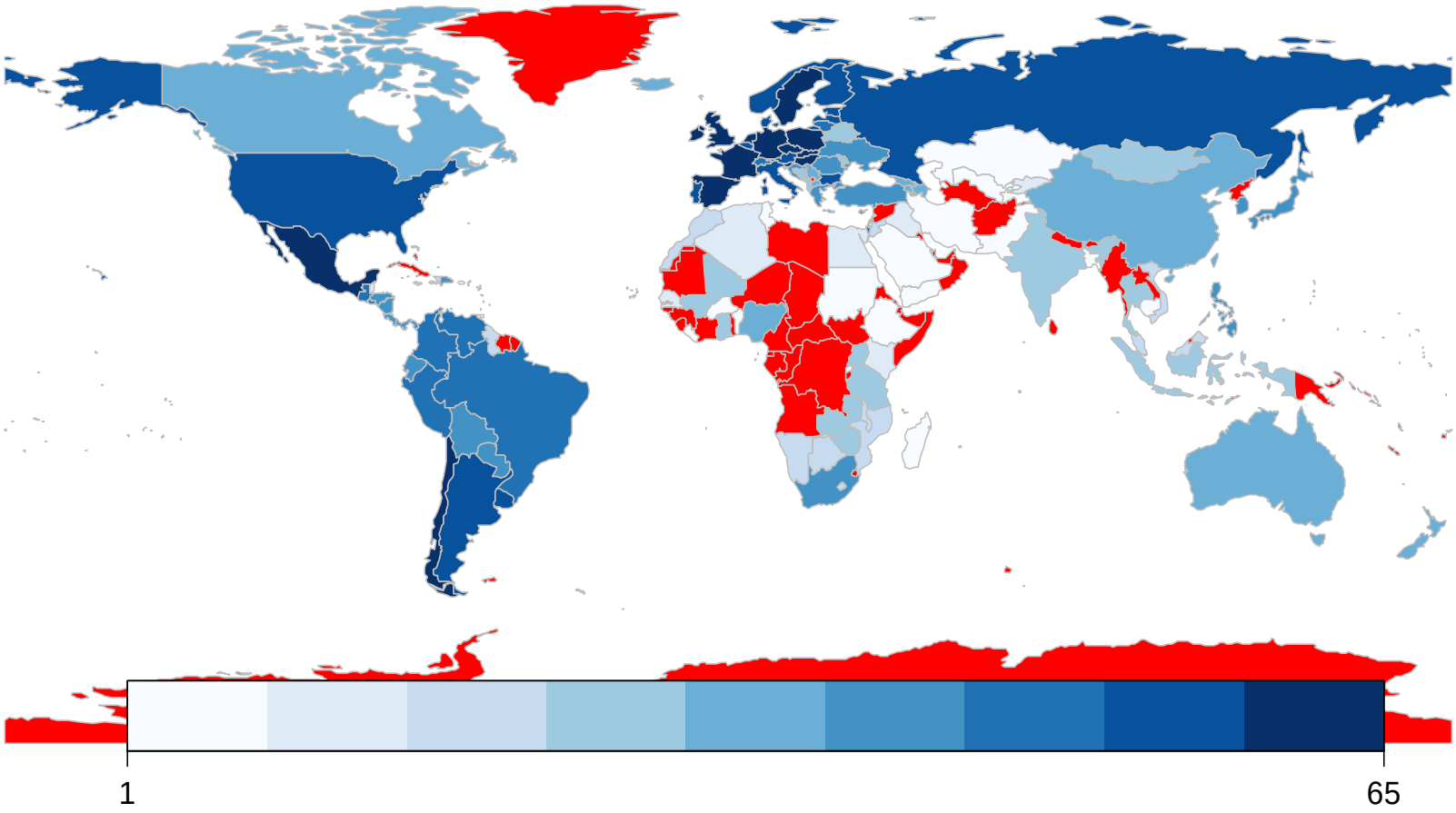
Weighting Data in Ex-Post Harmonization Process: The Consequences of Different Practices of Survey Organizations for Comparative Studies

Marcin W. Zieliński
Polish Academy of Sciences
University of Warsaw
Przemek Powałko
Polish Academy of Sciences

