HBM4EU

science and policy
for a healthy future

Taking a representative sample for all age groups

Hanna Tolonen

National Institute for Health and Welfare (THL), Finland

General sampling theory

Terminology related to sampling

- Target population
- Sampling frame
- Sample size
- Eligibility
- Sample selection
- Representativeness of the sample

Target population

”The collection of individuals, items, measurements, etc., about which inferences are desired. The term is sometimes used to **indicate the population or group from which a sample or study population is drawn** and sometimes to denote a reference population about which inference is desired.”

- Porta (2008) A Dictionary of Epidemiology

Sampling frame

- Sampling frame is a list of all units/individuals of the target population.
- Characteristics of a sampling frame
  - Updated regularly
  - Includes everyone in the target population
  - Contains contact information

Examples of sampling frames

- Population register
- Census
- Patient register
- List of employees of a company
- Post code address files
- Pupil/Student lists
- Maps

Sample size

- Number of individuals to be invited to the survey
- For determination of sample size
  - *p* = Estimated prevalence of outcome of interest in the population
  - *e* = margin of error/ precision of the estimate
  - *z* = the z score representing the desired level of confidence/probability of error

are needed.

$$\mathrm{N} = \frac{Z^2 * (p) * (1 - p)}{e^2}$$

Example of sample size calculation

- Estimated prevalence: p = 10%
- Precision of the estimate: e = 4%
- 95% confidence level: z = 1.96

$$\frac{1.96^2 * (0.1) * (1 - 0.1)}{0.04^2} = 216$$

Note! Sample size calculation needs to be adjusted for expected response/participation rate.

If 216 persons are expected to participate and response rate is expected to be 60%, 360 persons need to be invited.

Eligibility (inclusion/exclusion criteria)

- Depends on aim of the study
- Needs to be defined before sample is selected and study started
- Criteria may include
  - practical limitations such as language skills, access to the person (prisoners, nursing homes, etc.)
  - death before examination date
  - moving out of study area
  - research area (hotspots, regions, etc.)
  - baseline health condition/behaviours (smoking, etc.)
  - Etc.

Probability sample vs non-probability sample

Probability sample

Each subject of the target population has a known probability to be selected to the sample

Also called random sampling

Methods:

- Simple random sampling
- Stratified sampling
- Cluster sampling
- Systematic sampling

