Vehovar, 2007: 343*)



Post-stratification (4)

Step 2. compare *usual suspected* attitudinal variables before and after post-stratification adjustment

■ effect of PS computed for 45 target variables

- definition of *NR bias*: $bias(\bar{y}) = | \bar{y} - \bar{y}_w |$

large amount of information 45 vars x 25 country's

- *relative bias (Rbias)* = $\frac{bias(\bar{y})}{\bar{y}_w}$

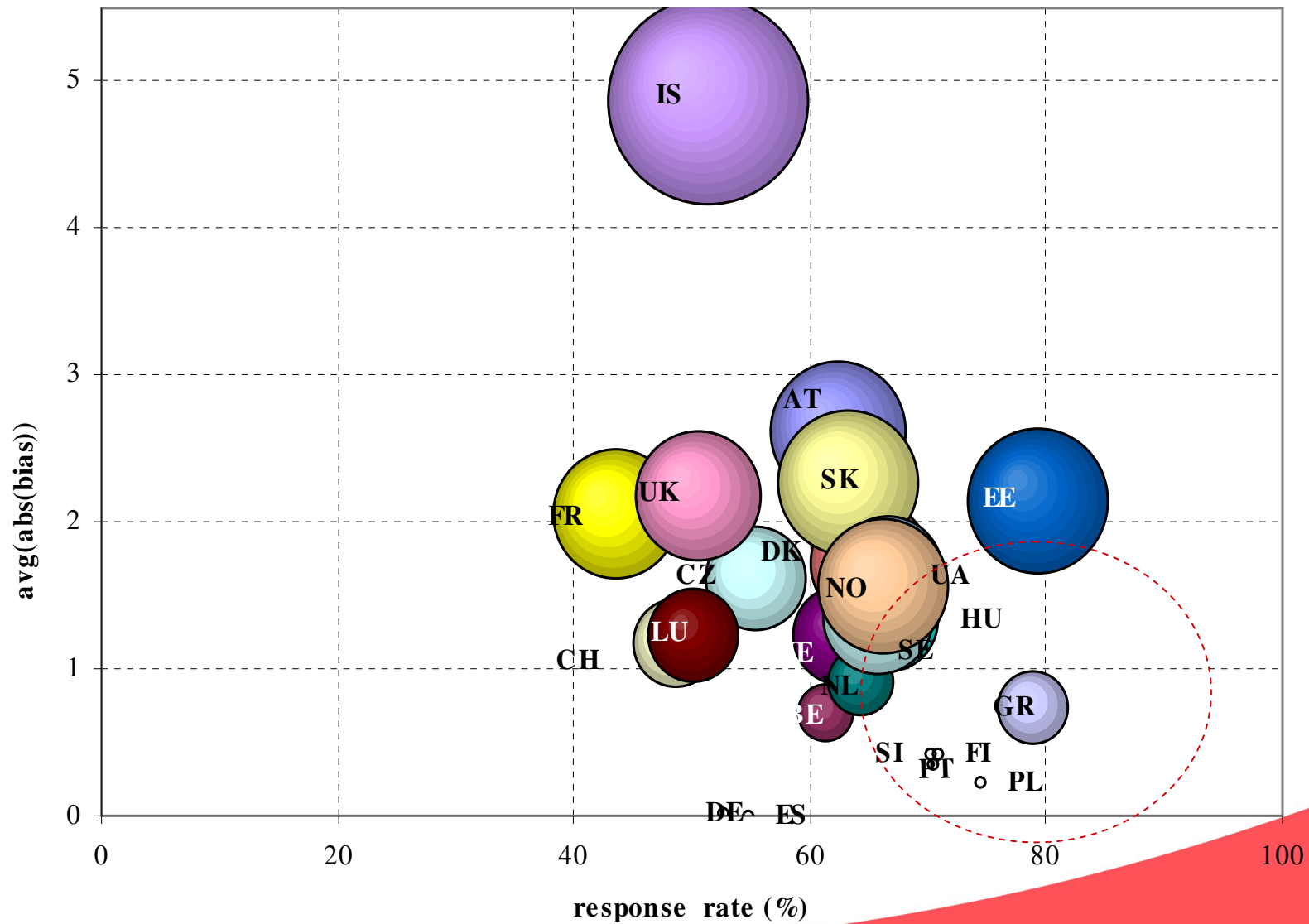
- *standardised bias (Sbias)* = $\frac{bias(\bar{y})}{se(\bar{y}_{SRS})}$

