Comparing sample quality: challenges of applying external and internal criteria of representativeness to cross-national survey projects

Piotr Cichocki, AMU Poznań
Piotr Jabkowski, AMU Poznań
(in collaboration with Marta Kołczyńska, ISP PAN)

Building Multi-Source Databases for Comparative Analyses
Warsaw, December 16-20, 2019

Funding information

This work was supported by the grant

Reflecting Europeanisation: cumulative databases of cross-country surveys as a tool for monitoring European public opinion trends

awarded by the National Science Centre, Poland
(grant no. 2018/31/B/HS6/00403)

Approaches to evaluating sample quality

Desirable properties of survey quality measures

The perspective of the secondary data users:

1. Procedures of survey quality assessment should lead to direct quantitative measures of representativeness without resorting to background survey information (Mohler and Uher 2003; Mohler et al. 2008);
2. Evaluation process should be easy applicable at large scale not only in the within- but also cross-project assessments;
3. Sample assessment procedures – to be broadly applicable – must rely on: a) information that is routinely made publicly available and b) survey variables omnipresence in questionnaires .

Demographic representativeness: Focus on Gender

1. Distributions of gender or age are most typically used in sample quality evaluations (Kobilanski et al. 2019; Struminskaya et al. 2014; Groves and Peytcheva 2008)
2. Note their omnipresence in questionnaires, straightforward measurement, low item non-response, as well as the availability of reasonably reliable population statistics in most countries of the world.
3. Although representativeness in terms of gender does not in itself preclude biases regarding other characteristics (Voogt and Van Kempen 2002), it does constitute a confidence-booster for overall sample quality.

Existing approaches for evaluating survey quality

Procedures that do not require individual-level auxiliary data:

1. External criteria of representativeness, i.e., comparisons of sample estimates to “gold standard” benchmarks from external sources (Eckman and Koch 2019; Koch 2016);
2. Internal criteria of representativeness, i.e., comparisons of survey estimate from a specific subsample to a parameter known by definition (Sodeur 1997). The most common applications: Menold (2014), Kohler (2007), Eckman and Koch (2019), Jabkowski and Cichocki (2019);
3. Comparisons of weighted and unweighted estimators (Vehovar 2007; Billiet et al. 2009).

Note that: all these procedures refer to the concept of the *Total Survey Error*!

