Health 2011 Survey: An overview of the design, missing data and statistical analyses examples

Tommi Härkänen

Department of Health, Functional Capacity and Welfare
The National Institute for Health and Welfare (THL)

June 17, 2013
1. The Health 2000 Survey
2. The Health 2011 Survey
3. Clustering
4. Missing data (nonresponse)
5. Analysis designs
6. Summary
Sampling design

- **Target population:** people aged 18 years or older and living in mainland Finland July 1, 2000.

The design was a stratified two-stage cluster sampling design. 20 strata based on 15 largest towns and the rest of the continental Finland divided by the 5 university hospital regions. 80 health center districts (HCD) including 15 largest towns with probability 1 and systematic PPS sampling of the 65 smaller HCDs so that there were in total 16 HCD’s in each university hospital regions. Systematic sampling of people so that the sample size in each stratum was proportional to the corresponding population base. The sample size in the smaller HCD’s were equal within each stratum, which yielded an approximate equal probability of selection (EPSEM) sample. Oversampling of people aged 80 and older using double inclusion probabilities. Total sample size was 9,922.
Sampling design

- **Target population:** people aged 18 years or older and living in mainland Finland July 1, 2000.

- The design was a stratified two-stage cluster sampling design.

- 20 strata based on 15 largest towns and the rest of the continental Finland divided by the 5 university hospital regions.

- Systematic sampling of people so that the sample size in each stratum was proportional to the corresponding population base.

- The sample size in the smaller health center districts were equal within each stratum, which yielded an approximate equal probability of selection (EPSEM) sample.

- Oversampling of people aged 80 and older using double inclusion probabilities.

- Total sample size was 9,922.
Sampling design

- Target population: people aged 18 years or older and living in mainland Finland July 1, 2000.
- The design was a stratified two-stage cluster sampling design.
- 20 strata based on 15 largest towns and the rest of the continental Finland divided by the 5 university hospital regions.
- 80 health center districts (HCD) including
  - 15 largest towns with probability 1 and
  - systematic PPS sampling of the 65 smaller HCDs so that there were in total 16 HCD’s in each university hospital regions.
Sampling design

- **Target population**: people aged 18 years or older and living in mainland Finland July 1, 2000.

- The design was a stratified two-stage cluster sampling design.

- 20 strata based on 15 largest towns and the rest of the continental Finland divided by the 5 university hospital regions.

- 80 health center districts (HCD) including
  - 15 largest towns with probability 1 and
  - systematic PPS sampling of the 65 smaller HCDs so that there were in total 16 HCD’s in each university hospital regions.

- Systematic sampling of people so that the sample size in each stratum was proportional to the corresponding population base.
Sampling design

- Target population: people aged 18 years or older and living in mainland Finland July 1, 2000.
- The design was a stratified two-stage cluster sampling design.
- 20 strata based on 15 largest towns and the rest of the continental Finland divided by the 5 university hospital regions.
- 80 health center districts (HCD) including
  - 15 largest towns with probability 1 and
  - systematic PPS sampling of the 65 smaller HCDs so that there were in total 16 HCD’s in each university hospital regions.
- Systematic sampling of people so that the sample size in each stratum was proportional to the corresponding population base.
- The sample size in the smaller HCD’s were equal within each stratum, which yielded an approximate equal probability of selection (EPSEM) sample.
Target population: people aged 18 years or older and living in mainland Finland July 1, 2000.

The design was a stratified two-stage cluster sampling design.

20 strata based on 15 largest towns and the rest of the continental Finland divided by the 5 university hospital regions.

80 health center districts (HCD) including
- 15 largest towns with probability 1 and
- systematic PPS sampling of the 65 smaller HCDs so that there were in total 16 HCD’s in each university hospital regions.

Systematic sampling of people so that the sample size in each stratum was proportional to the corresponding population base.

The sample size in the smaller HCD’s were equal within each stratum, which yielded an approximate equal probability of selection (EPSEM) sample.

Oversampling of people aged 80 and older using double inclusion probabilities.

Total sample size was 9,922.
Mini-Finland Resurvey data

- **Mini-Finland Survey** was conducted in 1978-80
  - Sample was based on a complex sampling design similar to the Health 2000 sample, and it represented the population
  - Sample size 8,000 and participation rate 90%
Mini-Finland Resurvey data

- Mini-Finland Survey was conducted in 1978-80
  - Sample was based on a complex sampling design similar to the Health 2000 sample, and it represented the population
  - Sample size 8,000 and participation rate 90%
- In Health 2000 Survey, 1,278 participants of Mini-Finland Survey, who lived in 9 towns in 2000, were invited.
Mini-Finland Resurvey data

- Mini-Finland Survey was conducted in 1978-80
  - Sample was based on a complex sampling design similar to the Health 2000 sample, and it represented the population
  - Sample size 8,000 and participation rate 90%
- In Health 2000 Survey, 1,278 participants of Mini-Finland Survey, who lived in 9 towns in 2000, were invited
- Does not represent population due to left truncation
  - We do not know who were the members of the baseline sample, who would have
    - moved to (or stayed in) one of the 9 towns between 1978-80 and 2000, and
    - not died
  - This would require further investigations but could be done
Missing data in Health 2000 Survey

- Participation rates in age group 30 years and above:
  - Home interview 89%
  - Health examination 85%
  - Any part of the survey 93%

Participation rates in age group 18 to 29 years:
- Any interview or questionnaire 90%

