Design and Analysis of Experiments embedded in Probability Samples

Jan A. van den Brakel

jbrl@cbs.nl

IASS Webinar, 22 September, 2022

1 Statistics Netherlands, Department of Statistical Methods
2 Maastricht University, Department of Quantitative Economics
<table>
<thead>
<tr>
<th></th>
<th>Introduction</th>
<th>Design</th>
<th>Analysis</th>
<th>Factorial designs</th>
<th>Special cases</th>
<th>Application</th>
<th>Discussion</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><strong>Introduction</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td><strong>Design</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td><strong>Analysis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td><strong>Factorial designs</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td><strong>Special cases</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td><strong>Application</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td><strong>Discussion</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Probability sampling and randomized experiments

- Two separated fields of applied statistics
- Unique similarity: design plays a crucial role
- Kish (1965), pp 595: "The separation of sample surveys from experimental designs is an accident in the recent history of science, and it will diminish."
Introduction

Response to Kish:

Experimental Methods in Survey Research
Techniques that Combine Random Sampling with Random Assignment

Edited by
Paul J. Lavrakas • Michael W. Traugott • Courtney Kennedy
Allyson L. Holbrook • Edith D. de Leeuw • Brady T. West
Embedded experiments:

Population U

Sample s

General complex sample design

($\pi_i, \pi_{ij}$)

Randomized experiment (CRD, RBD, factorial design, ...)

Sub sample s_1

Sub sample s_k
Embedded experiments:

- Probability sample from a finite population
- Randomize the sampling units over different treatments

Purpose:

- Estimation population parameters under different survey implementations
- Testing hypotheses about differences or treatment effects

Applications in survey methodology and official statistics:

- Improving survey process of repeated samples, e.g. new questionnaire, data collection modes
- Detecting and quantifying discontinuities
- Quantify non-sampling error sources
Introduction

Literature:

- Interpenetrating subsampling: Mahalanobis (1946), Hartley and Rao (1978)
- Split-ballot designs: Cochran (1977, section 13.15)
- Lavrakas et al (2019)
- ...
Embedding randomized experiments in probability samples:

- Combine strong external validity with strong internal validity
- Design-based or model-based inference?
- Design-based inference generalizes results to a target population
- This paper: design-based inference framework for
  - single factor experiments (CRD, RBD)
  - factorial designs (CRD, RBD)
Design considerations

Clear specifications of / decisions on

- Type and number of treatment factors and levels
- Hypotheses to be tested
  - Main effects
  - Interaction effects
- A prespecified set of target variables
- Minimum required sample size to observe
  - a prespecified treatment effect,
  - at a prespecified significance and power level
Design considerations

- Experimental design (Fienberg and Tanur, 1987, 1988, 1989)
  - Completely Randomized Design (CRD)
  - Randomized Block Design (RBD):
    - strata,
    - clusters,
    - PSU’s
    - interviewers
- Factorial designs versus single factor designs
  - Factorial design combines two or more factors in one experiment
  - Efficient (less sampling units)
  - Analyze interactions
- Experimental units:
  - Ultimate sampling units
  - Cluster of sampling units (households, PSU’s, sampling units assigned to the same interviewer)
Design considerations

- Field staff
  - Selection of interviewers
  - Assignment of interviewers to treatments
  - Are interviewers informed that they participate in an experiment
  - Additional training for the different treatments
  - ...

Field staff
Selection of interviewers
Assignment of interviewers to treatments
Are interviewers informed that they participate in an experiment
Additional training for the different treatments
...
One-factor experiment at $K$ levels

Probability sample:
- Select sample $s$ size $n$ from $U$ of size $N$ ($\pi_i, \pi_{ij'}$)

Experimental design:
- Factor $A$ at $K \geq 2$ levels
- Randomize $n$ units over $K$ treatments
- $\rightarrow K$ subsamples $s_k$ of size $n_k$
- Designs:
  - Completely randomized design
  - Randomized block designs (strata, PSU’s, clusters, interviewers as blocks)
Analysis

Purpose:
- Estimation finite population parameters under different survey implementations
- Testing hypothesis about differences between these subsample estimates

Design-based inference that accounts for:
- Sample design
- Experimental design
Analysis - Measurement error model