- In order to compare countries: *Average standardised absolute bias (ASbias)*: average over 45 items per country (*Vehovar, 2007*)

Post-stratification (5)

- Compact overview in *Figure 3*: relation between amount of bias (*vertical axis*), country response level (*horizontal axis*), and # of items with $S_{bias} > 1.96$ (*bubbles*)
- correlation between AS_{bias} and response rate = **-0.26**
(the higher response rate the lower the AS_{bias})
- Items most subjected to bias are items on Political interest, immigration items, media use (S_{bias} 2.10---3.65)

Figure 3: The absolute average standardised bias in relation to the response rate of the country samples (*source: Vehovar, 2007 adapted*)



Post-stratification (6)

Illustration

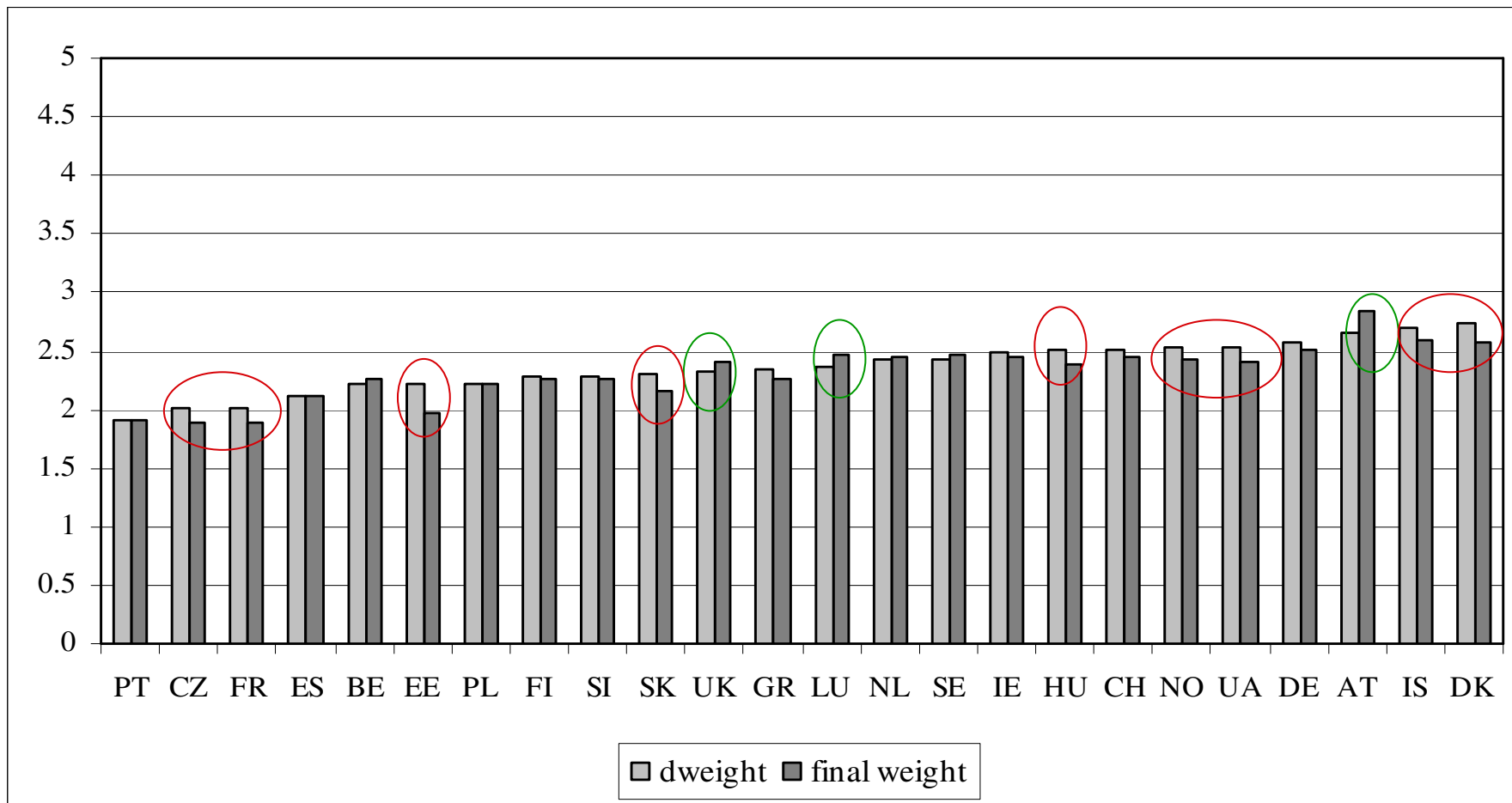
Multiple indicator *Latent* variable on politics and ethnic threat (contains indicators with $AS_{bias} > 2.5$)

