



The Netherlands Institute for Social  
Research



## Q2010 Training Courses

Cross-national surveys

Sampling in a comparative  
perspective

Based on Sabine Häder and  
her sampling team



# Sampling

- Principles
- Sample size
- Sampling frames
- Sampling designs
- Sampling frame and contactability
- Effective sample size
- Design effects
- ESS procedures



# Sampling principles

Identical population definition

Complete coverage

- 15+
- Institutional population
- Non-native language speakers
- Areas (Isle of Man)

Probability sampling

- Known probability greater than zero of being selected
- No substitution

Effective sample size



## Sample size

European sample (proportional)

- Don't bother about Malta and Slovenia
- Not in cross-national surveys

Cross-national comparison (equal)

- Precision of estimate of differences maximal when sample sizes are equal

Regions (sort of proportional)

- Large countries have many regions, small countries may be one region



## ESS R2 sampling frames (Häder and Lynn)

Country	Frame	Remarks
Austria	Telephone book	Additional non-telephone households sampled in the field
Belgium	National register	
Czech Republic	Address register UIR-ADR	Select streets, followed by field enumeration
Denmark	Danish Central Person Register	
Estonia	Population register	
Finland	Population register	
France	None	Area-based sampling
Germany	Registers from local residents' registration office	
Greece	None	Area-based sampling
Hungary	Central registry	
Iceland	National register	
Ireland	National Electoral Register	
Luxembourg	Social Security Register	
Netherlands	Postal address list	
Norway	Population register	
Poland	National register of Citizens	
Portugal	None	Area-based sampling
Slovakia	Central Register of Citizens	
Slovenia	Central Register of Population	
Spain	Continuous Census	
Sweden	Register of population	
Switzerland	Telephone register	
Turkey	Cluster of addresses	
UK	Postal address list	
Ukraine	None	Area-based sampling

 Simple random sampling, or simple stratified sampling (no design effect)



# Sampling design

Simple random sampling

Systematic random sampling

Stratified unclustered random sampling

Clustered random sampling (stages)

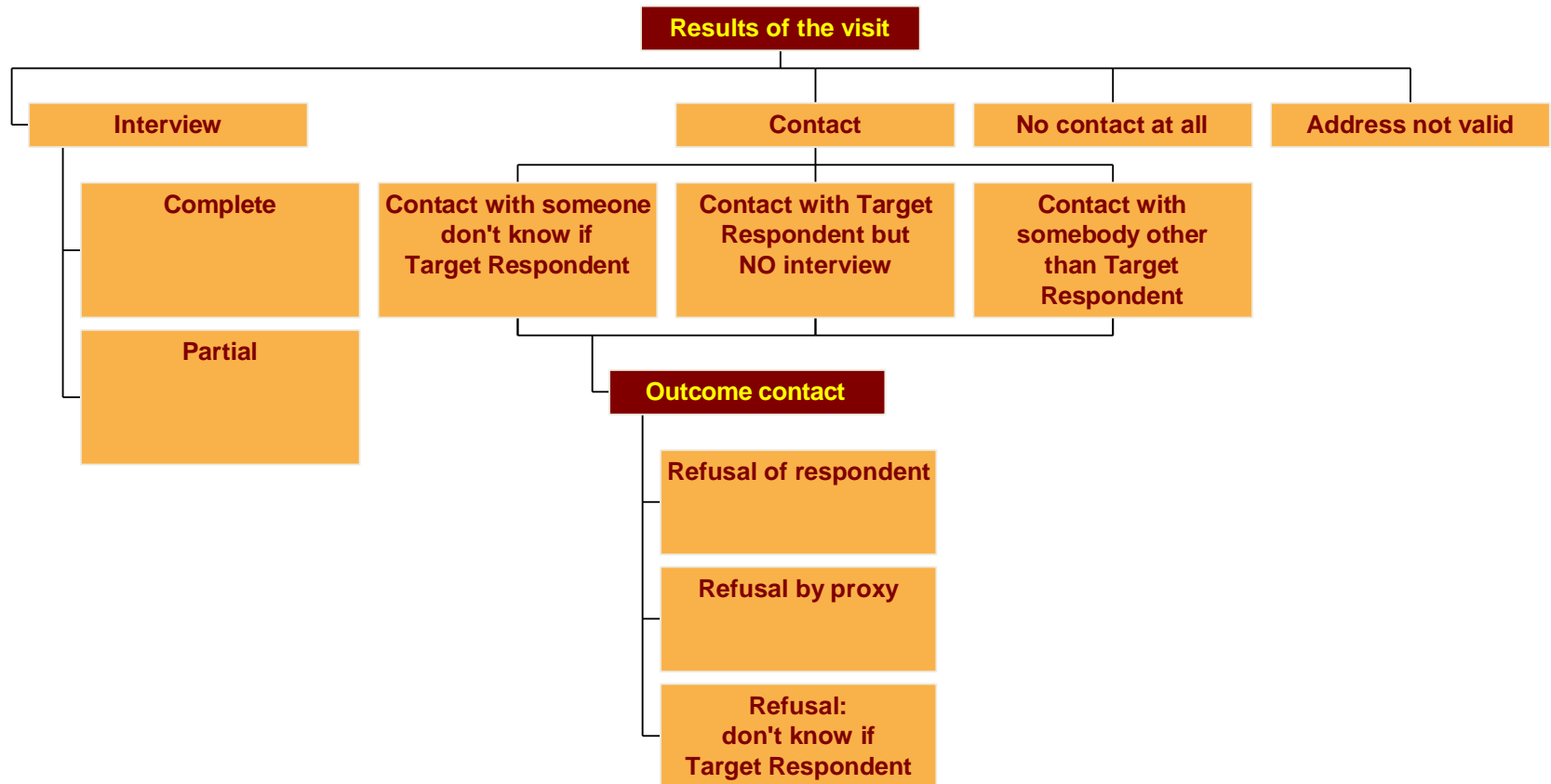
- Communities, streets, addresses, persons