Participation rates in the Mini-Finland Resurvey:
- Health examination 80%
- Any part of the survey 89%
Missing data in Health 2000 Survey

- Participation rates in age group 30 years and above:
  - Home interview 89 %
  - Health examination 85 %
  - Any part of the survey 93 %

- Participation rate in age group 18 to 29 years:
  - Any interview or questionnaire 90 %
Missing data in Health 2000 Survey

- Participation rates in age group 30 years and above:
  - Home interview 89%
  - Health examination 85%
  - Any part of the survey 93%

- Participation rate in age group 18 to 29 years:
  - Any interview or questionnaire 90%

- Participation rates in the Mini-Finland Resurvey:
  - Health examination 80%
  - Any part of the survey 89%
Handling of missing data in the Health 2000 Survey

- **Post-stratification weights** were calibrated by Statistics Finland
  - Calibration was based on register information on age, gender, area and language
Handling of missing data in the Health 2000 Survey

- **Post-stratification weights** were calibrated by Statistics Finland
  - Calibration was based on register information on age, gender, area and language
  - Participation was defined for
    - All respondents any section of the survey or separate non-response interviews or inquiries $n = 9,125$
    - Respondents’ union any section of the survey $n = 8,617$
      - Nutrition several sections and especially in the nutrition inquiry $n = 6,794$
      - Intersection interviews or corresponding inquiries and in most clinical sections and inquiries $n = 6,774$
Handling of missing data in the Health 2000 Survey

- **Post-stratification weights** were calibrated by Statistics Finland
  - Calibration was based on register information on age, gender, area and language
  - Participation was defined for
    - All respondents any section of the survey or separate non-response interviews or inquiries $n = 9,125$
    - Respondents’ union any section of the survey $n = 8,617$
    - Nutrition several sections and especially in the nutrition inquiry $n = 6,794$
    - Intersection interviews or corresponding inquiries and in most clinical sections and inquiries $n = 6,774$

- **Multiple imputation** has also been applied in several analyses
Options for analyses

- Most analyses were **design-based analyses** based on
  - generalized estimating equations and
  - post-stratification weights
- Some **model-based** analyses using hierarchical models have also been conducted
Sampling design of The Health 2011 Survey

**Health 2000 Survey** data (aged 29 or older)
- The Health 2000 sampling design (strata and PSUs)
- Repeated measurements on the members of the Health 2000 sample
- 7,964 were invited in the age group 30 years or older
- 59% participated in the health examination
- 73% participated in any part of the survey

**Mini-Finland Resurvey** data (aged 61 years or older)
- 922 were invited and 81% participated

**New sample** of 1,994 young adults (aged 18 to 28)
- Same geographical areas as in 2000
- Amalgamations of municipalities/HCDs were handled using GIS coordinates
- 415 were invited in the health examination and 52% participated
- 1,579 were sent a questionnaire and 40% participated

---

1 Population Register Centre (VRK)
Sampling design of The Health 2011 Survey

- **Health 2000 Survey** data (aged 29 or older)
  - The Health 2000 sampling design (strata and PSUs)
  - Repeated measurements on the members of the Health 2000 sample
  - 7,964 were invited in the age group 30 years or older
  - 59% participated in the health examination
  - 73% participated in any part of the survey

- **Mini-Finland Resurvey** data (aged 61 years or older)
  - 922 were invited and 81% participated

---

1. Population Register Centre (VRK)
Sampling design of The Health 2011 Survey

- **Health 2000 Survey** data (aged 29 or older)
  - The Health 2000 sampling design (strata and PSUs)
  - Repeated measurements on the members of the Health 2000 sample
  - 7,964 were invited in the age group 30 years or older
  - 59% participated in the health examination
  - 73% participated in any part of the survey

- **Mini-Finland Resurvey** data (aged 61 years or older)
  - 922 were invited and 81% participated

- **New sample** of 1,994 young adults (aged 18 to 28)
  - Same geographical areas as in 2000
  - Amalgamations of municipalities/HCDs were handled using GIS coordinates\(^1\) in some cases
  - 415 were invited in the health examination and 52% participated
  - 1,579 were sent a questionnaire and 40% participated

---

\(^1\)Population Register Centre (VRK)
Representativeness of the original Health 2000 sample

- The original Health 2000 sample represented the population
Representativeness of the original Health 2000 sample

- The original Health 2000 sample represented the population
- The same selection mechanisms, which apply to the population in 2000, also apply to the sample:
  - Mortality, emigration and migration in Finland, but
  - not immigration after 2000 (there is a separate survey\(^2\) on the immigrants)

\(^2\) www.thl.fi/maamu
Representativeness of the original Health 2000 sample

- The original Health 2000 sample represented the population
- The same selection mechanisms, which apply to the population in 2000, also apply to the sample:
  - Mortality, emigration and migration in Finland, but
  - not immigration after 2000 (there is a separate survey\(^2\) on the immigrants)
- We consider that the original sample still provides representative results in cross-sectional analyses in 2011

\(^2\) [www.thl.fi/maamu](http://www.thl.fi/maamu)
Additional protocols were applied on randomly selected subgroups.

Table: Sample sizes in different age and substudy groups.