In **simple** comparison and in **multiple** regression

- 1.** Simple comparison (next Figure 4 on interest in politics)

Post-stratification (7)

Figure 4. Mean scores on latent variable *interest in politics* (pol_int: 0-5) in country W1 samples with (*dweight*) and in samples W2 samples(*final_weight*) in ESS Round 2



Post-stratification (8)

- Differences not very large
- In different directions

2. Effect on multivariate regression models with dependent variables political interest and ethnic threat (latent variables with several biased indicators)

comparison of W1 (only dweight) and W2 (final_weight)
model with inflated variances (*multiply the standard errors by the square root of VIF*)

in samples with different response rates,
with higher ASbias (EE) and one with lower ASbias (DE)
and samples with very high ASbias (EE) and same
response rate (DE)

Post-stratification (9)

Results:

no differences in conclusions for ethnic threat in IS and DE (although large differences in response and ABias

only two differences in EE for political interest (gender has effect in final_weighted sample, and ever_job has no effect)

no difference in DE

Post-stratification (10)

The post-stratification approach: Critical appraisal

- **Pro:** +/- uniform for all countries
- **Pro:** correct for all non-response (refusal + noncontact)
- **Pro:** at least a view on direction of nr bias
- **But:** Deviation between sample and population distribution not only caused by non-response but also by sampling error (overestimation of NR bias)
- Using PS-weights given rather small effects and amount of *VIF*?

Question: outweighs the reduction in bias the loss in precision because of *VIF*?* (*accuracy versus precision?*)

**Mean Square Error* (MSE) of estimates
= the sum of the *variance* + the *square of bias* (Groves: 1989).

Post-stratification (11)

- **Neg:** not very effective when target variables not strongly correlated with PS-variables (target variables too heterogenous within strata)
- **Neg:** correct only for small fraction of NR bias
- **Neg:** PS-estimators are biased in themselves (errors in '*population distributions*')
- Better PS variables needed that stronger covary with target variables (several methods proposed in recent literature... *but what is their quality?* Makes no sense to exchange very critical assessed PS-weighting for un-validated new sources)

Approaches to NR bias applied in ESS

2. Bias as difference between cooperative and reluctant respondents

cooperative vs. reluctant respondents (1)

Assumption: converted refusals (reluctant respondents) are informative for final non-respondents

■ Method:

- convert as much as possible non-respondents
- data for analysis: call record data files (explain...)
- analysis: compare distributions of *cooperative* and *reluctant* respondents on demographics and attitudinal variables (“*usual suspects*”), and study effects of variables on probability ratio *reluctant/cooperative*