Data

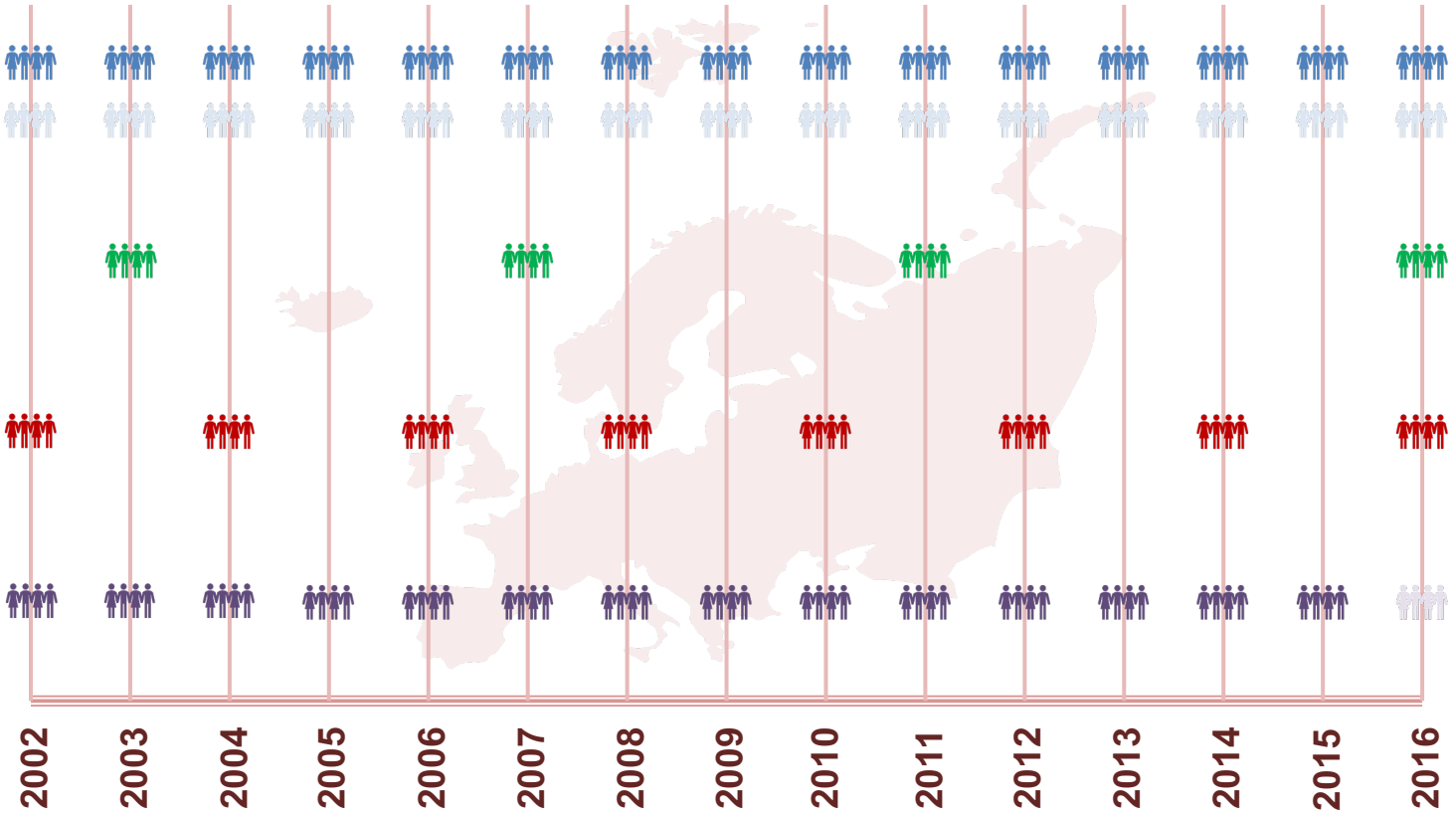
Data (1): scope of the sample quality assessment