<table>
<thead>
<tr>
<th>Age group</th>
<th>Sample</th>
<th>Courage</th>
<th>Exercise</th>
<th>Courage and Exercise</th>
</tr>
</thead>
<tbody>
<tr>
<td>18-28</td>
<td>1994</td>
<td>415</td>
<td>415</td>
<td>415</td>
</tr>
<tr>
<td>29-49</td>
<td>3306</td>
<td>884</td>
<td>2088</td>
<td>0</td>
</tr>
<tr>
<td>50-74</td>
<td>3840</td>
<td>1885</td>
<td>1908</td>
<td>0</td>
</tr>
<tr>
<td>75-79</td>
<td>384</td>
<td>181</td>
<td>193</td>
<td>5</td>
</tr>
<tr>
<td>80+</td>
<td>605</td>
<td>568</td>
<td>294</td>
<td>294</td>
</tr>
<tr>
<td>All</td>
<td>10129</td>
<td>3933</td>
<td>4898</td>
<td>714</td>
</tr>
</tbody>
</table>
Clustering effects can emerge from various sources, e.g.

- Original Health 2000 sampling design:
  - Stratification: largest towns (15) and university hospital district (5)
  - Primary sampling units (PSUs): health center districts (HCDs)
- Municipality vs. HCD
- Distances (spatial analyses using GIS coordinates)
Clustering effects can emerge from various sources, e.g.

- Original Health 2000 sampling design:
  - Stratification: largest towns (15) and university hospital district (5)
  - Primary sampling units (PSUs): health center districts (HCDs)
- Municipality vs. HCD
- Distances (spatial analyses using GIS coordinates)
- Individual level in longitudinal studies (study subject in the repeated measurements)
- Current HCD in 2000 and 2011
Clustering effects can emerge from various sources, e.g.

- Original Health 2000 sampling design:
  - Stratification: largest towns (15) and university hospital district (5)
  - Primary sampling units (PSUs): health center districts (HCDs)
- Municipality vs. HCD
- Distances (spatial analyses using GIS coordinates)
- Individual level in longitudinal studies (study subject in the repeated measurements)
- Current HCD in 2000 and 2011

The individual-level clustering is likely to be most important, but other clustering levels could be significant as well: this can depend on the variable of interest.
Clustering

Clustering effects can emerge from various sources, e.g.

- Original Health 2000 sampling design:
  - Stratification: largest towns (15) and university hospital district (5)
  - Primary sampling units (PSUs): health center districts (HCDs)
- Municipality vs. HCD
- Distances (spatial analyses using GIS coordinates)
- Individual level in longitudinal studies (study subject in the repeated measurements)
- Current HCD in 2000 and 2011

The individual-level clustering is likely to be most important, but other clustering levels could be significant as well: this can depend on the variable of interest.

For example, blood pressure appeared to have strong clustering effects in Health 2000, but BMI did not.
Clustering in statistical analyses

**Design-based methods**  Traditional methods to handle clustering, usually methods based on linearization and estimating equations (GEE)

**Model-based methods**  Clustering has been handled by e.g. mixed effects (hierarchical) models
Clustering in statistical analyses

Design-based methods  Traditional methods to handle clustering, usually methods based on linearization and estimating equations (GEE)

Model-based methods  Clustering has been handled by e.g. mixed effects (hierarchical) models

Some differences exist:
- Model-based methods can be useful when having several levels of hierarchy ⇒ variance on different levels can be compared.
- Model-based methods can be more efficient but also more sensitive to model assumptions.
Clustering in statistical analyses

**Design-based methods** Traditional methods to handle clustering, usually methods based on linearization and estimating equations (GEE)

**Model-based methods** Clustering has been handled by e.g. mixed effects (hierarchical) models

Some differences exist:

- Model-based methods can be useful when having several levels of hierarchy ⇒ variance on different levels can be compared.
- Model-based methods can be more efficient but also more sensitive to model assumptions.
- Design-based methods can effectively handle only one level of clustering. In Health 2011 repeated measurements analyses it is advisable to choose the individual level as the primary sampling unit (PSU).
Linear mixed effects models produce similar point estimates as GEE.

\[^3\text{E.g. } \text{http://courses.washington.edu/b571/lectures/set5.pdf (Johnson and Kotz, 1970).}\]
Similarities and differences of hierarchical models and GEE

- Linear mixed effects models produce similar point estimates as GEE.

- In non-linear models the point estimates generally differ.
  - Generalized linear mixed effects (GLME) models (e.g. logistic regression analyses with random effects) estimate **conditional effects**.
  - GEE estimates **marginal effects**, which are attenuated towards zero (towards OR=1).

---

Linear mixed effects models produce similar point estimates as GEE.

In non-linear models the point estimates generally differ!

- Generalized linear mixed effects (GLME) models (e.g. logistic regression analyses with random effects) estimate **conditional effects**.
- GEE estimates **marginal effects**, which are attenuated towards zero (towards OR=1).

What is a conditional/marginal effect (in case of a logistic regression model)???

---

Similarities and differences of hierarchical models and GEE

- Linear mixed effects models produce similar point estimates as GEE.
- In non-linear models the point estimates generally differ.
  - Generalized linear mixed effects (GLME) models (e.g. logistic regression analyses with random effects) estimate conditional effects.
  - GEE estimates marginal effects, which are attenuated towards zero (towards OR=1).
- What is a conditional/marginal effect (in case of a logistic regression model)???
  
  Marginal effect of a covariate is the average effect of a covariate in the population.