- Explains differences between finite population parameter under different survey implementations
- \( y_{iqk} = u_i + \beta_k + \gamma_q + \epsilon_{ik} \)
- Mixed interviewer effect: \( \gamma_q = \psi + \epsilon_q \)
- Treatment effects: \( \beta_k, k = 1, \ldots, K \)
- Model assumptions:

\[
E_m(\epsilon_{ik}) = 0, \quad E_m(\epsilon_q) = 0
\]

\[
\text{Cov}_m(\epsilon_q, \epsilon_{q'}) = \begin{cases} 
\tau_q^2 & : \quad q = q' \\
0 & : \quad q \neq q'
\end{cases}
\]

\[
\text{Cov}_m(\epsilon_{ik}, \epsilon_{i'k'}) = \begin{cases} 
\sigma_{ik}^2 & : \quad i = i', k = k' \\
0 & : \quad i \neq i', k = k'
\end{cases}
\]

\[
\text{Cov}_m(\epsilon_{ik}, \epsilon_q) = 0
\]
Analysis - Hypotheses testing

- \( \bar{Y}_k = 1/N \sum_{i \in U} y_{iqk} \)
- \( \bar{Y} = (\bar{Y}_1, \ldots, \bar{Y}_K)^t \)
- Hypothesis: \( H_0 : \mathbf{C}\bar{Y} = 0, \; H_1 : \mathbf{C}\bar{Y} \neq 0 \)
- Main effects: \( K - 1 \) contrasts:

\[
\mathbf{C} = \begin{pmatrix}
1 & -1 & 0 & \ldots & 0 \\
1 & 0 & -1 & \ldots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
1 & 0 & 0 & \ldots & -1
\end{pmatrix} \equiv (\mathbf{j}_{(K-1)} | - \mathbf{I}_{(K-1)})
\]

- Under measurement error model:

\( \mathbf{C}\hat{Y} = (\beta_1 - \beta_2, \ldots, \beta_1 - \beta_K)^t \)

- Wald test: \( W = \hat{Y}^t \mathbf{C}^t \text{Cov}(\mathbf{C}\hat{Y})^{-1} \mathbf{C}\hat{Y} \)
Analysis - First order inclusion probabilities

**Completely randomized designs:**
- \( P(i \in s_k|s) = \frac{n_k}{n} \)
- First order inclusion probability \( s_k: \pi_i^* = \pi_i(\frac{n_k}{n}) \)

**Randomized block designs:**
- \( s \) divided in \( B \) blocks of size \( n_b \)
- Randomize \( n_b \) units over \( K \) subsamples of size \( n_{bk} \)
- \( P(i \in s_k|s) = \frac{n_{bk}}{n_b} \)
- First order inclusion probability \( s_k: \pi_i^* = \pi_i(\frac{n_{bk}}{n_b}) \)
Analysis - Parameter estimation

General regression estimator for $\bar{Y}_k$ (Särndal et al. (1992)):

- $\hat{Y}_{k;r} = \hat{Y}_{k;\pi} + \hat{b}_k (\bar{X} - \bar{X}_\pi)$
  - $\hat{Y}_{k;\pi} = \frac{1}{N} \sum_{i=1}^{n_k} \frac{Y_{ik}}{\pi_i^*}$
  - $\hat{X}_\pi = \frac{1}{N} \sum_{i=1}^{n_k} \frac{x_i}{\pi_i^*}$
  - $\hat{b}_k = \left( \sum_{i=1}^{n_k} \frac{x_i x_i^t}{\omega_i^2 \pi_i^*} \right)^{-1} \sum_{i=1}^{n_k} \frac{x_i y_{ik}}{\omega_i^2 \pi_i^*}$

General regression estimator for $\bar{Y}$:

- $\hat{Y}_R = (\hat{Y}_{1;r}, \ldots, \hat{Y}_{K;r})^t$
Covariance matrix of $\mathbf{C\hat{Y}_R}$

Sources of variation:
- measurement error model ($m$)
- sample design ($s$)
- experimental design ($e$)

$$\text{Cov}(\mathbf{C\hat{Y}_R}) = \text{Cov}_m E_s E_e(\mathbf{C\hat{Y}_R}) + E_m \text{Cov}_s E_e(\mathbf{C\hat{Y}_R}) + E_m E_s \text{Cov}_e(\mathbf{C\hat{Y}_R})$$
Analysis - Variance estimation

\[
\begin{align*}
&\text{Cov}_m E_s E_e(\mathbf{C\hat{Y}_R}) = \frac{1}{N^2} \sum_{i=1}^{N} \mathbf{C}\Sigma_i \mathbf{C}^t \\
&E_m \text{Cov}_s E_e(\mathbf{C\hat{Y}_R}) = \frac{1}{N^2} \sum_{i=1}^{N} \left( \frac{1}{\pi_i} - 1 \right) \mathbf{C}\Sigma_i \mathbf{C}^t \\
&E_m E_s \text{Cov}_e(\mathbf{C\hat{Y}_R}) = E_m E_s \mathbf{C} \Delta \mathbf{C}^t - \frac{1}{N^2} \sum_{i=1}^{N} \frac{\mathbf{c}\Sigma_i \mathbf{c}^t}{\pi_i}
\end{align*}
\]

where

\[
\begin{align*}
\Sigma_i &= \text{Cov}_m(\epsilon_i), \quad \epsilon_i = (\epsilon_{i1}, \ldots, \epsilon_{iK})^t \\
\mathbf{D} &= \text{Diag}(d_1, \ldots, d_K)
\end{align*}
\]