EUROBAROMETER

Eurofound

European Social Survey

ISSP

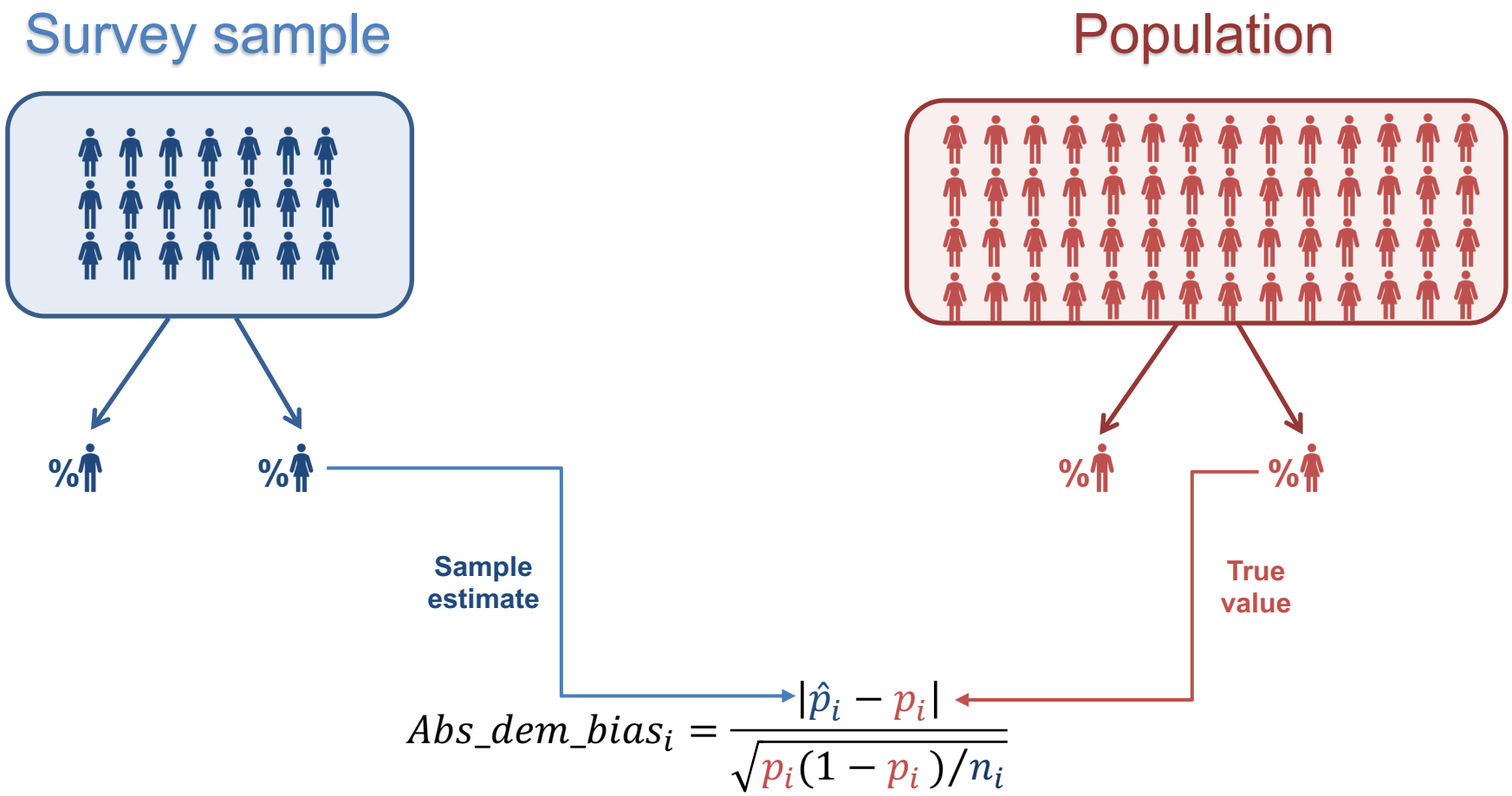


Data (2): scope of the sample quality assessment

EUROBAROMETER	15	462	412,594
Eurofound	4	125	129,964
European Social Survey	8	199	333,797
ISSP	14	339	265,415
	41	1,125	1,141,770

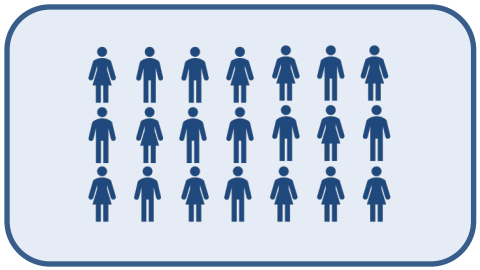
Methods

Methods (1): External criteria of representativeness (gender bias)



Methods (2): Internal criteria of representativeness (gender bias)