---

Similarities and differences of hierarchical models and GEE

- Linear mixed effects models produce similar point estimates as GEE.
- In non-linear models the point estimates generally differ.³
  - Generalized linear mixed effects (GLME) models (e.g. logistic regression analyses with random effects) estimate **conditional effects**.
  - GEE estimates **marginal effects**, which are attenuated towards zero (towards OR=1).

What is a conditional/marginal effect (in case of a logistic regression model)???

**Marginal effect** of a covariate is the **average** effect of a covariate in the population.

**Conditional effect** is the effect of a covariate for a particular **individual** (or a cluster of observations cf. the hierarchical model).

Conditional OR = 2.7
Marginal OR = 1.9

- Note that e.g. near probability 1 the individual curves of those individuals, who are in higher risk, bend. ("ceiling effect")
- On the other hand the curves of the individuals with low risk are much lower.

⇒ The average of the curves (marginal effect) also bends.
⇒ Covariate X has a smaller marginal effect on the outcome

Example by Jon Wakefield (2009)
Participation rates

Table: Participation rates in different age and substudy groups. HES refers to health examination.

<table>
<thead>
<tr>
<th></th>
<th>18-28</th>
<th>29-49</th>
<th>50-74</th>
<th>75-79</th>
<th>80+</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total sample</td>
<td>1994</td>
<td>3306</td>
<td>3840</td>
<td>384</td>
<td>605</td>
<td>10129</td>
</tr>
<tr>
<td>Participation (%)</td>
<td>42</td>
<td>68</td>
<td>79</td>
<td>74</td>
<td>59</td>
<td>67</td>
</tr>
<tr>
<td>HES sample</td>
<td>415</td>
<td>3306</td>
<td>3840</td>
<td>384</td>
<td>605</td>
<td>8550</td>
</tr>
<tr>
<td>HES participation (%)</td>
<td>29</td>
<td>50</td>
<td>66</td>
<td>61</td>
<td>50</td>
<td>57</td>
</tr>
<tr>
<td>Courage sample</td>
<td>415</td>
<td>884</td>
<td>1885</td>
<td>181</td>
<td>568</td>
<td>3933</td>
</tr>
<tr>
<td>Courage participation (%)</td>
<td>52</td>
<td>67</td>
<td>78</td>
<td>76</td>
<td>58</td>
<td>70</td>
</tr>
<tr>
<td>Physical exercise sample</td>
<td>415</td>
<td>2088</td>
<td>1908</td>
<td>193</td>
<td>294</td>
<td>4898</td>
</tr>
<tr>
<td>Physical exercise participation (%)</td>
<td>52</td>
<td>68</td>
<td>80</td>
<td>73</td>
<td>59</td>
<td>71</td>
</tr>
</tbody>
</table>
Nonresponse in 2000 vs. 2011


<table>
<thead>
<tr>
<th>Participated in 2000 (%)</th>
<th>Participated in 2011 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td>Yes</td>
<td>69</td>
</tr>
<tr>
<td>No</td>
<td>4</td>
</tr>
<tr>
<td>Sum</td>
<td>73</td>
</tr>
</tbody>
</table>
Factors which are often associated with nonresponse

- Low social activity, low education
Factors which are often associated with nonresponse

- Low social activity, low education
- **Oldest age groups:** Illnesses, disabilities, weak functional capacity
Factors which are often associated with nonresponse

- Low social activity, low education
- **Oldest age groups:** Illnesses, disabilities, weak functional capacity
- **Young age groups:** Male

Note that generally only part of these factors can be/have been observed!

Missing completely at random (MCAR) Nonresponse can be ignored in all analyses, only statistical power is lower. Usually unrealistic.

Missing at random (MAR) Probability of nonresponse depends only on observed data. Effects of nonresponse can be corrected.

Missing not at random (MNAR) Probability of nonresponse depends also on unobserved data. Untestable assumptions are needed.
Factors which are often associated with nonresponse

- Low social activity, low education
- **Oldest age groups:** Illnesses, disabilities, weak functional capacity
- **Young age groups:** Male

Note that generally only part of these factors can be/have been observed!
Factors which are often associated with nonresponse

- Low social activity, low education
- **Oldest age groups:** Illnesses, disabilities, weak functional capacity
- **Young age groups:** Male

Note that generally only part of these factors can be/have been observed!

**Missing completely at random (MCAR)** Nonresponse can be ignored in all analyses, only statistical power is lower. Usually unrealistic.
Factors which are often associated with nonresponse

- Low social activity, low education
- **Oldest age groups:** Illnesses, disabilities, weak functional capacity
- **Young age groups:** Male

Note that generally only part of these factors can be/have been observed!

**Missing completely at random (MCAR)** Nonresponse can be ignored in all analyses, only statistical power is lower. Usually unrealistic.

**Missing at random (MAR)** Probability of nonresponse depends only on observed data. Effects of nonresponse can be corrected.
Factors which are often associated with nonresponse

- Low social activity, low education
- **Oldest age groups:** Illnesses, disabilities, weak functional capacity
- **Young age groups:** Male

Note that generally only part of these factors can be/have been observed!

**Missing completely at random (MCAR)** Nonresponse can be ignored in all analyses, only statistical power is lower. Usually unrealistic.

**Missing at random (MAR)** Probability of nonresponse depends only on observed data. Effects of nonresponse can be corrected.

**Missing not at random (MNAR)** Probability of nonresponse depends also on unobserved data. **Untestable assumptions** are needed.
How to handle effects of missing data?