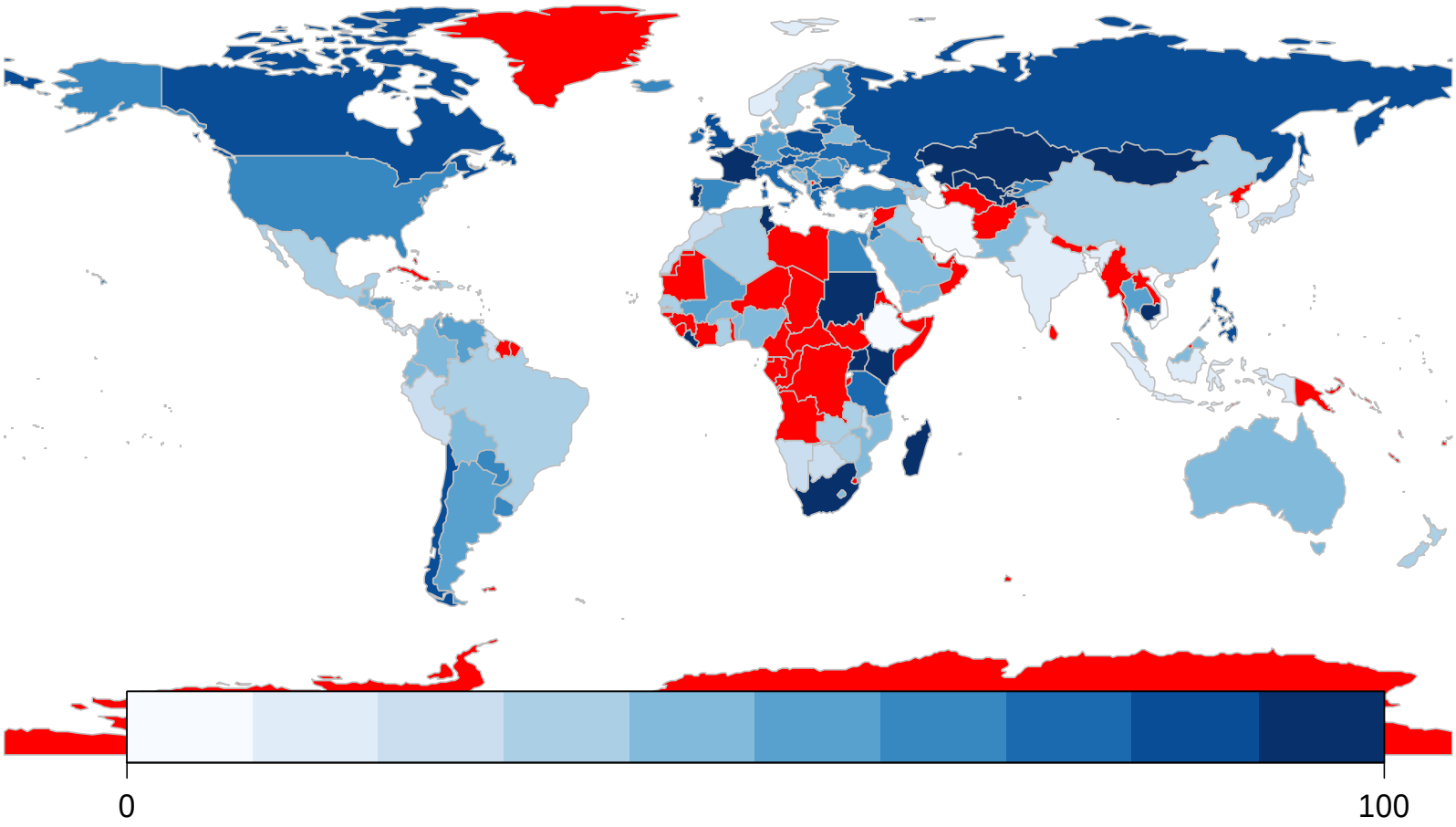
About the project…

- 22 international projects
- 130 countries
- years: 1966-2013
- 1721 studies

No of studies per country



Weights: avlbl in 1035 studies (60%)



Weights in time (cor=0,22)