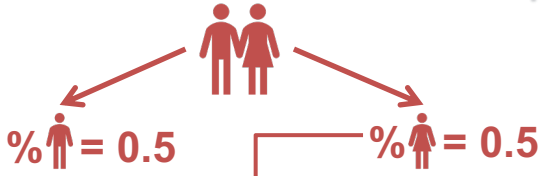
Survey sample



Subsample of people living in 2-person households of heterosexual couples

Population

Subpopulation of people living in 2-person households of heterosexual couples



Subsample estimate

True value

$$Abs\_unit\_nonresponse\_bias_i = \frac{|\hat{p}_i - 0.5|}{\sqrt{0.25/n_i}}$$

Methods (3): Variables & data for gender-based external and internal evaluation

External evaluation

1. Gender
2. Age

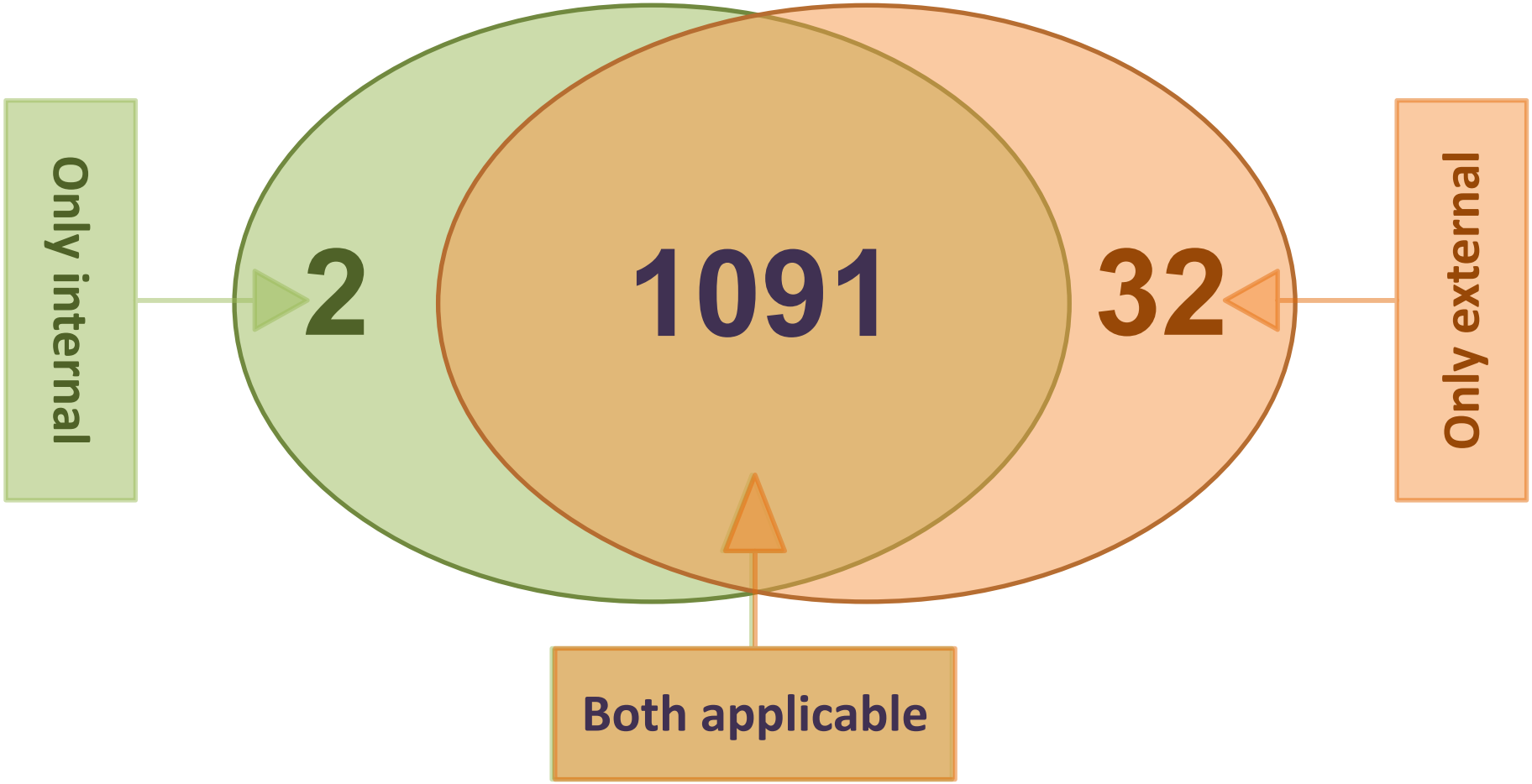
+

Population data on gender and age distribution, e.g., United Nations Population Division statistics

Internal evaluation

1. Gender	1. Gender
2. HH size	2. HH size
3. Living with heterosexual spouse or partner	3'. Respondent's marital status
Strict approach EQLS, ESS & post-2010 ISSP	**Lanient approach** EB & pre-2010 ISSP