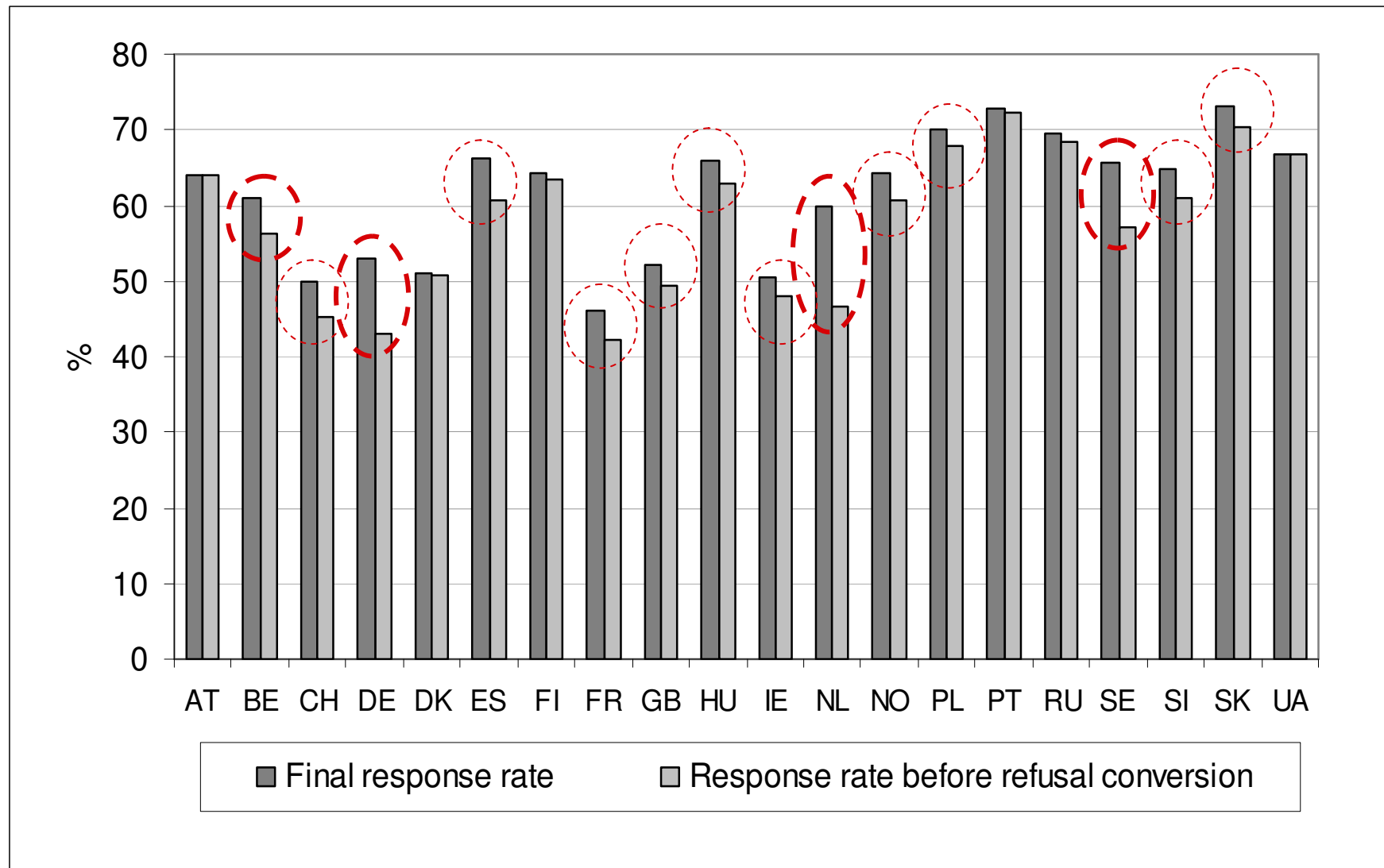
= this might provide information about direction and amount of bias (not accurate estimation!)