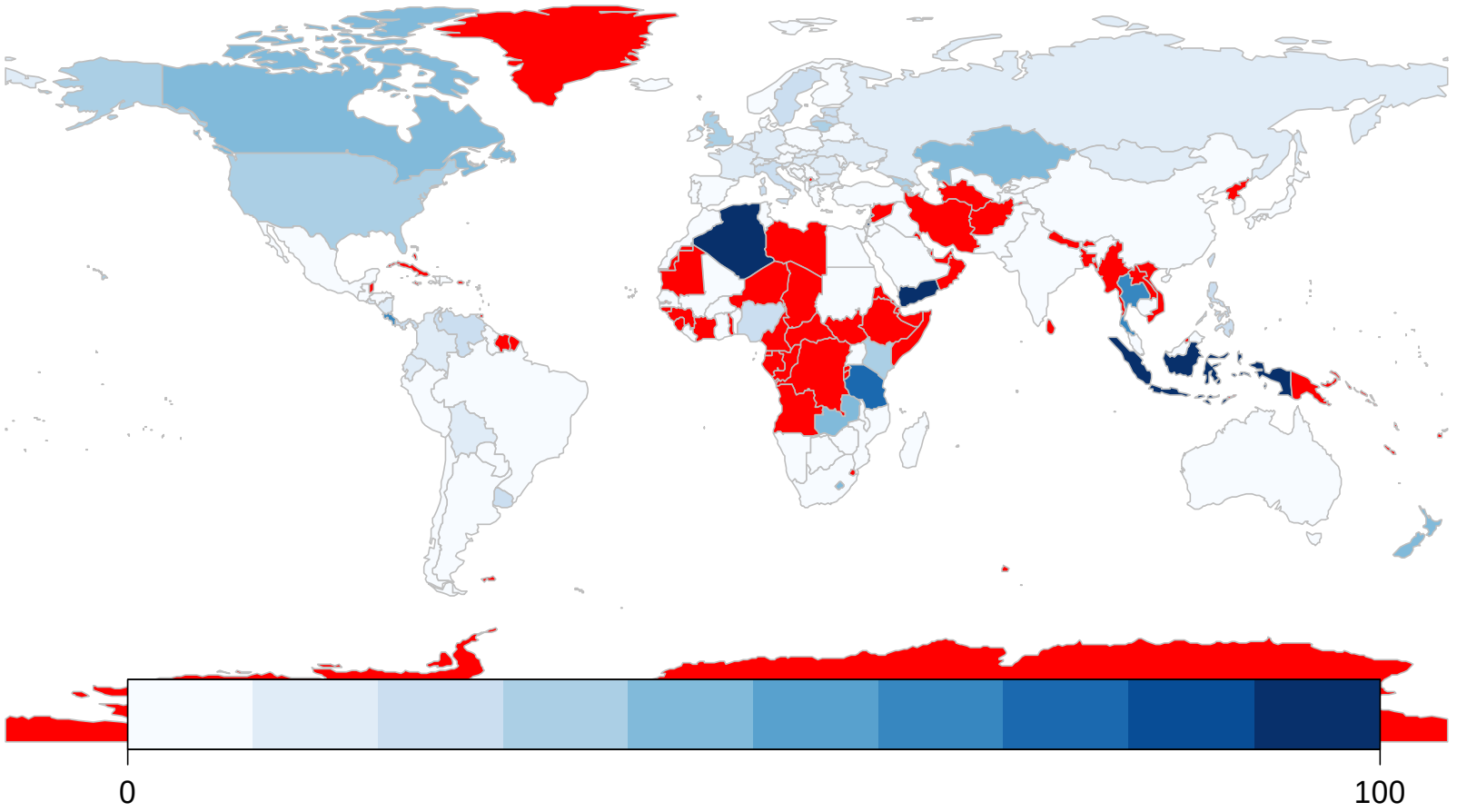
Properties of weights

- mean
- standard deviation
- minimum
- maximum

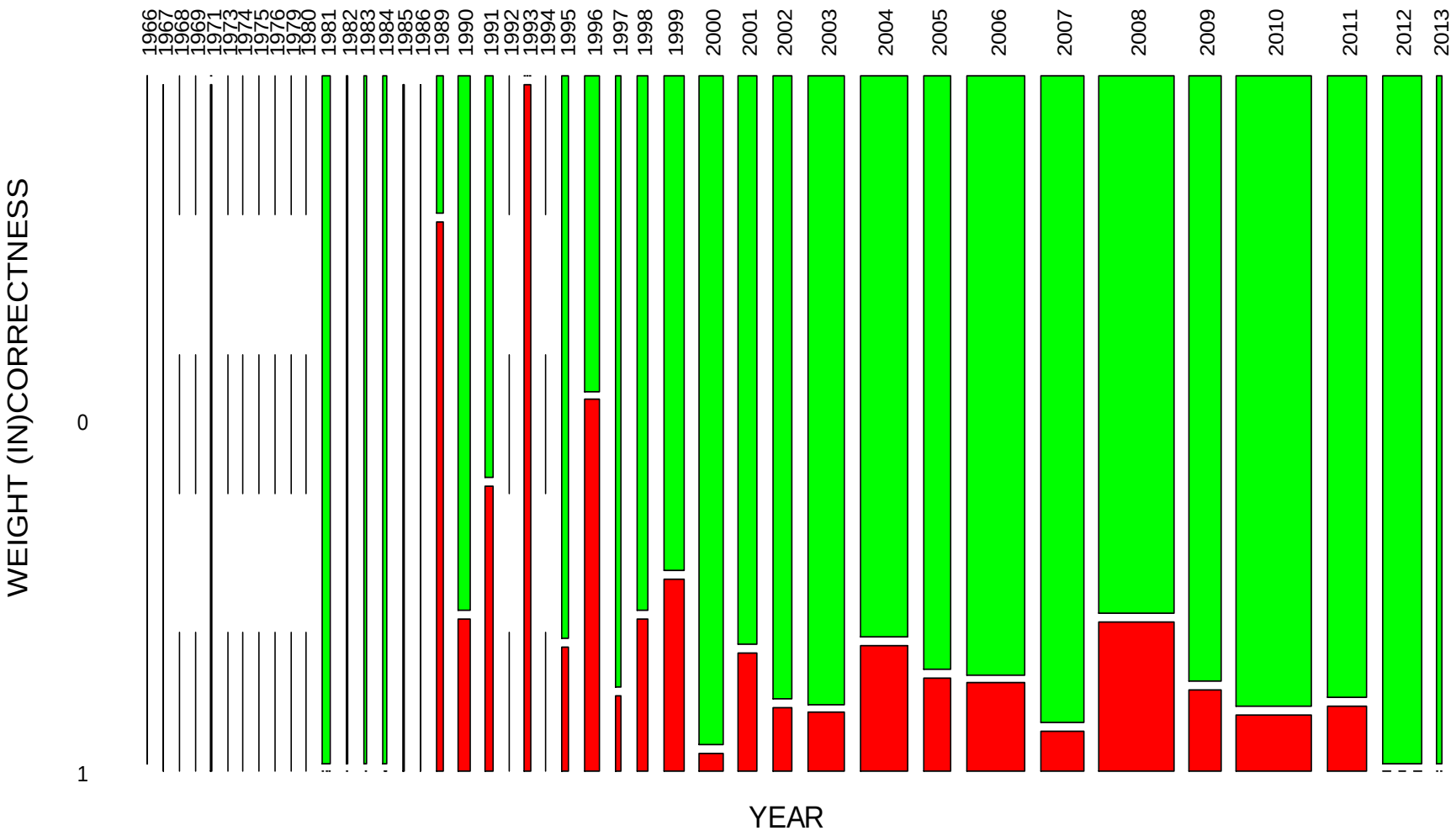
Mean of the weights

- Incorrect weights mean*(wght)* ≠ 1: 70% of studies
- Assuming tolerance of error <0,001: 15% incorrect
- Range mean(wght):
  - 3,29 (Philippines, ISSP 1996)
  - 0,83 (Philippines, ASB 2010)