Methods (4): Applicability of gender-based internal and external evaluations

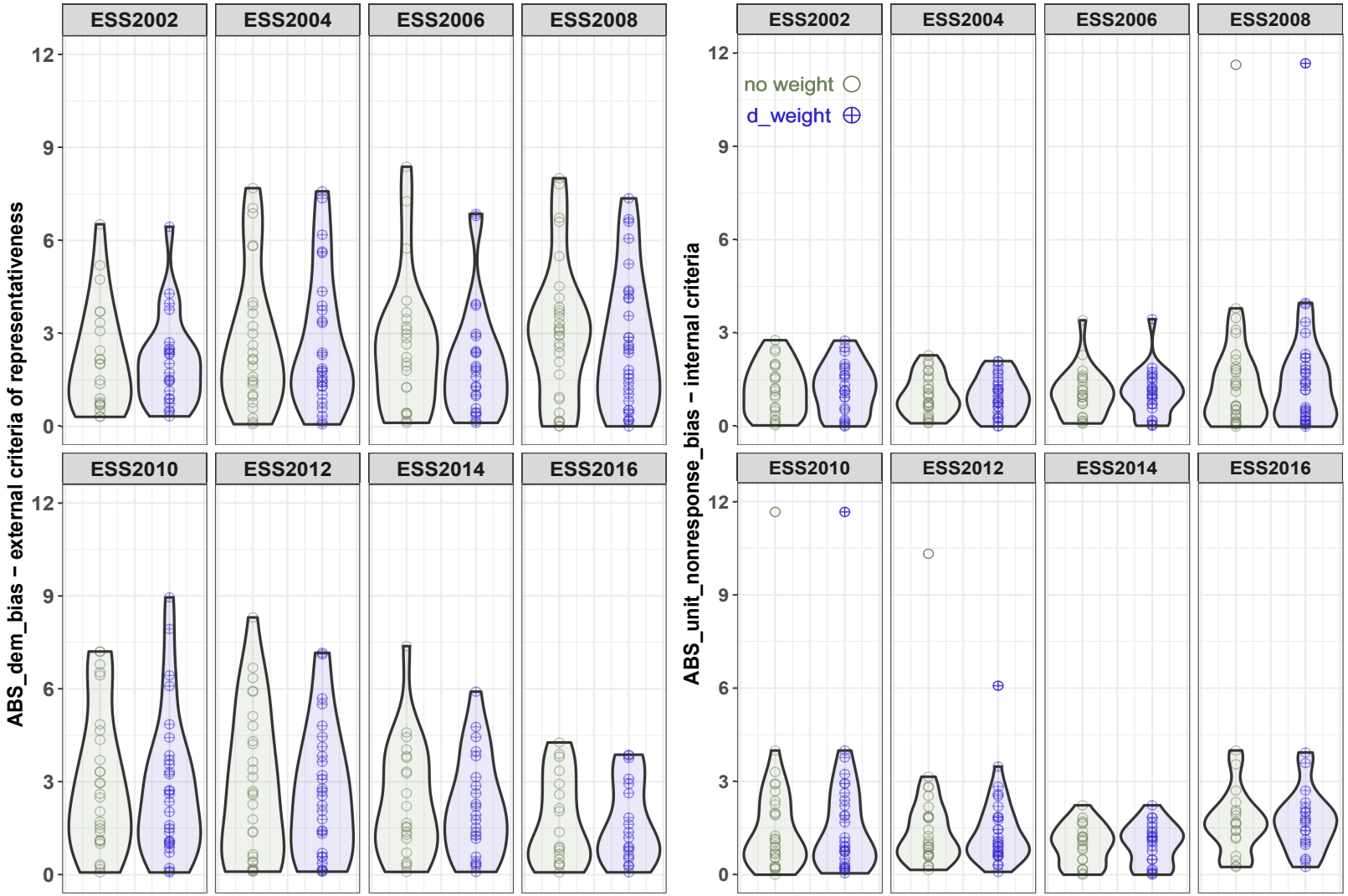


Results

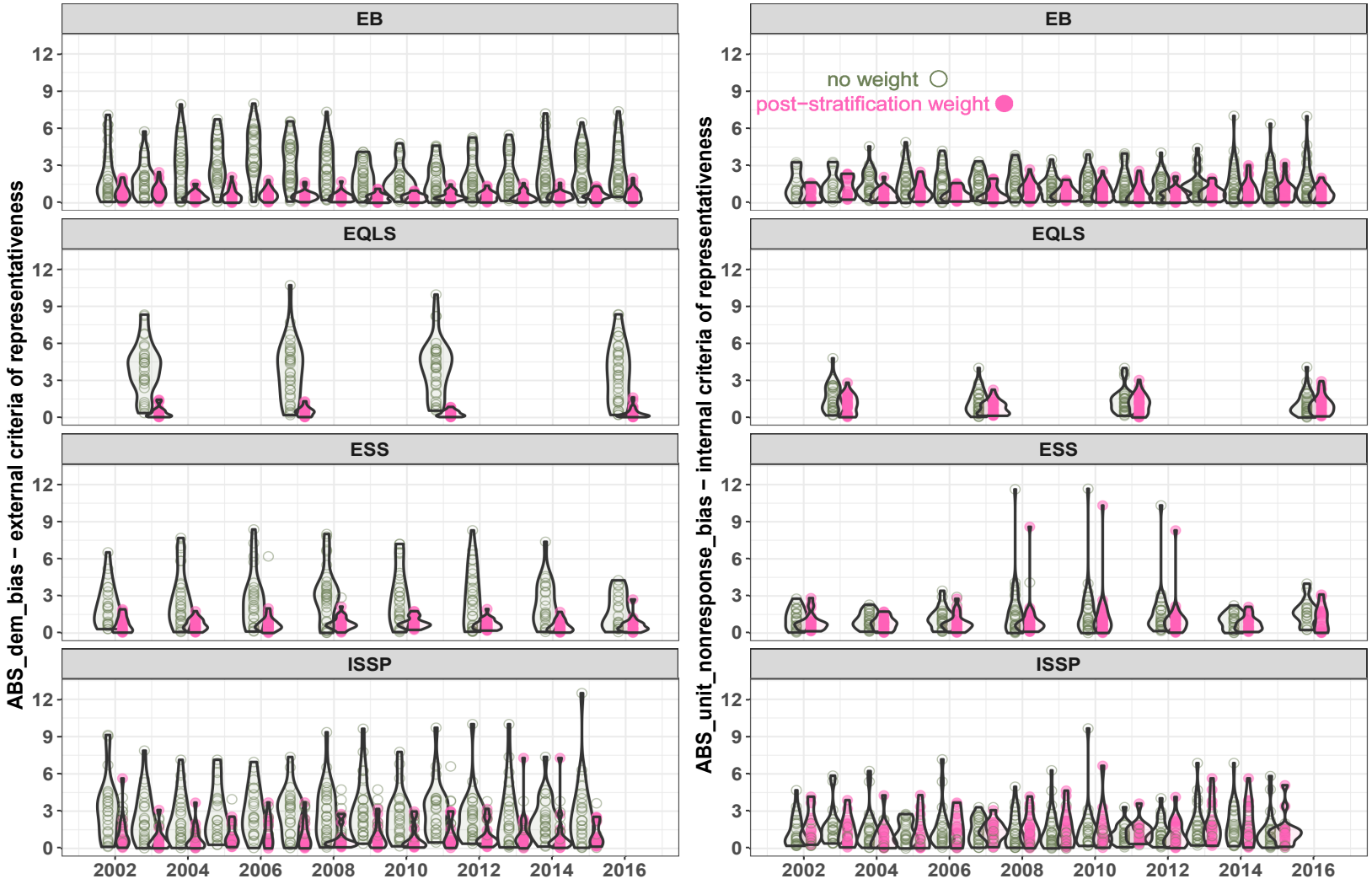
Results (1): heterogeneity of weights

Project	Design weights *dweight*	Poststratification weights *psweight*	Population size weights	Standardisation of weighting procedures
EB	No	Yes	Yes	Yes
EQLS	No: EQLS 1&2 Yes: EQLS 3&4	Yes EQLS 3&4: *psweight* combined with dweight	Yes	Yes
ESS	Yes / No weights: 3 out of 199	Yes *psweight* combined with dweight	Yes	Yes
ISSP	No	Yes / No weights: 132 out of 339	No	No