cooperative vs. reluctant respondents (2)

Problems

- Only few countries with large numbers of reluctant respondents (Billiet et al., *JOS* 2007) (see figure 5)
- Based on assumption of “*continuum of resistance model*” but alternative assumption is “*classes of nonparticipants model*” (Stoop, 2004)
- Definition of reluctant respondent (*converted refusal*) much dependent on decisions of interviewers
solution: attempts to differentiate between “*easy*” and “*difficult*” to convert respondents based on info in call record data (Beullens et al. *ESRA* 2007)
- Amount of reluctant respondents depends on practice of survey organisation (*comparability over countries?*)

Figure 5 – Effects on final response rates of refusal conversion in ESS R3



cooperative vs. reluctant respondents (3)

- Effects very different according to country (R1, R2, R3?)
- Most effect in the Netherlands (most biased? or **bias best measured?** (R1, R2)
- Selection of survey organisation & interviewers not comparable over countries (R1, R2, R3)
 - some re-approach **all** refusals
 - some only selection, but **not** random
based on **subjective estimation** about later cooperation

cooperative vs. reluctant respondents (4)

- 'reluctant' respondents very different in DE and NL (R2)
- implications for bias estimation... reluctant respondents representative for final respondents??? See NL
'soft' and *'hard'* refusals = refused *'once'* or *more*
= same biased variables as in post-stratification (ASbias)
see Table 1)

Table 1. Multinomial baseline logit estimates (β) and odds ratio's of belonging to soft or hard refusals versus cooperative respondents (reference) with respect to background, attitudinal, and media use variables (ESS R2, the Netherlands)

| Predictors | Refused once | | Refused twice | |
|--|--------------|------------|---------------|------------|
| | β | Odds ratio | β | Odds ratio |
| Male | -0.1301 | 0.878 | 0.9279*** | 0.395*** |
| Single | -0.2908 | 0.748 | -0.4962** | 0.609** |
| Level of education | | | | |
| Low education | -0.2019 | 0.817 | -0.4006* | 0.670* |
| Middle education | -0.2314* | 1.260* | 0.1146 | 1.121 |
| High education | -0.0295 | 0.971 | 0.2860* | 1.331* |
| Minutes watching television / day | 0.0028* | 1.003* | 0.0023 | 1.002 |
| Minutes reading newspaper / day | -0.0013 | 0.999 | 0.0077*** | 1.008*** |
| Perceived threat by immigrants | 0.0781 | 1.081 | 0.1562*** | 1.169*** |
| Trust in political institutions | 0.0782 | 1.081 | 0.1450** | 1.156** |
| Social isolation | -0.0028 | 0.997 | 0.1123** | 1.119** |
| Satisfied about own life | -0.0929 | 0.911 | -0.2040*** | 0.815*** |
| Max-rescaled R-Square | 0.0755 | | | |
| -2 Log Likelihood | 2786.151 | | | |

cooperative vs. reluctant respondents (5)

Conclusion:

- part of the converted refusals are ‘*soft*’ and other part are ‘*hard*’. Bias only detected among ‘*hard*’ refusals
- What are proportions of “*hard*” and “*soft*” among *final refusals*?
- How to adjust the data for bias given these problems?
Estimations of response propensities that emerge from model based adjustments not to be trusted until now

cooperative vs. reluctant respondents (6)

- Not so useful in cross-nation context (strategies and measurement too different, not enough data)
- Practice in field organisations too diverse (more standardization needed, and there is already a lot... realistic?)
- **NEG:** Only refusal component of nonresponse, nothing on non-contacts (*How does non-contacts differ from reluctant respondents and final nonrespondents?*)

PRO: substantial response enhancement (R3) + better survey climate in survey organisation: refusal conversion communicates to interviewers that response is important and serious

Approaches to NR bias applied in ESS

3. Information from observable data in CF

Enriched sample with observable data (1)

What? Interviewer observations about housing type and neighbourhood recorded in Contact Forms (CF) for all selected sample units (respondent + nonrespondents) + estimating of gender and age category in CF