Non-probability sample

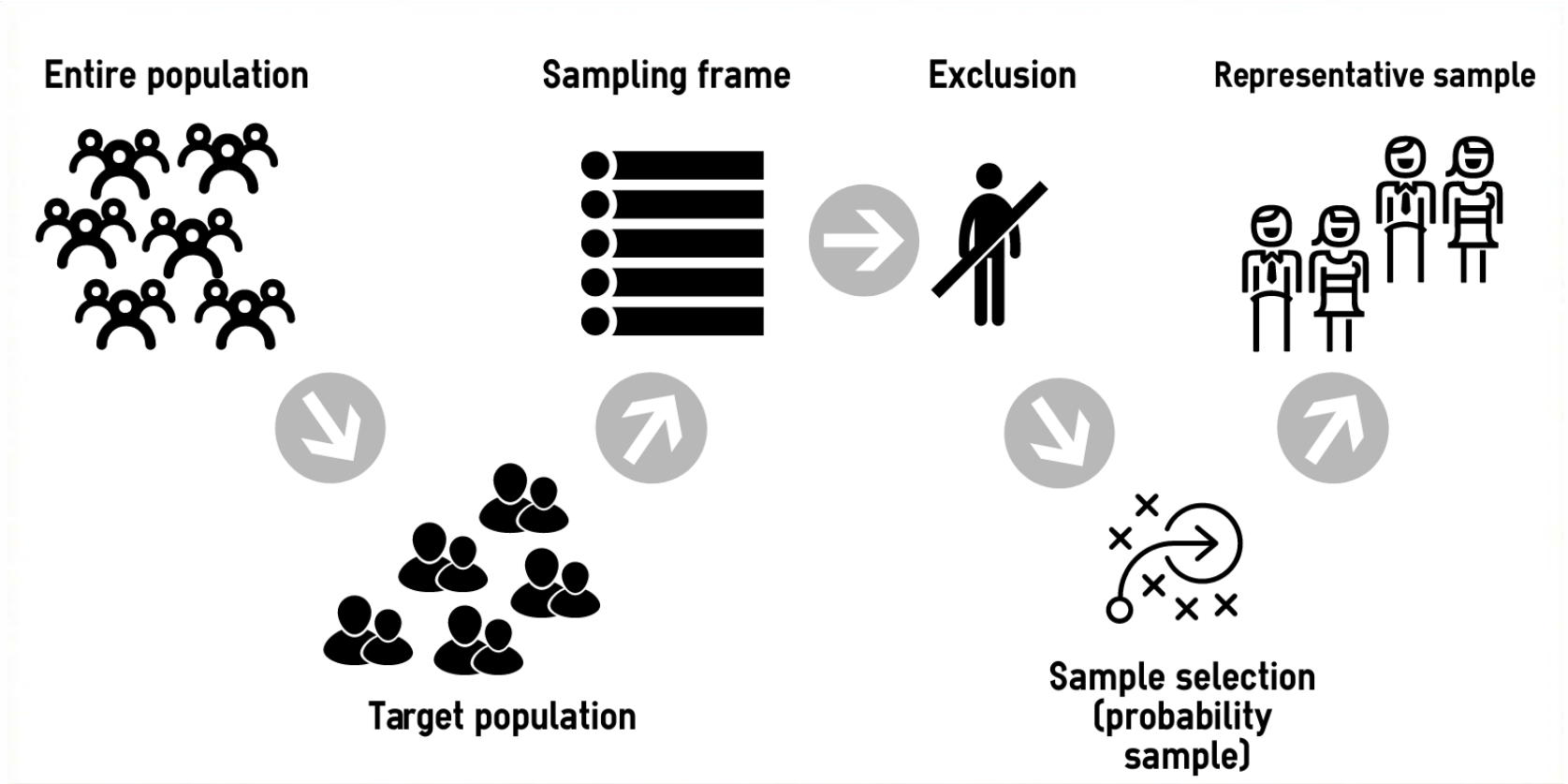
It is not known which individuals from the population are selected to the sample

Also called non-random sampling

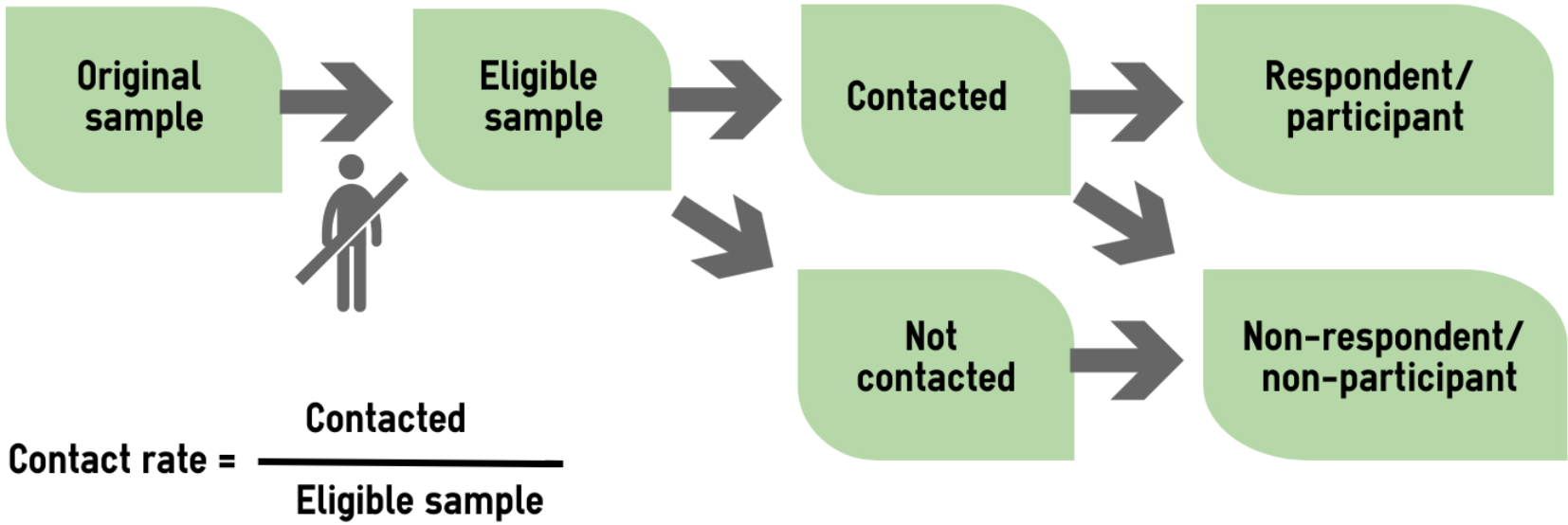
Methods:

- Convenience sampling
- Quota sampling
- Purposive sampling
- Snowball sampling

Obtaining representative sample



Indicators for representativeness of the sample



$$\text{Contact rate} = \frac{\text{Contacted}}{\text{Eligible sample}}$$

$$\text{Response rate (A)} = \frac{\text{Respondents}}{\text{Eligible sample}}$$

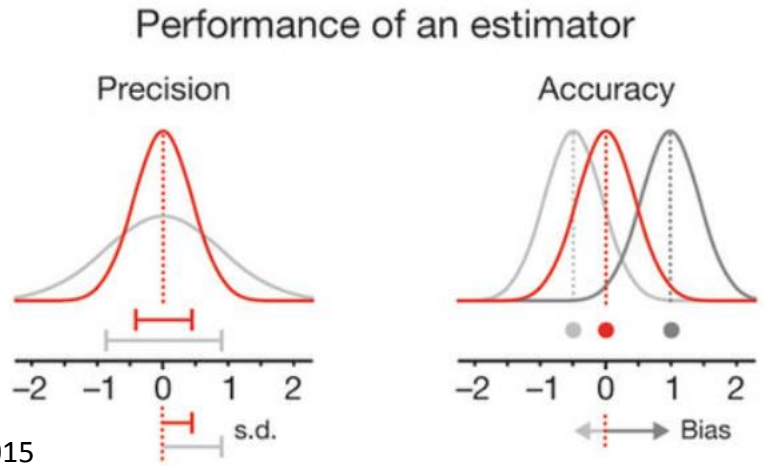
$$\text{Response rate (B)} = \frac{\text{Respondents}}{\text{Eligible sample - Not contacted}}$$

Representativeness – precision vs accuracy

Precision = variation of repeated measurements of the value of interest, i.e. how wide the sampling distribution is, which can be expressed as the standard deviation of the sampling distribution

Accuracy = difference between measured value and actual value in target population, i.e. how close the mean of the sampling distribution is to the mean of the population distribution (bias)

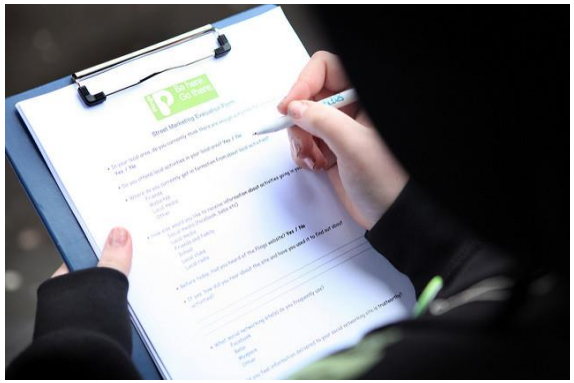
Increase in sample size, increases precision