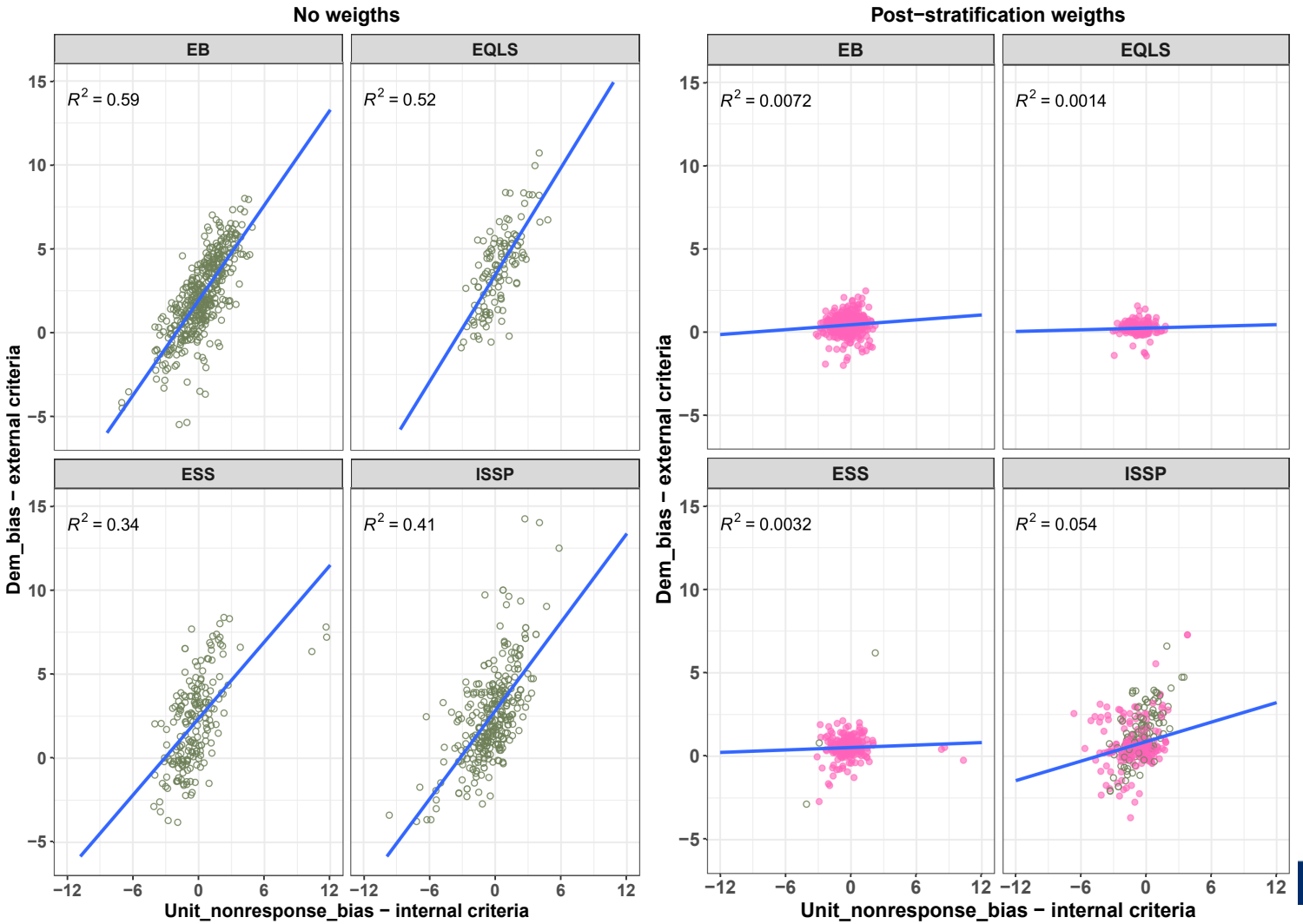
Results (2): impact of design weights



Results (3): impact of post-stratification weights



Results (4): correlation between external and internal evaluation



Conclusions

Challenges with incorporating weights into external and internal evaluation

1. Assessments of sample quality in terms of both external and internal criteria should be performed on data weighted by design factors (corrections for the unequal probabilities of selection stemming from sampling design);
2. Most cross-national survey projects do not provide separate variables with design factors;
3. Most cross-national surveys provide some kind of post-stratification weights; however, these must not be used in assessments of sample representativeness based on gender distribution;
4. Since design weights are routinely unavailable and application of post-stratification weights distorts sample assessments, it seems reasonable to consider the possibility of abstaining from weighting altogether.