Collecting results:

\[
\text{Cov}(\mathbf{C\hat{Y}_R}) = E_m E_s \mathbf{C} \Delta \mathbf{C}^t
\]
Randomized block design:

\[
d_k = \sum_{b=1}^{B} \frac{1}{n_{bk}} \frac{1}{n_b-1} \sum_{i=1}^{n_b} \left( \frac{n_b e_{ik}}{N_{\pi_i}} - \frac{1}{n_b} \sum_{i'=1}^{n_b} \frac{n_b e_{i'k}}{N_{\pi_{i'}}} \right)^2
\]

\[
e_{ik} = y_{ik} - b_k^t x_i
\]

Estimator: \( \hat{\text{Cov}}(\hat{C\hat{Y}_R}) = \hat{C\hat{D}\hat{C}}^t \)

\[
\hat{d}_k = \sum_{b=1}^{B} \frac{1}{n_{bk}} \frac{1}{n_{bk}-1} \sum_{i=1}^{n_{bk}} \left( \frac{n_{bk} \hat{e}_{ik}}{N_{\pi_i}} - \frac{1}{n_{bk}} \sum_{i'=1}^{n_{bk}} \frac{n_{bk} \hat{e}_{i'k}}{N_{\pi_{i'}}} \right)^2
\]

\[
\hat{e}_{ik} = y_{ik} - \hat{b}_k^t x_i
\]

Completely randomized design:

\( B = 1, n_b = n, n_{bk} = n_k \)
Analysis - Variance estimation

Variance structure:

- Assumption of the measurement error model:
  - additive treatment effects equal for all elements in the population
  - measurement errors between sampling units are independent

- Weighting model that meets the condition that there is a vector $\mathbf{a}$ such that $\mathbf{a}^t \mathbf{x}_i = 1$ of all $i \in U$

- Super imposition of experimental design on the sampling design

- Variances of contrasts
Analysis - Wald test

Wald statistic

\[ W = \hat{Y}_R^t C^t (C\hat{D}C^t)^{-1} C\hat{Y}_R \text{ with } \hat{D} = \text{Diag}(\hat{d}_1, \ldots, \hat{d}_K) \]

If \( C\hat{Y}_R \rightarrow \mathcal{N} \left( (C\tilde{Y}), \text{Cov}(C\hat{Y}_R) \right) \) then \( W \rightarrow \chi^2_{(K-1)} \)
Analysis - Hypotheses about ratio's

- \( \mathbf{R} = (R_1, \ldots, R_K)^t \) with \( R_k = \frac{\bar{Y}_k}{Z_k} \)

- Hypothesis: \( H_0 : \mathbf{CR} = 0, \; H_1 : \mathbf{CR} \neq 0 \)

- Wald test: \( W = \hat{\mathbf{R}}^t \mathbf{C} \mathbf{t} \; \text{Cov}(\mathbf{CR})^{-1} \mathbf{CR} \)

  - with

    - \( \hat{\mathbf{R}} = (\hat{R}_1, \ldots, \hat{R}_K)^t \), \( \hat{R}_k = \frac{\hat{Y}_{k;r}}{\hat{Z}_{k;r}} \)
    
    - \( \hat{\text{Cov}}(\mathbf{CR}) = \hat{\mathbf{C}} \hat{\mathbf{D}} \hat{\mathbf{C}}^t \), \( \hat{\mathbf{D}} = \text{Diag}(\hat{d}_1, \ldots, \hat{d}_K) \)
    
    - RBD:
      
      \[
      \hat{d}_k = \frac{1}{\hat{Z}_{k;r}^2} \sum_{b=1}^{B} \frac{1}{n_{bk}} \frac{1}{n_{bk}-1} \sum_{i=1}^{n_{bk}} \left( \frac{n_b \hat{e}_{ik}}{N\pi_i} - \frac{1}{n_{bk}} \sum_{i'=1}^{n_{bk}} \frac{n_b \hat{e}_{i'k}}{N\pi_{i'}} \right)^2
      \]
      
      \[
      \hat{e}_{ik} = (y_{ik} - \hat{b}_y^{t} x_i) - \hat{R}_{k;r}(z_{ik} - \hat{b}_z^{t} x_i)
      \]
    
    - CRD: \( B = 1, \; n_b = n, \; n_{bk} = n_k \)
Analysis - randomizing clusters over treatments

- Population $U$ of size $N$
  - $M$ PSU’s of size $N_j$ (and $N = \sum_{j=1}^{M} N_j$)

- Two stage sample $s$
  - First stage: $m$ PSU’s with selection probabilities $\pi_j^I$
  - Second stage: $n_j$ SSU’s with selection probabilities $\pi_{i,j}^{II}$
  - $n = \sum_{j=1}^{m} n_j$