Problems: estimation of *age* and *gender* not among non-contacts plus bad quality (tresholt < 10% missing observations) = useful for only **5** countries in R2
observed data about *housing* and *neighbourhood* somewhat better quality, however useful for only **14** countries in R2
(quality R1 < R2 < R3)

Enriched sample with observable data (2)

Variables measured:

- (1) pre-coded housing type
- (2) physical state of building/dwellings: 5 cat;
- (3) commonness of litter or rubbish in immediate area: 5 cat;
- (4) vandalism, graffiti or damage to property: 5 cat

Two single indicator vars (1 & 2) and 1 equivalent two indicator (latent) variable (3+4)

Method = logistic regression

dept var 1: initial refusal vs cooperative

dept var 2: reluctant, re-approached but not-contacted, re-approached but second refusal (all versus cooperative)

Enriched sample with observable data (3)

Some results: summary

- Living in **apartment** increases likelihood to refuse and not to be contacted
- More likely to refuse or non-contact when physical state is worse
- Litter and vandalism in environment: less and weaker effects
- Interaction between housing type (apartment) and physical condition in ES and SK (more likely not to participate and non-contact in apartment in bad state)

Enriched sample with observable data (4)

Observable data: critical appraisal:

- **PRO:** all sample units used (cooperative, refusals, non contacted)
- Chosen variables moderate correlate with target variables as social status (education)
- **CON:** measurement problems = missing data
- Measurement problems: quality of coding of observation (interviewers not specifically trained = much to do)
- Adjustment for bias not possible (model based weights, propensity scores estimation) until data quality improved for more (all) countries

Approaches to NR bias applied in ESS

4. Difference between respondents and non-respondents

Survey among nonrespondents (1)

- New survey among refusals with very small & easy questionnaire (some crucial variables) (*Voogt, 2004; Saris*)
- Implemented in ESS Round 3
 - 4 participating countries
 - Full mail survey (12 questions) after main survey in NO (*medium rr*), CH (*low rr*) & PL (*high rr*)
 - At moment of refusal **7** crucial questions in BE
response of 44.7% among non-respondents

Survey among nonrespondents (2)

Some examples from OL

■ Basic questions procedure (PL)

Short 7 question module (+ at door): work situation, highest level of education, # of members in household, frequency of social activities, feeling (un)safe, interest in politics, attitude towards surveys