Three types of sampling frames:

- Address, household, individual
- Address and household: selection of target respondent in the field

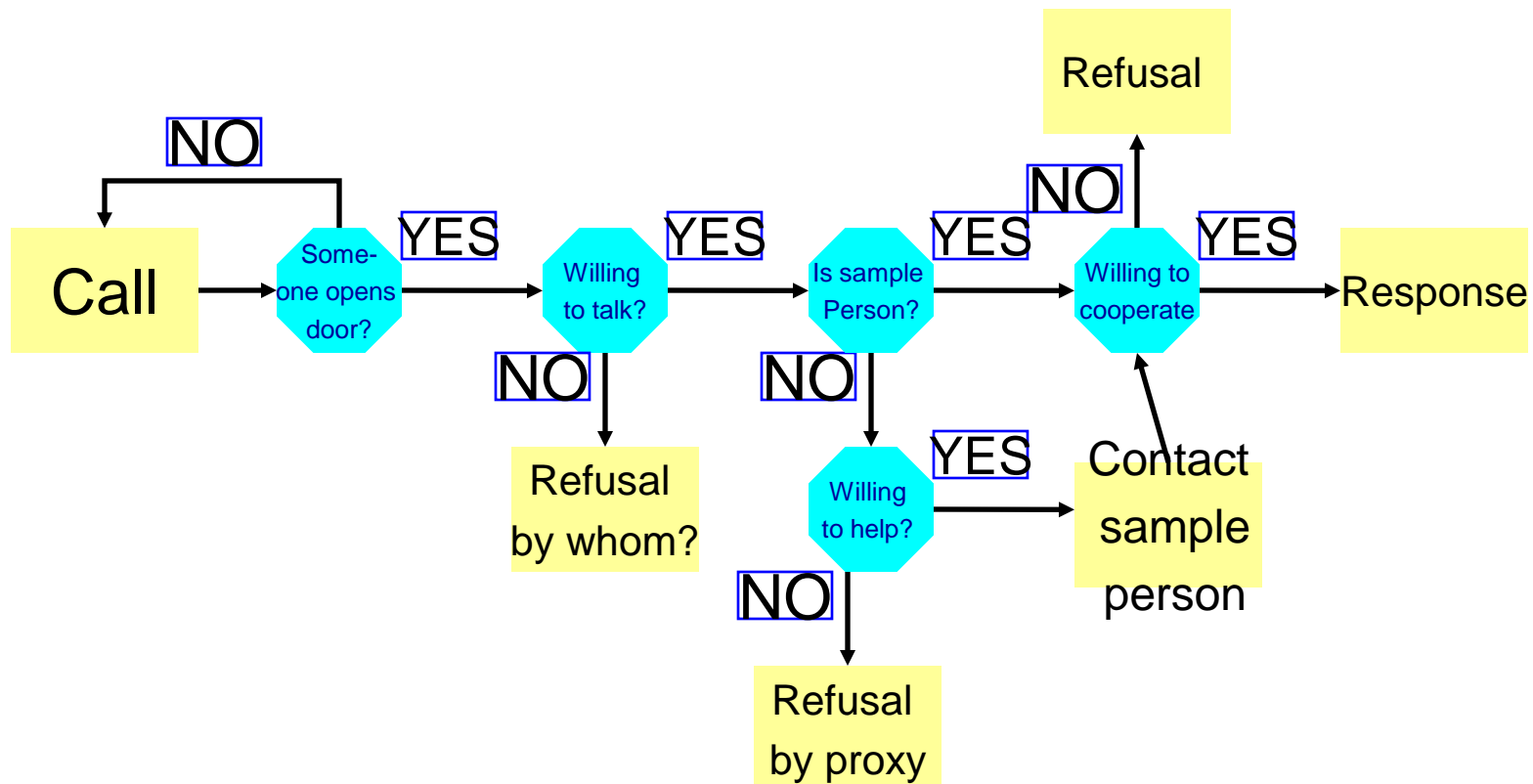


# R2 individual sample



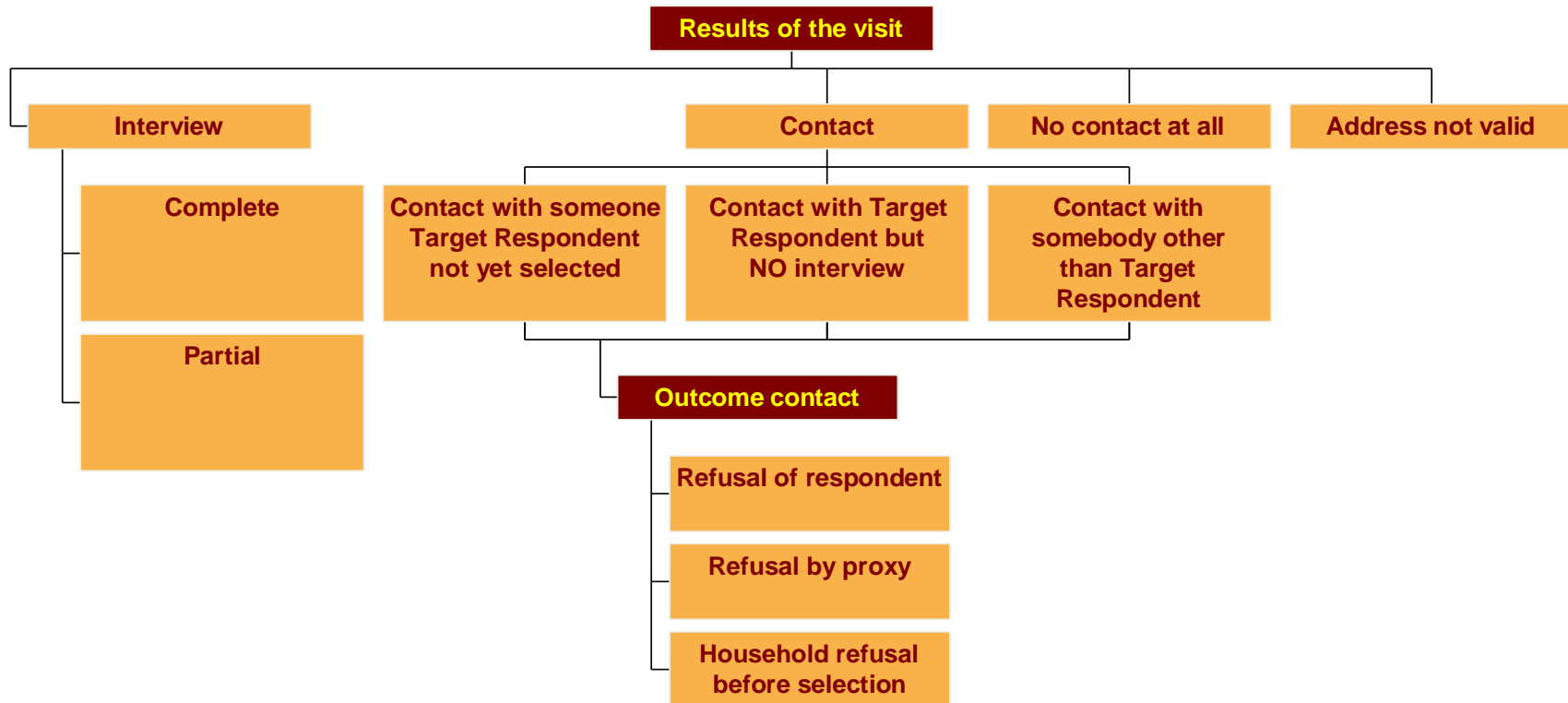


## F2f, individual sample