Increase in response rate, increases accuracy

Kulesa et al. Nature Methods 2015

Obtaining high response rates



Questionnaire design

Timing and length of the visit

Incentives

Access to the examination clinic

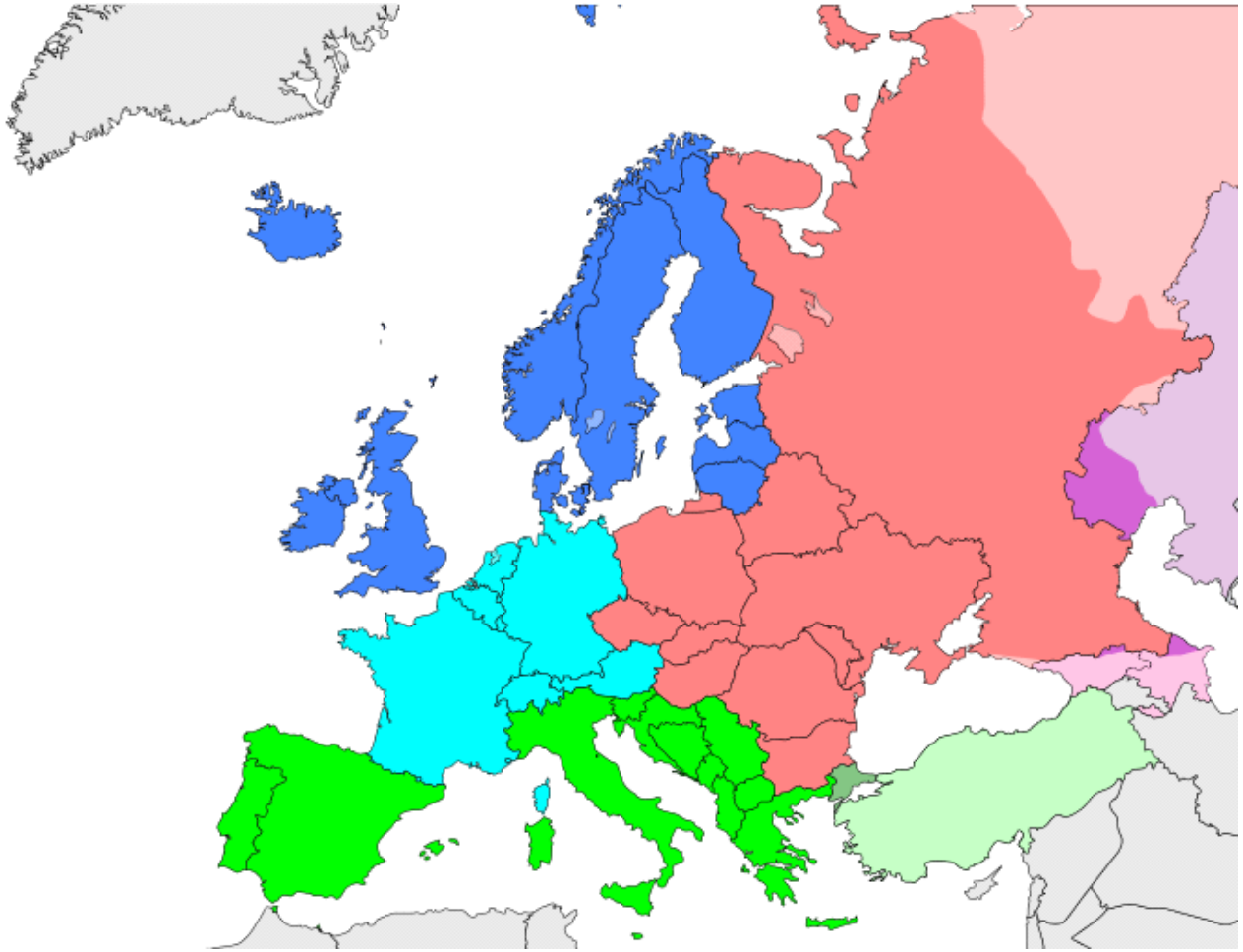
Sample selection for HBM4EU

Required samples within HBM4EU

Depending on the objective as formulated in scoping documents, different sampling approaches could be followed:

- Objective 1: Assessing the actual exposure of EU population or differences between EU countries
- Objective 2: Monitoring time trends in exposure
- Objective 3: Assessing the impact of policy