- Experimental design:
  - Randomize $m$ PSU’s over $K$ subsamples $s_k$ of size $m_k$
  - CRD:
    - $P(j \in s_k | s) = \frac{m_k}{m}$
    - $\pi_j^* = P(j \in s_k) = \frac{m_k}{m} \pi_j^I$
  - RBD: $m$ PSU’s grouped in $B$ blocks of size $m_b$
    - $P(j \in s_k | s) = \frac{m_{bk}}{m_b}$
    - $\pi_j^* = P(j \in s_k) = \frac{m_{bk}}{m_b} \pi_j^I$
Analysis - randomizing clusters over treatments

Measurement error model:

- $Y_{ijqk} = u_{ij} + \beta_k + \gamma_q + \epsilon_{ijk}$
- Mixed interviewer effect: $\gamma_q = \psi + \varepsilon_q$
- Treatment effects: $\beta_k, k = 1, \ldots, K$
- Model assumptions:

$$E_m(\epsilon_{ijk}) = 0 \quad E_m(\varepsilon_q) = 0,$$

$$\text{Cov}_m(\varepsilon_q, \varepsilon_{q'}) = \begin{cases} \tau_q^2 & : \ q = q' \\ 0 & : \ q \neq q' \end{cases}$$

$$\text{Cov}_m(\epsilon_{ijk}, \epsilon_{i'j'k'}) = \begin{cases} \sigma_{ijk}^2 + \sigma_{jk}^2 & : \ i = i', j = j', k = k' \\ \sigma_{jk}^2 & : \ i \neq i', j = j', k = k' \\ 0 & : \ i \neq i', j \neq j', k = k' \end{cases}$$

$$\text{Cov}_m(\epsilon_{ijk}, \varepsilon_q) = 0$$
Analysis - randomizing clusters over treatments

- **Hypothesis:** \( H_0 : \mathbf{C} \hat{\mathbf{Y}} = 0, \ H_1 : \mathbf{C} \hat{\mathbf{Y}} \neq 0 \)

- **Wald test:** \( W = \hat{\mathbf{Y}}_R^t \mathbf{C}^t \text{Cov}(\mathbf{C} \hat{\mathbf{Y}}_R)^{-1} \mathbf{C} \hat{\mathbf{Y}}_R \)

- \( \hat{\mathbf{Y}}_R = (\hat{\mathbf{Y}}_1; r, \ldots, \hat{\mathbf{Y}}_K; r)^t \)

- \( \hat{\mathbf{Y}}_k; r = \hat{\mathbf{Y}}_k; \pi + \hat{\mathbf{b}}_k^t (\bar{\mathbf{X}} - \hat{\mathbf{X}}_\pi) \)

  - \( \hat{\mathbf{Y}}_k; \pi = \frac{1}{N} \sum_{j=1}^{m_k} \sum_{i=1}^{n_j} \frac{y_{ijk}}{\pi_{ij}^* \pi_{i|j}^*} \)

  - \( \hat{\mathbf{X}}_\pi = \frac{1}{N} \sum_{j=1}^{m_k} \sum_{i=1}^{n_j} \frac{x_{ij}}{\pi_{ij}^* \pi_{i|j}^*} \)

  - \( \hat{\mathbf{b}}_k = \left( \sum_{j=1}^{m_k} \sum_{i=1}^{n_j} \frac{x_{ij}^t x_{ij}}{\omega_{ij}^2 \pi_{ij}^* \pi_{i|j}^*} \right)^{-1} \sum_{j=1}^{m_k} \sum_{i=1}^{n_j} \frac{x_{ij} y_{ijk}}{\omega_{ij}^2 \pi_{ij}^* \pi_{i|j}^*} \)
Analysis - randomizing clusters over treatments

\[
\widehat{\text{Cov}}(\mathbf{Y}_R) = \mathbf{C}\hat{\mathbf{D}}\mathbf{C}^t, \quad \hat{\mathbf{D}} = \text{Diag}(\hat{d}_1, \ldots, \hat{d}_K)
\]

- RBD:
  \[
  \hat{d}_k = \sum_{b=1}^{B} \frac{1}{m_{bk}} \frac{1}{m_{bk}-1} \sum_{j=1}^{m_{bk}} \left( \frac{m_b \hat{e}_{ik}}{N\pi_j} - \frac{1}{m_{bk}} \sum_{j'=1}^{m_{bk}} \frac{m_b \hat{e}_{j'k}}{N\pi_{j'}} \right)^2
  \]
  \[
  \hat{e}_{jk} = \sum_{i=1}^{n_j} \frac{(y_{ijk} - \hat{b}_k^t x_{ij})}{\pi_{i|j}}
  \]

- CRD: \( B = 1, m_b = m, m_{bk} = m_k \)