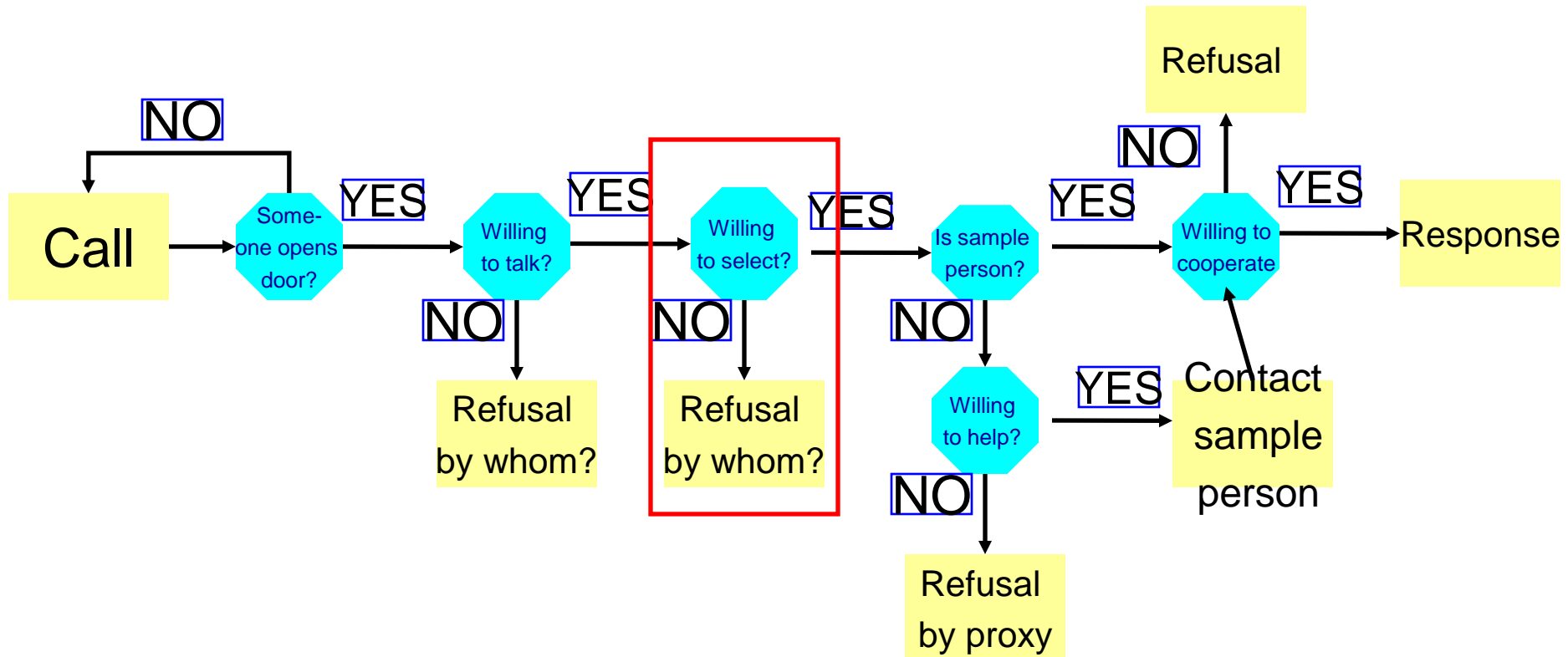


# R2 address/household sample





# F2f, household or address sample





## Proxy rate

In