Proposed geographical regions



Northern Europe Western Europe Eastern Europe Southern Europe

(source: https://en.wikipedia.org/wiki/United_Nations_geoscheme_for_Europe)

Objective 1. Assessing the actual exposure in EU population

- A particular age group
- Organize sampling at national level (representative studies)
- Collect data in minimum of 3 countries per geographical region (N, E, S, W)
- 150 male and 150 female participants*

→ min. 3600 EU participants

*Based on sample size calculations

Objective 2. Monitoring time trends in exposure

- Homogeneous age group
- Country representative population or subpopulation from a subregion of a country
- At least one country or subregion per geographical region (N, E, S, W)
- Include 150 male and 150 female participants
- At least 3 time points (2 should be available already)

→ min. 1200 EU participants

Objective 3. Assessing impact of policy

- 3 (groups of) countries with different policy, i.e.:
  - No regulation
  - Phasing out
  - Complete restriction

N = 900 EU participants

- 1 country (or more): before and after a legislation

Theoretical concept: EU representative samples

! All sampling strategies are compromises

Domains for which reliable data are **needed:**

Age

- ✓ **Toddlers (0-3y)**
- ✓ **Children (3-6y)**
- ✓ **Children (6-11y)**
- ✓ **Teenagers (12-19y)**
- ✓ **Adults (20-39y)**
- ✓ **Elderly (60-?y)**

Sex

- ✓ **Male**
- ✓ **Female**

Geographical coverage

- ✓ **North**
- ✓ **East**
- ✓ **South**
- ✓ **West Europe**

Domains for which reliable data are **wanted**:

Education: ISCED-classification

- ✓ **UNESCO:**
- ✓ **Level 0-6**

Subject living environment

Inhabitants of

- ✓ **Low**
- ✓ **Medium**

- ✓ **high density communities**

Other?

- ✓ **SES**
- ✓ **…**

Age domains for which reliable data are needed

- Age groups considered in HBM4EU (~HealthCanada)

Younger children (3-5 years)
Children (6-11 years)
Adolescents (12-19 years)
Adults (20-39 years)
Adults (40-59 years)
Elderly (60-79 years)

Depending on the chemical, a different age group will be selected!

Reason not including younger/older age groups:

- ↑ logistic feasibility of the field work
- ↓ noise e.g. physiology may differ more within those age groups

Inclusion/exclusion criteria for HBM4EU studies

Exclusion:

Hotspot areas

Specific targeted populations e.g. patient groups

Inclusion:

General population

Harmonized sampling – uniform parameters between countries

Recruitment & population selection

- **Number individuals**: N= 300/age group/country (cf. sampling scheme task 7.2)
- **Age group**: which?
- **Gender**: Men and women
- **Participants:** 150/gender group/country
- **Smoking, Alcohol/drugs**: inclusion/exclusion
- **Socio-economic status**: -> ESCID (education) info to be collected in questionnaire
- **Residential history**: minimal 5 years living in country?
- **Geographical coverage**: distribution over country + selection of secondary sampling units (SSU) with a probability of being selected according to population size?

Collection of biological samples:

- **Matrices** collected
- **Time of sampling** (time of day, spot/repeated sampling)
- **Recipients**
- **Field blanks**
- **Sample transport** conditions

Harmonized sampling – non-uniform parameters between countries

Recruitment & population selection

- Restrictions on **literacy** (country language knowledge)
- **Life (style) factors:** e.g. diet, housing conditions, hobbies and occupation
- **Recruitment strategy and sampling locations** (via schools, work, registries)
- **Sampling time period** (within 2014- 2019, no seasonal restrictions)
- Links with other national HIS/HES programs

Sampling & field work

- **Ethical approval** strategies
- **Communication** with participants
- Field workers training
- Biobanking

Facing reality - Can we only include national representative studies?

Table 1. Alternatives for including only nationally representative studies

Alternatives	Pros	Cons
National	• General population • Can be used for comparisons between countries	• Only few countries • Few consortium partners are responsible for carrying out national studies
Regional	• More study centres can be involved • Aggregation per EU region is still possible	• No comparison possible between countries

No financial consequences: We will strengthen regional studies, this may delay the onset of national representative studies?