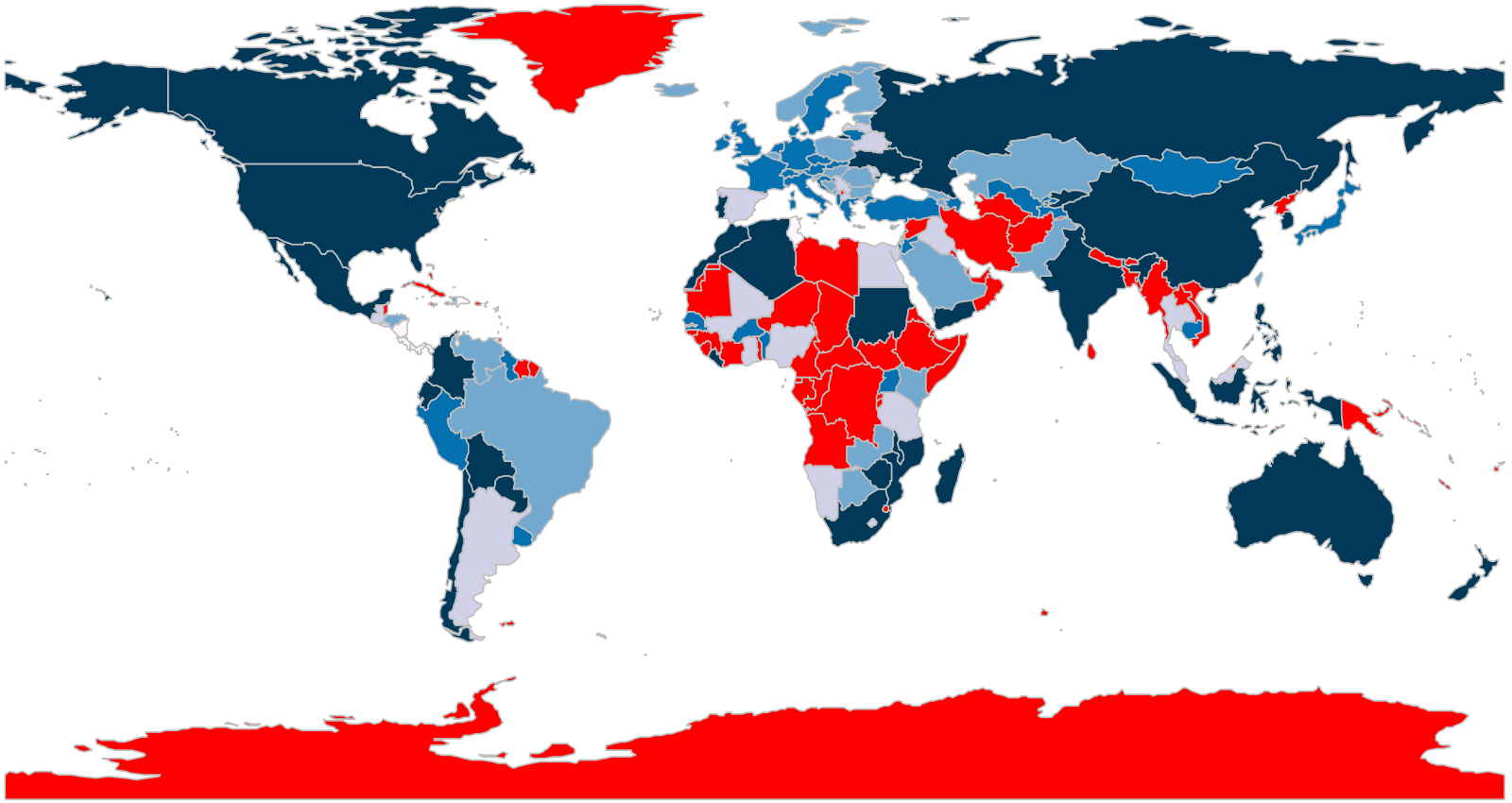
wghts ≠1 (with tolerance) per country (%)



wghts ≠ 1 (with tolerance) per year (cor=-0,20)



mean(std) per country



MIN; MAX

MINIMUM (range):

- 0.000 (42 studies)
- 1.900 (Philippines, ISSP 1991)

MAXIMUM (range):

- 0.921 (Lithuania, NBB 2001)
- 90.320 (New Zealand, ISSP 2007)

Weight components

- Poststratification factors:
  - Gender (70%)
  - Age (57%)
  - Country region (42%)
  - Education (36%)
  - Other (42%)
- Design factors:
  - Household+other factors (25%)

Summary of results

- Weighting is a common practice
- Data are weighted using different procedures
- Some weights have errors or are suspicious
- Weights are constructed using different factors

What to do?

- Calculate new weights using external sources of information (like e.g. UN data)

Advantages:

- The same factors taken into account
- Avoiding mistakes

Disadvantages:

- Practically weighting only by age and gender
- Loosing information about design weight factor
- Loosing information about other factors

What to do?

An example:

AFB, 2000, Republic of South Africa:

weight components:

province, race, gender, residental area, language (among Whites), housing type (among Blacks)

What to do?

- Do we know better than the authors of the study what factors should be included? Leave as it is?

Advantages:

  - Taking into account authors perspective
  - Design component where it was available

Disadvantages:

  - Comparability problem