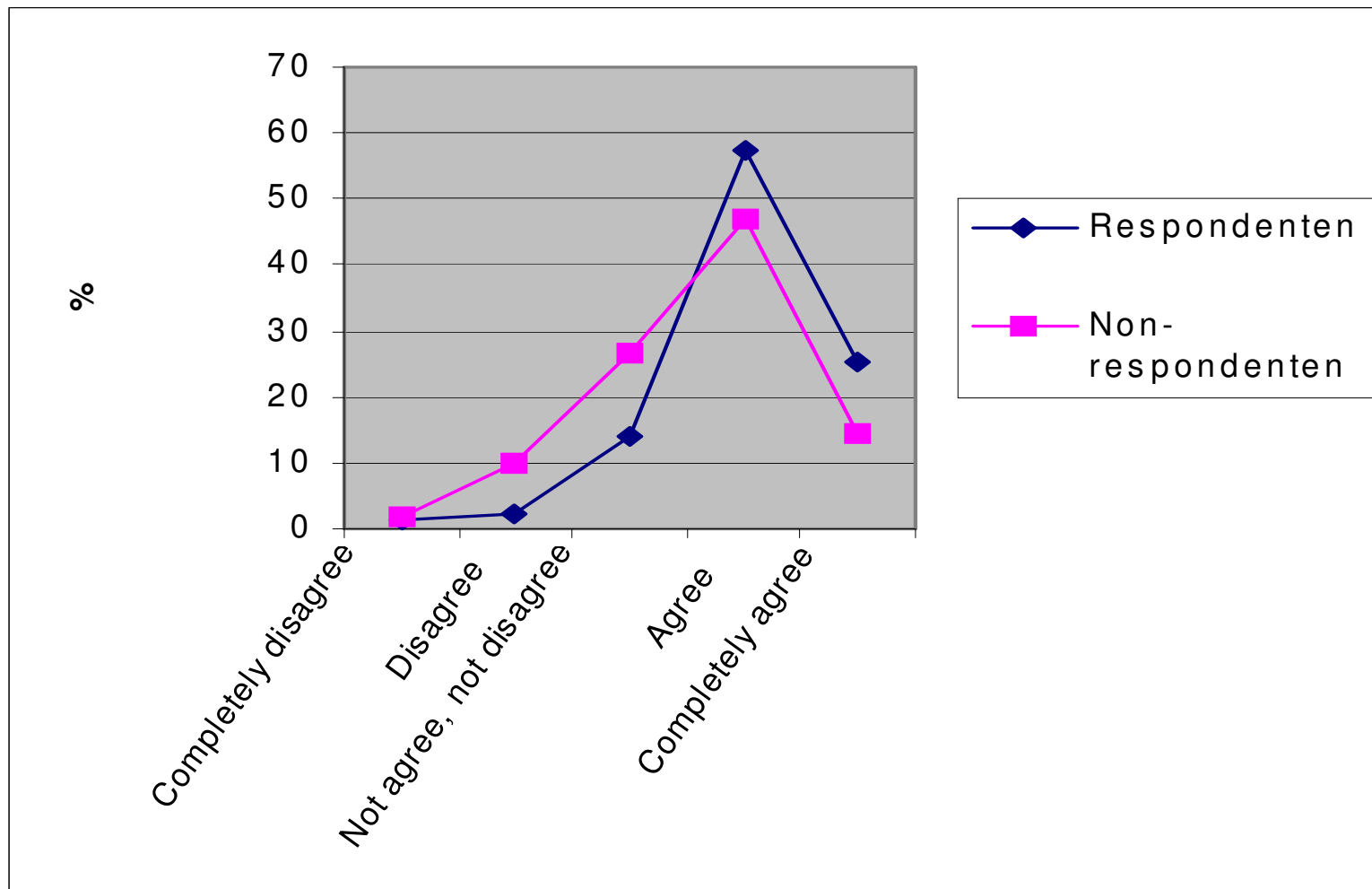
Normal 15 questions module (version with reason for refusal): same as short + gender, year of birth, TV watching, voluntary work, trust in people, satisfied with democracy, trust in politics, immigration good/worse for country + reasons for refusal (closed)

Normal 15 questions module (version with survey climate): same as 15 questions before – reasons, + attitude towards surveys

Survey among nonrespondents in PL (3)

| Questionnaire version | Sent questionnaires | Filled in questionnaires |
|---------------------------------|---------------------|--------------------------|
| Respondents – long | 838 | 480 (57.28%) |
| Respondents – short | 830 | 520 (62.65%) |
| Converted refusers – long | 28 | 8 (28.57%) |
| Converted refusers – short | 25 | 16 (64%) |
| Non-respondents – long | 404 | 89 (22.03%) |
| Non-respondents – short | 230 | 54 (23.48%) |
| Non-respondents not able – long | 192 | 49 (25.52%) |
| TOTAL | 2547 | 1216 (47.74%) |

Survey Climate: *surveys are valuable for the whole society, as we all want to know what the [inhabitants of Poland] think and what opinions they have on various important matters*



Survey among nonrespondents (4)

Preliminary appraisal:

- Seems possible but small number of responses among prior non-respondents in PL, CH and NO (22%-30%)
- Only useful for specific countries (limited use in cross-nation perspective if not used in all cntry's)
- No privacy concerns until now? (yes in some countries)
- Expensive (120,000 ERO more for 4 contries + research time)
- Analysis has just been started... critical appraisal = later



discussion

See appraisal after each section

thank you