household/address samples more refusal by proxy

$$N_{\text{proxy}}/N_{\text{refusals}}$$

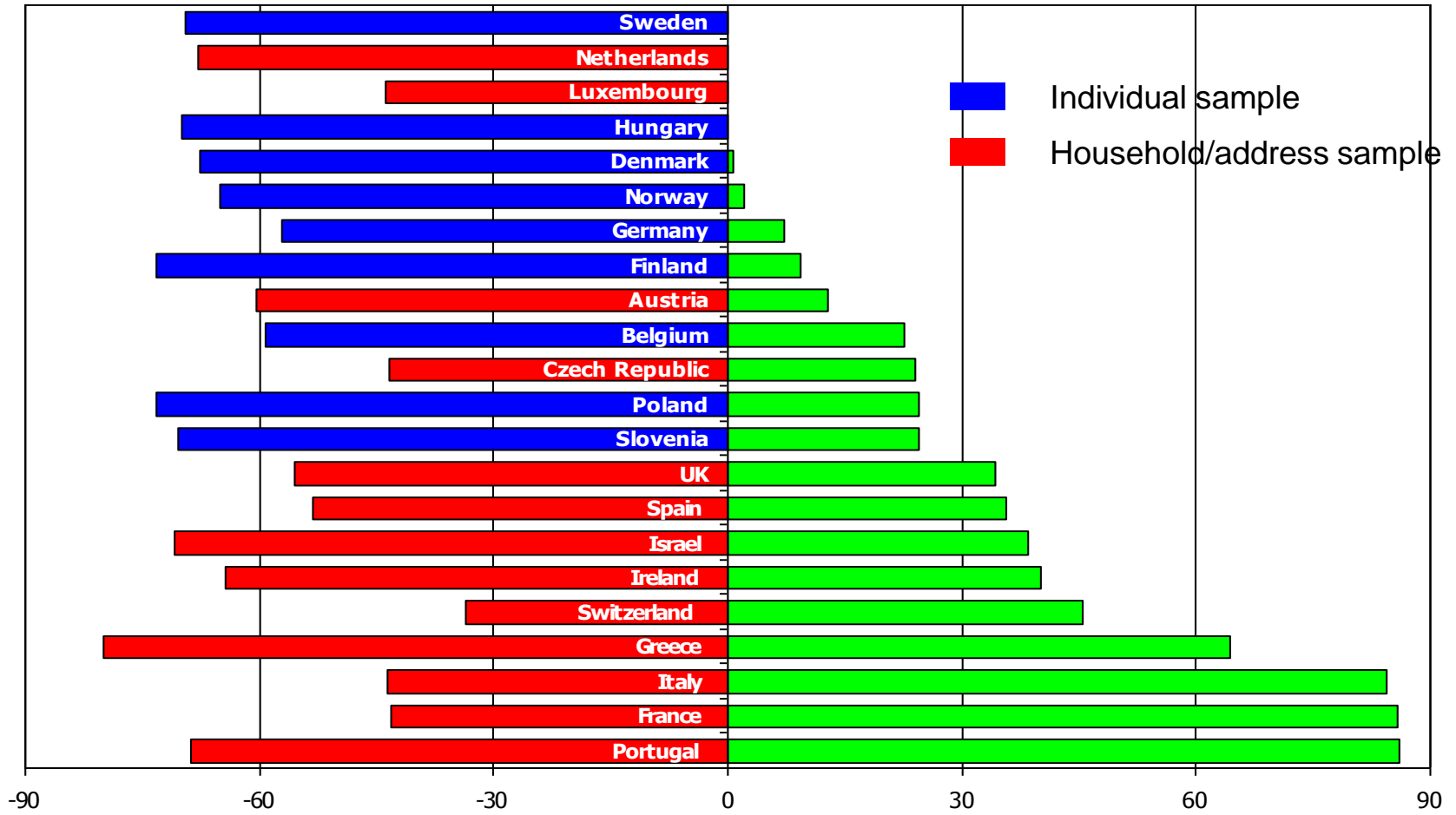
or