- Randomizing clusters of sampling units assigned to the same interviewer

- Extension to analysis of ratio’s
## Analysis - simulation

### Artificial population

<table>
<thead>
<tr>
<th>Stratum</th>
<th>Nr PSU’s</th>
<th>Nr SSU’s</th>
<th>Value target parameter</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>St.Dev.</td>
<td>Min</td>
</tr>
<tr>
<td>1</td>
<td>70</td>
<td>6,250</td>
<td>22,183</td>
</tr>
<tr>
<td>2</td>
<td>130</td>
<td>18,250</td>
<td>6,128</td>
</tr>
<tr>
<td>3</td>
<td>250</td>
<td>85,000</td>
<td>1,407</td>
</tr>
<tr>
<td>Total</td>
<td>450</td>
<td>109,500</td>
<td>3,380</td>
</tr>
</tbody>
</table>
Sample design: stratified two-stage sampling

<table>
<thead>
<tr>
<th>Stratum</th>
<th>Nr PSU’s</th>
<th>Nr SSU’s</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>25</td>
<td>900</td>
</tr>
<tr>
<td>2</td>
<td>30</td>
<td>1,080</td>
</tr>
<tr>
<td>3</td>
<td>50</td>
<td>1,800</td>
</tr>
<tr>
<td>Total</td>
<td>105</td>
<td>3,780</td>
</tr>
</tbody>
</table>
## Analysis - simulation

### Experimental design

<table>
<thead>
<tr>
<th>Experimental design</th>
<th>Treatment effects $\beta_1$</th>
<th>$\beta_2$</th>
<th>$\beta_3$</th>
<th>$\beta_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRD</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>RBD</td>
<td>0</td>
<td>20</td>
<td>40</td>
<td>60</td>
</tr>
<tr>
<td>CRD</td>
<td>0</td>
<td>40</td>
<td>80</td>
<td>120</td>
</tr>
<tr>
<td>RBD</td>
<td>0</td>
<td>80</td>
<td>160</td>
<td>240</td>
</tr>
</tbody>
</table>
Analysis - simulation

Simulation:

- Repeatedly drawing samples from the finite population
- For each replicate generate measurement errors with variances proportional to the target variables
- Randomize the SSU’s of the sample over 4 treatments with a CRD and RBD
- Add treatment effects
- Estimate contrasts, variances and evaluate the Wald test
- Number of replicates: $R = 100,000$
Analysis - simulation

Analysis of the simulation:

- For each replicate $r$:
  - Hájek’s ratio estimator for sample mean:
    $$\hat{Y}_k = \left(\sum_{i \in S_k} \frac{1}{\pi_i^*}\right)^{-1} \sum_{i \in S_k} \frac{y_i}{\pi_i^*}, \quad k = 1, \ldots, 4.$$  
  - Sample means: $\hat{Y}^r = (\hat{Y}^r_1, \ldots, \hat{Y}^r_4)^t$
  - Contrasts $C\hat{Y}^r$ with variance: $C\hat{D}^rC^t$
  - Wald statistic: $W^r = (C\hat{Y}^r)^t(C\hat{D}^rC^t)^{-1}(C\hat{Y}^r)$

- Means over the $R = 100,000$ resamples:
  - $\bar{Y} = \frac{1}{R} \sum_{r=1}^R \hat{Y}^r$
  - Contrasts $C\bar{Y}$
  - Variance: $C\bar{D}C^t = \frac{1}{R} \sum_{r=1}^R C\hat{D}^rC^t$
  - Wald statistic: $\hat{W} = \frac{1}{R} \sum_{r=1}^R W^r$
Analysis - simulation

Analysis of the simulation:

- Point estimates: $\mathbf{C}\tilde{\mathbf{Y}} \approx \mathbf{C}\beta$
- Variance: $\mathbf{C}\tilde{\mathbf{D}}\mathbf{C}^t$ should be approximately equal to:

$$\mathbf{C}\mathbf{V}\mathbf{C}^t = \frac{1}{R} \sum_{r=1}^{R} \mathbf{C}(\hat{\mathbf{Y}}^r - \tilde{\mathbf{Y}})(\hat{\mathbf{Y}}^r - \tilde{\mathbf{Y}})^t \mathbf{C}$$

- Distribution of the Wald test:
  - If $\mathbf{C}\hat{\mathbf{Y}} \rightarrow \mathcal{N}(\mathbf{C}\beta, \mathbf{C}\mathbf{V}\mathbf{C}^t)$, then $W \rightarrow \chi^2_{[K-1][\delta]}$ with n.c.p. $\delta = (1/2)(\mathbf{C}\beta)^t(\mathbf{C}\mathbf{V}\mathbf{C}^t)^{-1}(\mathbf{C}\beta)$
  - Power of the Wald test: $P(\chi^2_{[K-1][\delta]} > \chi^2_{[1-\alpha][K-1]})$
    should be approximately equal to simulated power:

$$P^{\text{sim}}(W) = \frac{1}{R} \sum_{r=1}^{R} I(W^r > \chi^2_{[1-\alpha][K-1]})$$

- $\tilde{W} \approx E(\chi^2_{[K-1][\delta]}) = (K - 1) + 2\delta$
Results RBD $\boldsymbol{\beta} = (0, 80, 160, 240)^t$
Population mean: 3,380