**Weighting**  Probability of participation varies individually. Each participant is given a weight indicating the number of similar sample members, which the participant represents:

- Ideally, participation rate 100% ⇒ weight equals 1.

Low participation rates ≪ 100% ⇒ weight can be much larger than 1 ⇒ unstable results.

Weighting works best in case of unit nonresponse.

Imputation: Missing values are replaced by predictive values. Better than weighting especially in the case of item nonresponse.

Other methods: E.g. data augmentation using Bayesian inference can allow flexible incorporation of prior/expert information on missing data mechanisms.
How to handle effects of missing data?

**Weighting** Probability of participation varies individually. Each participant is given a weight indicating the number of similar sample members, which the participant represents:

- Ideally, participation rate 100% ⇒ weight equals 1.
- Low participation rates ≪ 100% ⇒ weight can be much larger than 1 ⇒ unstable results.
How to handle effects of missing data?

**Weighting** Probability of participation varies individually. Each participant is given a weight indicating the number of similar sample members, which the participant represents:

- Ideally, participation rate 100% ⇒ weight equals 1.
- Low participation rates ≪ 100% ⇒ weight can be much larger than 1 ⇒ unstable results.

Weighting works best in case of unit nonresponse.
How to handle effects of missing data?

Weighting  Probability of participation varies individually. Each participant is given a weight indicating the number of similar sample members, which the participant represents:
  - Ideally, participation rate 100% ⇒ weight equals 1.
  - Low participation rates ≪ 100% ⇒ weight can be much larger than 1 ⇒ unstable results.

Weighting works best in case of unit nonresponse

Imputation  Missing values are replaced by predictive values
  - Better than weighting especially in the case of item nonresponse
How to handle effects of missing data?

**Weighting**  
Probability of participation varies individually. Each participant is given a weight indicating the number of similar sample members, which the participant represents:

- Ideally, participation rate 100% $\Rightarrow$ weight equals 1.
- Low participation rates $\ll$ 100%  
  $\Rightarrow$ weight can be much larger than 1  
  $\Rightarrow$ **unstable results**.

Weighting works best in case of **unit nonresponse**

**Imputation**  
Missing values are replaced by predictive values

- Better than weighting especially in the case of **item nonresponse**

**Other methods**  
E.g. data augmentation using Bayesian inference can allow flexible incorporation of prior/expert information on missing data mechanisms.
Weights

Separate weights for each wave:

- **The Health 2000 Survey** Use the original Health 2000 poststratification weights.

...
Weights

Separate weights for each wave:

- **The Health 2000 Survey** Use the original Health 2000 poststratification weights.
- **The Health 2011 Survey**
Weights

Separate weights for each wave:

- **The Health 2000 Survey** Use the original Health 2000 poststratification weights.

- **The Health 2011 Survey**
  - Participants in 2000 Inverse probability weights (IPW) method based on logistic regression model. Predictors based on observed information in 2000, e.g.
    - risk factors of diseases and disabilities
    - various lifestyle factors (social activity etc.)
Weights

Separate weights for each wave:

- **The Health 2000 Survey** Use the original Health 2000 poststratification weights.

- **The Health 2011 Survey**
  
  **Participants in 2000** Inverse probability weights (IPW) method based on logistic regression model. Predictors based on observed information in 2000, e.g.
  
  - risk factors of diseases and disabilities
  - various lifestyle factors (social activity etc.)

  **Nonparticipants in 2000** A small number of individuals without observed baseline information
  
  ⇒ Poststratification weights based on age groups.
Weights

Separate weights for each wave:

- **The Health 2000 Survey** Use the original Health 2000 poststratification weights.

- **The Health 2011 Survey**
  
  **Participants in 2000** Inverse probability weights (IPW) method based on logistic regression model. Predictors based on observed information in 2000, e.g.
  
  - risk factors of diseases and disabilities
  - various lifestyle factors (social activity etc.)

  **Nonparticipants in 2000** A small number of individuals without observed baseline information
  
  ⇒ Poststratification weights based on age groups.

  **New sample of young adults** IPW weights based on age, gender and area (register data).
Weights

Separate weights for each wave:

- **The Health 2000 Survey** Use the original Health 2000 poststratification weights.

- **The Health 2011 Survey**
  
  Participants in 2000 Inverse probability weights (IPW) method based on logistic regression model. Predictors based on observed information in 2000, e.g.
  
  - risk factors of diseases and disabilities
  - various lifestyle factors (social activity etc.)

  Nonparticipants in 2000 A small number of individuals without observed baseline information
  
  ⇒ Poststratification weights based on age groups.

  New sample of young adults IPW weights based on age, gender and area (register data).

Resulting weights were further calibrated so that they match the population sizes in different areas.
Different definitions of weights

- Different definitions of participation (DoP) in 2011, e.g.
  1. any part of the survey
  2. the health examination (HES)
  3. the Courage subsample
  4. the physical exercise subsample (4 different criteria)
  5. (even more, e.g. food frequency questionnaire?)
Different definitions of weights

- Different definitions of participation (DoP) in 2011, e.g.
  1. any part of the survey
  2. the health examination (HES)
  3. the Courage subsample
  4. the physical exercise subsample (4 different criteria)
  5. (even more, e.g. food frequency questionnaire?)

- How to choose the correct weight for my analysis?
Different definitions of weights

- Different definitions of participation (DoP) in 2011, e.g.
  1. any part of the survey
  2. the health examination (HES)
  3. the Courage subsample
  4. the physical exercise subsample (4 different criteria)
  5. (even more, e.g. food frequency questionnaire?)