$$(N_{\text{refusals}} - N_{\text{refusals target respondent}})/N_{\text{refusals}}$$

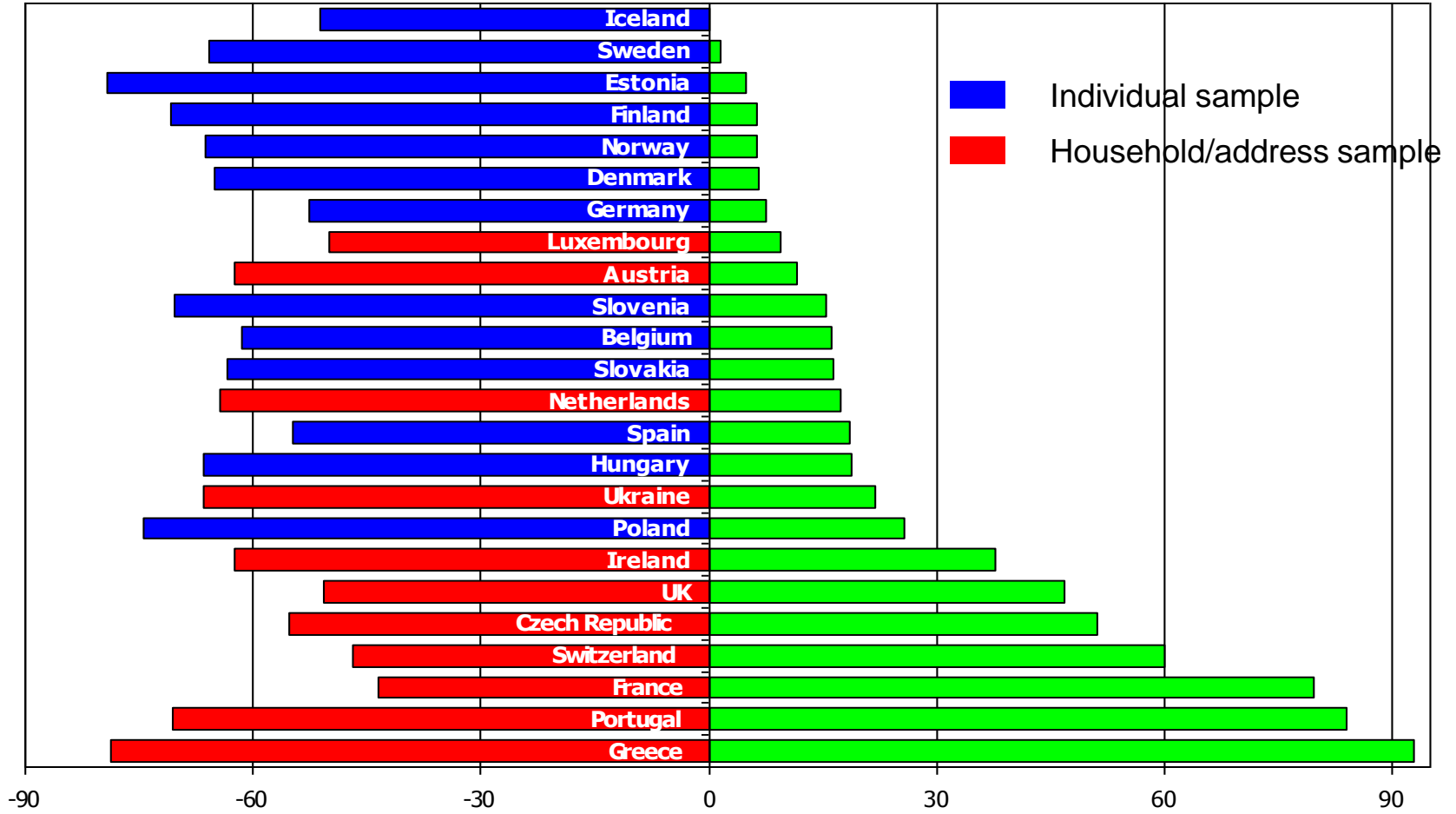
$N_{\text{proxy}}$ :