<table>
<thead>
<tr>
<th>Subsamples</th>
<th>$\beta_k$</th>
<th>$\bar{Y}_k$</th>
<th>$k - k'$</th>
<th>C$\bar{Y}$</th>
<th>CVC$^t$</th>
<th>CDC$^t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>3,390</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>80</td>
<td>3,470</td>
<td>1-2</td>
<td>-80</td>
<td>6,204</td>
<td>6,180</td>
</tr>
<tr>
<td>3</td>
<td>160</td>
<td>3,550</td>
<td>1-3</td>
<td>-160</td>
<td>6,210</td>
<td>6,181</td>
</tr>
<tr>
<td>4</td>
<td>240</td>
<td>3,630</td>
<td>1-4</td>
<td>-240</td>
<td>6,214</td>
<td>6,181</td>
</tr>
</tbody>
</table>
Analysis - simulation

Results RBD $\beta = (0, 80, 160, 240)^t$

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>$P(W)$</th>
<th>$P_{\text{sim}}(W)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.050</td>
<td>0.77340</td>
<td>0.77712</td>
</tr>
<tr>
<td>0.025</td>
<td>0.68135</td>
<td>0.68789</td>
</tr>
<tr>
<td>0.010</td>
<td>0.55796</td>
<td>0.56701</td>
</tr>
</tbody>
</table>

$\tilde{W} = 13.4859$

$\delta = 5.1331$

$E(\chi^2_{[K-1][\delta]}) = 13.2662$
Analysis - simulation

Results for one randomly chosen resample
\[ \text{RBD } \beta = (0, 80, 160, 240)^t \]

<table>
<thead>
<tr>
<th>Subsamples</th>
<th>Contrasts</th>
<th>Wald statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>( k ) ( \beta_k ) ( \hat{Y}_k )</td>
<td>( k - k' ) ( \hat{Y}<em>k - \hat{Y}</em>{k'} ) ( \sqrt{\hat{d}<em>k + \hat{d}</em>{k'}} )</td>
<td>( W ) ( df ) ( p )</td>
</tr>
<tr>
<td>1 0 3,395</td>
<td>1-2 -25</td>
<td>81.247</td>
</tr>
<tr>
<td>2 80 3,420</td>
<td>1-3 -120</td>
<td>80.697</td>
</tr>
<tr>
<td>3 160 3,515</td>
<td>1-4 -231</td>
<td>82.383</td>
</tr>
<tr>
<td>4 240 3,626</td>
<td>1-4 -231</td>
<td>82.383</td>
</tr>
</tbody>
</table>
### Analysis - simulation

Results for one randomly chosen resample

RBD $\beta = (0, 80, 160, 240)^t$

<table>
<thead>
<tr>
<th>Subsamples</th>
<th>Contrasts</th>
<th>Descriptives</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k$</td>
<td>$\beta_k$</td>
<td>$\hat{Y}_k$</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>8,815</td>
</tr>
<tr>
<td>2</td>
<td>80</td>
<td>8,150</td>
</tr>
<tr>
<td>3</td>
<td>160</td>
<td>8,566</td>
</tr>
<tr>
<td>4</td>
<td>240</td>
<td>8,746</td>
</tr>
</tbody>
</table>
Analysis - simulation

Results for one randomly chosen resample
RBD $\beta = (0, 80, 160, 240)^t$

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>MSS</th>
<th>$F$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between blocks</td>
<td>2</td>
<td>1.6773 E+11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between treatments</td>
<td>3</td>
<td>84,377,277</td>
<td>1.99</td>
<td>0.1126</td>
</tr>
<tr>
<td>Residual</td>
<td>3,774</td>
<td>42,310,035</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>3,779</td>
<td>131,089,505</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
$K \times L$ factorial experiments

Probability sample:
- Select sample $s$ size $n$ from $U$ of size $N$ ($\pi_i, \pi_{ij'}$)

Experimental design:
- Factor $A$ at $K \geq 2$ levels
- Factor $B$ at $L \geq 2$ levels
- Randomize $n$ units over $KL$ treatment combinations
- $\rightarrow KL$ subsamples $s_{kl}$ of size $n_{kl}$
- Designs:
  - Completely randomized design
  - Randomized block designs (strata, PSU’s, clusters, interviewers as blocks)
**K × L** factorial experiments

Measurement error model

- \( y_{ijkl} = u_i + \beta_{kl} + \gamma_q + \epsilon_{ikl} \)
- Mixed interviewer effect: \( \gamma_q = \psi + \varepsilon_q \)
- Treatment effects ANOVA: \( \beta_{kl} = A_k + B_l + AB_{kl} \)
- Restrictions:
  \[ \sum_{k=1}^{K} A_k = 0, \quad \sum_{l=1}^{L} B_l = 0, \]
  \[ \sum_{k=1}^{K} AB_{kl} = 0, \quad \sum_{l=1}^{L} AB_{kl} = 0, \]
- Model assumptions:

\[
E_m(\epsilon_{ik}) = 0, \quad E_m(\varepsilon_q) = 0 \\
 Cov_m(\varepsilon_q, \varepsilon_{q'}) = \begin{cases} 
\tau_q^2 & : q = q' \\
0 & : q \neq q'
\end{cases} \\
Cov_m(\epsilon_{ik}, \epsilon_{i'k'}) = \begin{cases} 
\sigma_{ik}^2 & : i = i', k = k' \\
0 & : i \neq i', k = k'
\end{cases} \\
Cov_m(\epsilon_{ik}, \varepsilon_q) = 0
\]
**$K \times L$ factorial experiments**