- **How to choose** the correct weight for my analysis?
  - The DoP should match one-to-one to the participation to the analysis variables.
    E.g. in analyses involving HES variables, the HES weights are the best choice.
Different definitions of weights

- Different definitions of participation (DoP) in 2011, e.g.
  1. any part of the survey
  2. the health examination (HES)
  3. the Courage subsample
  4. the physical exercise subsample (4 different criteria)
  5. (even more, e.g. food frequency questionnaire?)

- **How to choose** the correct weight for my analysis?
  - The DoP should match one-to-one to the participation to the analysis variables.
    E.g. in analyses involving HES variables, the HES weights are the best choice.
  - Can be problematic, if the variables were collected at different sections of the survey (item nonresponse).
Weights

Pros:
- weights are **easy to use** in analyses
- weights can work well in case of **unit-nonresponse** (no additional item-nonresponse)
Weights

Pros:
- weights are easy to use in analyses
- weights can work well in case of unit-nonresponse (no additional item-nonresponse)

Cons:
- There are generally considerable amounts of item-nonresponse thus participation indicator of the weight variable often differ from observed analysis variable ⇒ weights do not work optimally
- item-nonresponse in any analysis variable drops the individual out of the analysis ⇒ loss of information
Table: Comparison of prevalences (%) in age group 30 years and older.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Clustering</th>
<th>Missing data</th>
<th>Prevalence</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disability pension</td>
<td>SRS</td>
<td>None</td>
<td>8.8</td>
<td>0.4</td>
</tr>
<tr>
<td></td>
<td>Complex</td>
<td>Baseline weights</td>
<td>8.9</td>
<td>0.4</td>
</tr>
<tr>
<td></td>
<td>Complex</td>
<td>Resurvey weights</td>
<td>9.4</td>
<td>0.4</td>
</tr>
<tr>
<td></td>
<td>Complex</td>
<td>True prevalence</td>
<td>9.5</td>
<td>0.4</td>
</tr>
<tr>
<td>Hospitalization</td>
<td>SRS</td>
<td>None</td>
<td>16.6</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>Complex</td>
<td>Baseline weights</td>
<td>16.7</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>Complex</td>
<td>Resurvey weights</td>
<td>17.3</td>
<td>0.6</td>
</tr>
<tr>
<td></td>
<td>Complex</td>
<td>True prevalence</td>
<td>17.8</td>
<td>0.5</td>
</tr>
<tr>
<td>Reimbursement</td>
<td>SRS</td>
<td>None</td>
<td>40.2</td>
<td>0.7</td>
</tr>
<tr>
<td></td>
<td>Complex</td>
<td>Baseline weights</td>
<td>40.6</td>
<td>0.8</td>
</tr>
<tr>
<td></td>
<td>Complex</td>
<td>Resurvey weights</td>
<td>42.0</td>
<td>0.8</td>
</tr>
<tr>
<td></td>
<td>Complex</td>
<td>True prevalence</td>
<td>42.1</td>
<td>0.6</td>
</tr>
</tbody>
</table>
Multiple imputation (MI)

MI in short:

1. Create **several copies** of the original data set
2. Impute missing values using **predictive distributions** based on the associations and variable values of the observed data
   - Various methods exist, some are available in general-purpose statistical software
3. Analyze **separately** the copies using standard statistical methods
4. Join the results

---


Tommi Härkänen (THL)  The Health 2011 Survey  June 17, 2013  23 / 33
Multiple imputation (MI)

- MI in short:
  1. Create **several copies** of the original data set
  2. Impute missing values using **predictive distributions** based on the associations and variable values of the observed data
     - Various methods exist, some are available in general-purpose statistical software
  3. **Analyze separately** the copies using standard statistical methods
  4. **Join** the results

- For example, MI has been used in **NHANES** 4
  - MI based on multivariate **linear** mixed effects models
  - 67 health examination and lifestyle variables were imputed
  - MI separately in 9 age groups
  - clustering in the data was accounted for
  - R package **pan** by Schafer

---

**Multiple imputation (MI)**

- **Pros:**
  - Item-nonresponse can be better accounted for by replacing missing values with randomly drawn values from a predictive distribution
  - MI is generally more efficient than weighting
  - Imputation model does not have to be the same as the analysis model
Multiple imputation (MI)

Pros:
- Item-nonresponse can be better accounted for by replacing missing values with randomly drawn values from a predictive distribution
- MI is generally more efficient than weighting
- Imputation model does not have to be the same as the analysis model

Cons:
- Difficult to produce one MI data set for the whole Health 2011 survey
- More laborious as the imputation model should be tuned separately for most analyses:
  - variable selection, level of measurement, interactions, nonlinearities, heteroscedasticity etc.
Multiple imputation (MI)

Pros:
- Item-nonresponse can be better accounted for by replacing missing values with randomly drawn values from a predictive distribution
- MI is generally more efficient than weighting
- Imputation model does not have to be the same as the analysis model

Cons:
- Difficult to produce one MI data set for the whole Health 2011 survey
- More laborious as the imputation model should be tuned separately for most analyses:
  - variable selection, level of measurement, interactions, nonlinearities, heteroscedasticity etc.