Number of refusals by proxy or by household member that may or may not be Target Respondent or before selection Target Respondent

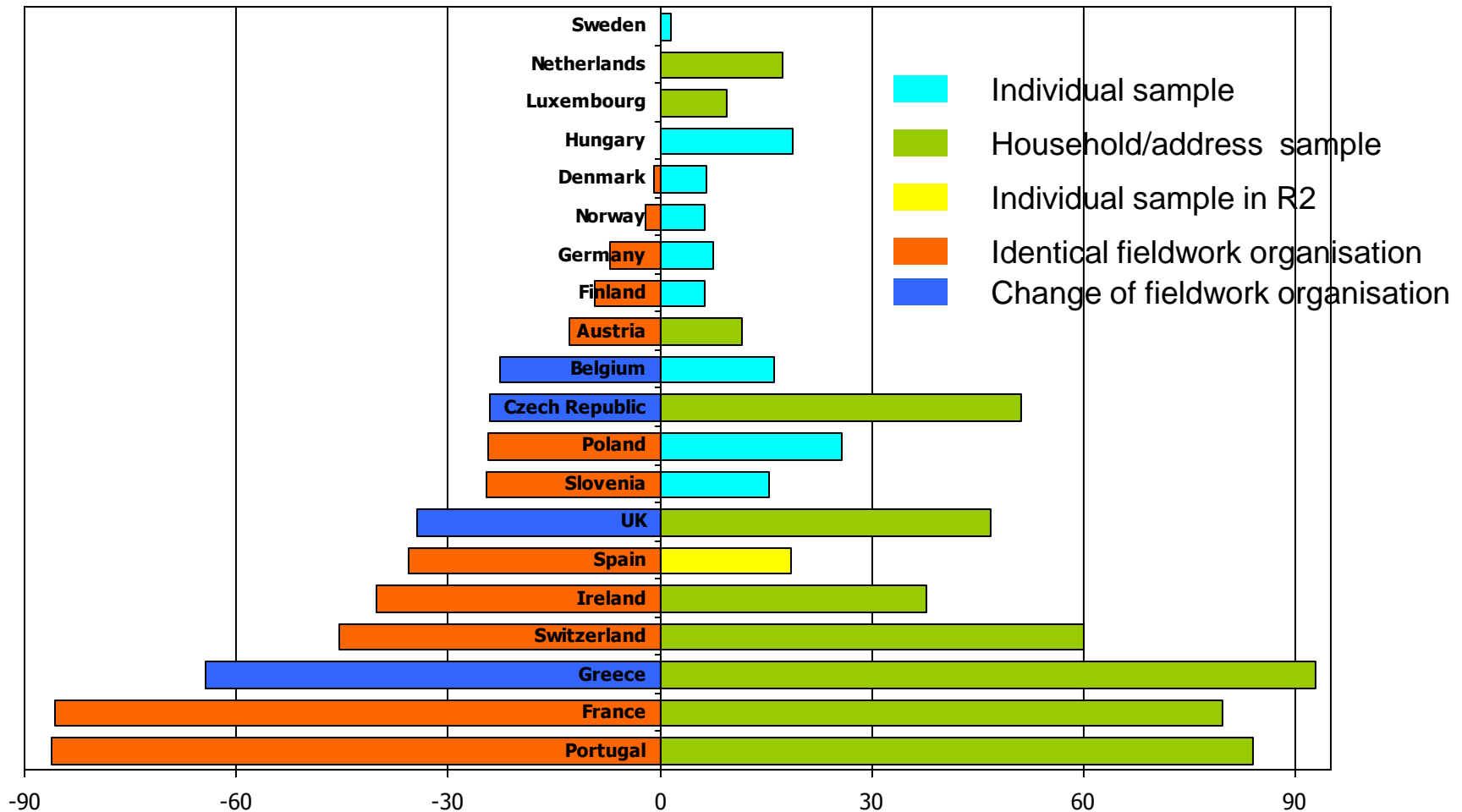
# Response rate (l) and proxy rate (r) ESS-R1