Hypotheses:

1. $\bar{Y}_{kl} = 1/N \sum_{i \in U} y_{iql}$
2. $\bar{Y} = (\bar{Y}_{11}, \ldots, \bar{Y}_{kl}, \ldots, \bar{Y}_{KL})^t$
3. Hypothesis: $H_0 : C \bar{Y} = 0, H_1 : C \bar{Y} \neq 0$

- Main effects A: $K - 1$ contrasts
  
  $C_A = \frac{1}{L} (j_{(K-1)} \mid - I_{(K-1)}) \otimes j_L^t \equiv \frac{1}{L} \tilde{C}_A \otimes j_L^t$

- Main effects B: $L - 1$ contrasts
  
  $C_B = \frac{1}{K} j_K^t \otimes (j_{(L-1)} \mid - I_{(L-1)}) \equiv \frac{1}{K} j_K^t \otimes \tilde{C}_B$

- Interactions: $(K - 1)(L - 1)$ contrasts
  
  $C_{AB} = (j_{(K-1)} \mid - I_{(K-1)}) \otimes (j_{(L-1)} \mid - I_{(L-1)}) = \tilde{C}_A \otimes \tilde{C}_B$
$K \times L$ factorial experiments

Under measurement error model:

- $C_A \bar{Y} = C_A \beta = (A_1 - A_2, \ldots, A_1 - A_K)^t$
- $C_B \bar{Y} = C_B \beta = (B_1 - B_2, \ldots, B_1 - B_L)^t$
- $C_{AB} \bar{Y} = C_{AB} \beta = (AB_{11} - AB_{12} - AB_{21} + AB_{22}, \ldots,$
  $AB_{11} - AB_{1L} - AB_{21} + AB_{2L}, \ldots,$
  $AB_{11} - AB_{12} - AB_{K1} + AB_{K2}, \ldots,$
  $AB_{11} - AB_{1L} - AB_{K1} + AB_{KL})^t$

Wald test: $W = \hat{Y}^t C^t Cov(\hat{C} \hat{Y})^{-1} C \hat{Y}$
**K × L factorial experiments**

Inclusion probabilities

- **CRD:** \( s_{kl}: \pi_i^* = \pi_i(n_{kl}/n) \)
- **RBD:** \( s_{kl}: \pi_i^* = \pi_i(n_{bkl}/n_b) \)

General regression estimator for \( \bar{Y}_{kl} \):

\[
\hat{\bar{Y}}_{kl; r} = \hat{\bar{Y}}_{kl; \pi} + \hat{\mathbf{b}}_{kl}(\bar{X} - \hat{\bar{X}}_{\pi})
\]

\[
\hat{\bar{Y}}_{kl; \pi} = \frac{1}{N} \sum_{i=1}^{n_{kl}} \frac{y_{ikl}}{\pi_i^*},
\]

\[
\hat{\bar{X}}_{\pi} = \frac{1}{N} \sum_{i=1}^{n_{kl}} \frac{x_i}{\pi_i^*},
\]

\[
\hat{\mathbf{b}}_{kl} = \left( \sum_{i=1}^{n_{kl}} \frac{x_i x_i^t}{\omega_i^2 \pi_i^*} \right)^{-1} \sum_{i=1}^{n_{kl}} \frac{x_i y_{ikl}}{\omega_i^2 \pi_i^*}
\]

\[
\hat{\bar{Y}}_R = (\hat{\bar{Y}}_{11; r}, \ldots, \hat{\bar{Y}}_{kl; r}, \ldots, \hat{\bar{Y}}_{KL; r})^t
\]
**K \times L** factorial experiments

Randomized block design:

\[
\text{Cov}(\mathbf{C} \hat{\mathbf{Y}}_R) = E_m E_s \mathbf{C} \mathbf{D} \mathbf{C}^t
\]

\[
d_{kl} = \sum_{b=1}^{B} \frac{1}{n_{bkl}} \frac{1}{n_{b} - 1} \sum_{i=1}^{n_b} \left( \frac{n_b e_{ikl}}{N_{\pi_i}} - \frac{1}{n_b} \sum_{i'=1}^{n_b} \frac{n_b e_{i'kl}}{N_{\pi_{i'}}} \right)^2
\]

\[
e_{ikl} = y_{ikl} - \mathbf{b}_{kl}^t \mathbf{x}_i
\]