The imputation models and methods must be documented in great detail in the publications!
This allows other researchers to apply the same imputation model (possibly with some additional variables etc.)
Example: Weighting vs. MI

Table: Kendall correlations of body mass index (BMI), systolic blood pressure (SBP) and walking speed (WSp).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>BMI 2000</td>
<td>1.00</td>
<td>0.65</td>
<td>0.24</td>
<td>0.21</td>
<td>-0.12</td>
<td>-0.18</td>
<td>7,585</td>
</tr>
<tr>
<td>BMI 2011</td>
<td>0.65</td>
<td>1.00</td>
<td>0.17</td>
<td>0.18</td>
<td>-0.11</td>
<td>-0.15</td>
<td>4,253</td>
</tr>
<tr>
<td>SBP 2000</td>
<td>0.24</td>
<td>0.17</td>
<td>1.00</td>
<td>0.40</td>
<td>-0.10</td>
<td>-0.20</td>
<td>5,561</td>
</tr>
<tr>
<td>SBP 2011</td>
<td>0.21</td>
<td>0.18</td>
<td>0.40</td>
<td>1.00</td>
<td>-0.02</td>
<td>-0.13</td>
<td>4,239</td>
</tr>
<tr>
<td>WSp 2000</td>
<td>-0.12</td>
<td>-0.11</td>
<td>-0.10</td>
<td>-0.02</td>
<td>1.00</td>
<td>0.43</td>
<td>1,833</td>
</tr>
<tr>
<td>WSp 2011</td>
<td>-0.18</td>
<td>-0.15</td>
<td>-0.20</td>
<td>-0.13</td>
<td>0.43</td>
<td>1.00</td>
<td>4,191</td>
</tr>
</tbody>
</table>
Example: Estimating BMI means using weighting vs. MI

Table: Comparison of different methods to handle missing data in the estimation of the body mass index (BMI) mean.

<table>
<thead>
<tr>
<th>Clustering</th>
<th>Missing data</th>
<th>Mean</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRS</td>
<td>None</td>
<td>27.06</td>
<td>0.08</td>
</tr>
<tr>
<td>Complex</td>
<td>Baseline weights</td>
<td>27.16</td>
<td>0.08</td>
</tr>
<tr>
<td>Complex</td>
<td>Resurvey weights(^5)</td>
<td>27.03</td>
<td>0.08</td>
</tr>
<tr>
<td>Complex</td>
<td>Resurvey HES weights(^6)</td>
<td>26.47</td>
<td>0.13</td>
</tr>
<tr>
<td>Complex</td>
<td>Baseline weights; Multiple imputation</td>
<td>26.84</td>
<td>0.08</td>
</tr>
</tbody>
</table>

\(^5\) Participation in any part of survey
\(^6\) Participation in health examination
Example: Estimating systolic blood pressure means using weighting vs. MI

<table>
<thead>
<tr>
<th>Clustering</th>
<th>Missing data</th>
<th>Mean</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRS</td>
<td>None</td>
<td>131.21</td>
<td>0.28</td>
</tr>
<tr>
<td>Complex</td>
<td>Baseline weights</td>
<td>131.77</td>
<td>0.32</td>
</tr>
<tr>
<td>Complex</td>
<td>Resurvey weights(^7)</td>
<td>131.04</td>
<td>0.32</td>
</tr>
<tr>
<td>Complex</td>
<td>Resurvey HES weights(^8)</td>
<td>128.36</td>
<td>0.43</td>
</tr>
<tr>
<td>Complex</td>
<td>Baseline weights; Multiple imputation</td>
<td>129.04</td>
<td>0.33</td>
</tr>
</tbody>
</table>

\(^7\) Participation in any part of survey

\(^8\) Participation in health examination
Assumptions in MI

Associations of continuous variables are not always linear and errors terms not homoscedastic and symmetric.
Cross-sectional analyses

- "What was the population distribution and/or the associations between variables?"

What needs to be accounted for in the analyses?

- Analyses for Mini-Finland Survey or the Health 2000 Survey:
  - Clustering in the data (strata and PSUs, especially the 80 health center districts)
  - Missing data handled using the weights

- Analyses for Health 2011:
  - More options in handling clustering effects (health center districts in 2000 vs. 2011) due to 
    environmental effects (which can influence the outcome after migration) or 
    selection effects due to migration
  - More missing data
Longitudinal analyses: Changes in population distributions from (1980 to) 2000 to 2011

- “What was the difference of means between 2000 and 2011 in age group X?”
- Repeated measurements design – standard methodology
  - Mini-Finland and Health 2000 were independent samples
  - At most one measurement per individual in Mini-Finland data
Longitudinal analyses: Changes in population distributions from (1980 to) 2000 to 2011

- ‘What was the difference of means between 2000 and 2011 in age group X?’
- Repeated measurements design – standard methodology
  - Mini-Finland and Health 2000 were independent samples
  - At most one measurement per individual in Mini-Finland data
  - In Health 2000 and Health 2011 the same individuals were examined
  - At most two measurements per individual in Health 2000 data

Example: A simple regression model could be defined as

\[ \text{MODEL } Y = \text{AGEGROUP} + \text{TIME} \]

Outcome: \( Y \)

Categorical covariates:

- \( \text{AGEGROUP} \) is age of study subject at measurement time, and measurement is \( \text{TIME} \) (note that subjects are 11 years older at the end of follow-up!)

The term \( \text{TIME} \) represents the overall change.
Longitudinal analyses: Changes in population distributions from (1980 to) 2000 to 2011

- ‘What was the difference of means between 2000 and 2011 in age group X?’