# Response rate (l) and proxy rate (r) ESS-R2



# Proxy rate R1 (I) and R2 (r)





## Effective sample size (Neff)

Size of a simple random sample which would produce the same precision (standard errors) as the design actually used.

Usually less than the actual number of achieved interviews as certain aspects of survey design

- Clustering
  - Use of differing selection probabilities
- tend to reduce the precision of estimates.

Reduction of precision is known as the design effect (DEFF)



## DEFF

DEFF = Actual sampling variance / Sampling variance with simple random samples of same size ( $m$ );

DEFF =  $m/\text{neff}$ , so  $\text{neff} = m/\text{DEFF}$

Value of DEFF for a proposed sample design has to be predicted, in order to determine how many interviews should be achieved so as to produce a particular value of  $\text{neff}$ .

Also expected response rate should be predicted in advance

DEFF depends on design effect arising from differing selection probabilities (DEFF<sub>p</sub>) and the design effect arising from clustering (DEFF<sub>c</sub>) .

DEFF = DEFF<sub>p</sub> x DEFF<sub>c</sub>.





## Differing selection probabilities

$$DEFF_p = \frac{m(\sum_i m_i w_i^2)}{(\sum_i m_i w_i)^2}$$

where there are  $m_i$  respondents in the  $i$ th selection probability class, each receiving a weight of  $w_i$ , where

$$w_i \propto \frac{N_i}{m_i}$$

where  $\propto$  means 'proportional to'

DEFF<sub>c</sub>



Multistage cluster sample (b – cluster size)

$$DEFF_c = 1 + (b-1) \rho$$

*Estimate of  $\rho$  0.02 (can differ for each variable)*



## Actual design effects ESS R2 (Ganninger)

Country	$\rho$	$deff_c$	$deff_p$	$deff$	N
Austria	0.10	1.47	1.25	1.84	2556
Belgium	0.04	1.19	1.00	1.19	1778
Czech Republic	0.14	2.60	1.50	3.91	3026
Denmark		1.00	1.00	1.00	1487
Estonia		1.00	1.06	1.06	1989
Finland		1.00	1.00	1.00	2022
France	0.04	1.36	1.19	1.62	1806
Germany	0.06	2.01	1.11	2.23	2870
Greece	0.10	1.36	1.20	1.64	2406
Hungary	0.05	2.41	2.16	5.21	1498
Iceland		1.00	1.00	1.00	581
Ireland	0.08	1.87	1.31	2.44	2286
Luxembourg		1.00	1.15	1.15	1635
Netherlands		1.00	1.20	1.20	1881
Norway		1.00	1.00	1.00	1761
Poland	0.07	1.57	1.01	1.59	1717
Portugal	0.16	1.88	1.39	2.61	2052
Slovakia		1.00	1.00	1.00	1512
Slovenia	0.05	1.39	1.00	1.39	1325
Spain	0.11	1.40	1.01	1.41	1663
Sweden		1.00	1.00	1.00	1948
Switzerland	0.04	1.25	1.22	1.53	2141
UK	0.04	1.34	1.26	1.69	1897
Ukraine	0.16	1.95	1.73	3.38	2031



# ESS sampling procedures

## Central specifications

- Sample size
- Effective sample size
- Random sampling
- Population definition

## Sampling panel

- Sabine Häder, Peter Lynn, Siegfried Gabler, Seppo Laaksonene

## Assigned to individual countries

Help and advise on sampling frames

Sign off sampling procedure

Compute actual design effects



# Sampling heaven?