We will also include regional studies (with large catching area IF no national studies are available

Facing reality - What number of countries per region?

Table 2. Alternatives for including 3 countries per EU region

Alternatives	Pros	Cons
3 countries per EU region	• 12 of the 26 countries that will contribute with data for a specific age group	• In the East and North we don't have enough countries with available studies in the selected age groups
Number of countries proportional to inhabitants of the region	• Feasibility proportional to capacity in the region	• In the East and North, strong country specific contribution

Number of countries per region was adapted to % inhabitants in the region: North 2, West 3-4, South 3, East 1

Facing reality - What age groups?

Table 3. Age specific groups to consider

Alternatives	Pros	Cons
6-11 y. 12-19 y. 20-40 y.	• Homogeneous adult age group with a more similar health status. • Adult group of reproductive age is considered separately.	• Age group 20-40 years is lacking • No older age group
6-11 y. 12-19 y. 20-60 y. 60+ y.	• Easier to fit in adult datasets (no problem of overlap of data with 2 age categories) • If this group serves as a baseline for occupational exposure groups, a wider age range is needed. • Older age group is present in several studies and may be vulnerable	• More heterogenous age groups will be less suitable for linkage to health data and effect markers. • No specific emphasis on reproductive age group (20-40 years)

Reason not including younger/older age groups:

↑ logistic feasibility of the field work

↓ noise e.g. physiology may differ more within those age groups

Sampling for HBM4EU studies

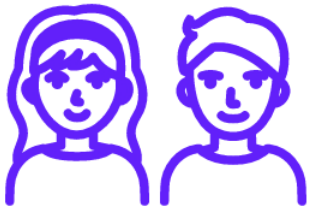
Implementation plan for assessing current EU exposure for 1st group of priority compounds

- **Timing of the sample collection:** 2014-2018
- **Geographical coverage:** Number was adapted to % inhabitants in the region: North 2, West 3, South 3, East 1
- **Expected number of participants/country/age group:** 300 (150 males + 150 females)
- **Representativeness:** We also include regional studies IF no national studies are available

Age groups and 1st priority substances covered by HBM4EU aligned studies

6-11 years

Phthalates+DINCH
Flame retardants



12-19 years

Phthalates+DINCH
Per- poly fluorinated compounds

20-40 years

Bisphenols
Cadmium + PAH

Aligned studies to be conducted in 2019

Age group		North	West	South	East	Total
Children (6-11 y)	Studies	NO-NEBII DK-OCC	FR-ESTEBAN DE-GerES V NL-Duch Youth cohort	SI-SLOCRP EL-CROME IT-NACII	HU-InAirQ SK-PCB cohort PL-POLAES	Aim: 2700-3000
	Participants	600	900	600	850	3950
Adolescents (12-19 y)	Studies	SE-Riksmaten NO-NEBII	FR-ESTEBAN BE-FLEHS IV DE-GerES V	SI-SLOCP EL-CROME ES-BEA	PL-POLAES CZ-School children SK-PCB cohort (follow-up)	Aim: 2700-3000
	Participants	500	900	600	900	2900
Adults (20-40 y)	Studies	DK-CPHMINIPUB IS-Nutrition survey FI-FinHealth	FR-ESTEBAN CH-New study DE-ESB LU-New study	HR-CIPH PT-INSEF	PL-POLAES CZ-ELSPAC	Aim: 2700-3000
	Participants	800	1165	600	600	3165

NATIONAL INSTITUTE FOR HEALTH AND WELFARE

More information at http://www.hbm4eu.eu

vito

is acknowledge about their contribution for preparation of sampling strategy for HBM4EU and providing related material for this presentation.

Speakers' information

Hanna Tolonen, PhD, Adjunct Professor, works as Research manager at the National Institute for Health and Welfare, Finland. She received training in statistics, public health and epidemiology. In HBM4EU she is a member of the Management Board and the Ethics Board, and the leader of WP11 Linking HBM, health surveys and registers.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 733032.