Repeated measurements design – standard methodology
- Mini-Finland and Health 2000 were independent samples
  - At most one measurement per individual in Mini-Finland data
- In Health 2000 and Health 2011 the same individuals were examined
  - At most two measurements per individual in Health 2000 data

Example: A simple regression model could be defined as
MODEL Y = AGEGROUP + TIME
- Outcome: Y
- Categorical covariates: AGEGROUP is age of study subject at measurement time, and measurement is TIME (note that subjects are 11 years older at the end of follow-up!)
- The term TIME represent the overall change.
Longitudinal analyses: Individual changes

- “What has been the average change of variable X during the follow-up?”
- Repeated measurements design – standard methodology, e.g. hierarchical models
Longitudinal analyses: Individual changes

- "What has been the average change of variable X during the follow-up?"
- Repeated measurements design – standard methodology, e.g. hierarchical models
- Example: A simple regression model could be defined as
  
  \[
  \text{MODEL \hspace{1em} Y = AGEGROUP + TIME}
  \]
  
  - Outcome: \( Y \)
  - Categorical covariates: AGEGROUP is age of study subject at baseline, and measurement is TIME (note that subjects have aged 11 during follow-up!)
  - The term TIME represent the average individual change between 2000 and 2011
Longitudinal analyses: Individual changes

- ‘What has been the average change of variable X during the follow-up?’
- Repeated measurements design – standard methodology, e.g. hierarchical models
- Example: A simple regression model could be defined as 
  \[ \text{MODEL } Y = \text{AGEGROUP} + \text{TIME} \]
  - Outcome: \( Y \)
  - Categorical covariates: \( \text{AGEGROUP} \) is age of study subject at \( \text{baseline} \), and measurement is \( \text{TIME} \) (note that subjects have aged 11 during follow-up!)
  - The term \( \text{TIME} \) represents the average individual change between 2000 and 2011
- But informative right-censoring can complicate the interpretation of the results
  - Mortality at older age can influence the results if the outcome is associated with risk of death
  - Individuals with weak condition are more likely to die, but those of them, who do not die, can have more positive progression of the outcome
  \( \Rightarrow \) Trend estimates can be too positive
Example on informative right-censoring

- Assume repeated measurements study with 2 measurements
- We want to estimate the average speed of change

Subjects in good condition are likely to survive until second wave ⇒ No problems in estimating the speed of change

Subjects in poor condition might well die before second wave ⇒ There is a selection process ⇒ Problems in estimating the speed of change speed of change is underestimated!

Possible solution is to use a selection model

Stenholm, Härkänen et al. (2012)
Example on informative right-censoring

- Assume repeated measurements study with 2 measurements
- We want to estimate the average speed of change
- Subjects in good condition are likely to survive until second wave
  \[ \Rightarrow \text{No problems in estimating the speed of change} \]

Subjects in poor condition might well die before second wave
\[ \Rightarrow \text{Problems in estimating the speed of change} \]

Possible solution is to use a selection model

\text{Stenholm, Härkänen et al. (2012)}
Example on informative right-censoring

- Assume repeated measurements study with 2 measurements
- We want to estimate the average speed of change
- Subjects in **good condition** are likely to survive until second wave
  $\Rightarrow$ No problems in estimating the speed of change
- Subjects in **poor condition** might well die before second wave

---

Stenholm, Härkänen et al. (2012)
Example on informative right-censoring

Assume repeated measurements study with 2 measurements.

We want to estimate the average speed of change.

Subjects in **good condition** are likely to survive until second wave
⇒ No problems in estimating the speed of change.

Subjects in **poor condition** might well die before second wave.

There is a **selection process**
⇒ Problems in estimating the speed of change – speed of change is underestimated!

---

Stenholm, Härkänen et al. (2012)
Example on informative right-censoring

- Assume repeated measurements study with 2 measurements
- We want to estimate the average speed of change
- Subjects in good condition are likely to survive until second wave
  ⇒ No problems in estimating the speed of change
- Subjects in poor condition might well die before second wave
- There is a selection process
  ⇒ Problems in estimating the speed of change – speed of change is underestimated!
- Possible solution is to use a selection model\(^a\)

\(^a\)Stenholm, Härkänen et al. (2012)
Summary

- Nonresponse has increased considerably from 2000 to 2011, and its effects should be corrected for
  - The Health 2011 organization will provide IPW weights to handle missing data
  - Researchers are encouraged to use better methods such as multiple imputation (MI) or data augmentation
Nonresponse has increased considerably from 2000 to 2011, and its effects should be corrected for:

- The Health 2011 organization will provide IPW weights to handle missing data.
- Researchers are encouraged to use better methods such as multiple imputation (MI) or data augmentation.

Clustering needs to be accounted for, and in longitudinal studies the individual variation is likely to be most important.
Summary

- Nonresponse has increased considerably from 2000 to 2011, and its effects should be corrected for.
  - The Health 2011 organization will provide IPW weights to handle missing data.
  - Researchers are encouraged to use better methods such as multiple imputation (MI) or data augmentation.

- Clustering needs to be accounted for, and in longitudinal studies the individual variation is likely to be most important.

- Register-based follow-up
  - Important in handling missing data (of the Health 2011 Survey).
  - Important also in cohort analyses as outcome variables.
  - Methodology for joint analysis of time-to-event data and repeated measurements is likely to be needed.