Estimator: \( \hat{\text{Cov}}(\mathbf{C} \hat{\mathbf{Y}}_R) = \mathbf{C} \hat{\mathbf{D}} \mathbf{C}^t \)

\[
\hat{d}_{kl} = \sum_{b=1}^{B} \frac{1}{n_{bkl}} \frac{1}{n_{bkl} - 1} \sum_{i=1}^{n_{bkl}} \left( \frac{n_b \hat{e}_{ikl}}{N_{\pi_i}} - \frac{1}{n_{bkl}} \sum_{i'=1}^{n_{bkl}} \frac{n_b \hat{e}_{i'kl}}{N_{\pi_{i'}}} \right)^2
\]

\[
\hat{e}_{ikl} = y_{ikl} - \hat{\mathbf{b}}_{kl}^t \mathbf{x}_i
\]

Completely randomized design:

\( B = 1, n_b = n, n_{bkl} = n_{kl} \)
Wald statistic

- \( W = \hat{Y}_R^t C^t (CDC^t)^{-1} C\hat{Y}_R \) with \( \hat{D} = \text{Diag}(\hat{d}_{11}, \ldots, \hat{d}_{KL}) \)
- Main effects \( A: C_A \hat{Y}_R \) with \( W \sim \chi^2_{K-1} \)
- Main effects \( B: C_B \hat{Y}_R \) with \( W \sim \chi^2_{L-1} \)
- Interaction effects between \( A \) and \( B: C_{AB} \hat{Y}_R \) with \( W \sim \chi^2_{(K-1)(L-1)} \)
Higher order factorial designs
  - Similar parameter and variance estimators
  - Different definitions for contrast matrices

Hypotheses about ratio’s of estimated population totals
  - Adaption variance estimators

Experimental units clusters of sampling units
  - E.g. households, or persons assigned to the same interviewer
  - Adaption variance estimators
Special cases

- Self-weighted sample design and balanced experimental designs
  - \( K \times L \) factorial CRD:
    Wald test main effect equal to F-test ANOVA two-way layout
  - \( K \times L \) factorial RBD:
    Wald test main effect equal to F-test ANOVA three-way layout

- Single factor designs and self-weighted sample design
  - CRD: Wald test equal to F-test ANOVA one-way layout
  - RBD: Wald test equal to F-test ANOVA two-way layout

- Single factor designs with two levels
  - Design-based version of t-test
  - Self-weighted sample design: Welch t-test or t-test
Mixed-mode experiment in the Dutch Crime Victimization Survey (CVS)

- **Survey design:**
  - Annual survey
  - Publication domains: 25 police regions
  - Stratified sample using police regions as strata
  - Sample size:
    - Minimum 750 respondents per domain
    - Net response of about 19,000 respondents
    - With a response of 65% a gross sample of 30,500
  - Data collection: mixed mode using CATI and CAPI
  - Inference: general regression estimator
Application

Mixed-mode experiment Dutch CVS:

- **Purpose**: effect of sequential mixed mode design on response rates and target parameters
- **Sequential mixed-mode design**
  - Starts with Web Interviewing
  - Non-respondent follow-up with CATI and CAPI
- **Embedded experiment**:
  - Control group: regular CVS with 30,500 persons (net response 19,000)
  - Treatment group: 3750 persons (net response 2350)
  - RBD using strata as block variables
Application

Mixed-mode experiment Dutch CVS:

- Target variables:
  - Satispol: percentage of people satisfied with police performance
  - Nuisance: Mean perceived amount of irritation due to antisocial behaviour (10 point scale)
  - Propvic: Percentage of people victim of property crime
  - Violvic: Percentage of people victim of violent crime
  - Repvic: Victimization rate of reported crimes
**Application**

Observable differences at 5% significance and different power levels

<table>
<thead>
<tr>
<th>Variable</th>
<th>Power 50%</th>
<th>Power 80%</th>
<th>Power 90%</th>
<th>Power 50%</th>
<th>Power 80%</th>
<th>Power 90%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Satisfol</td>
<td>4.55</td>
<td>6.49</td>
<td>7.51</td>
<td>2.85</td>
<td>4.07</td>
<td>4.70</td>
</tr>
<tr>
<td>Nuisance</td>
<td>0.06</td>
<td>0.09</td>
<td>1.00</td>
<td>0.04</td>
<td>0.05</td>
<td>0.06</td>
</tr>
<tr>
<td>Propvic</td>
<td>2.06</td>
<td>2.94</td>
<td>3.41</td>
<td>1.29</td>
<td>1.84</td>
<td>2.13</td>
</tr>
<tr>
<td>Violvic</td>
<td>2.30</td>
<td>3.29</td>
<td>3.81</td>
<td>1.44</td>
<td>2.06</td>
<td>2.38</td>
</tr>
<tr>
<td>Repvic</td>
<td>4.49</td>
<td>6.41</td>
<td>7.41</td>
<td>2.81</td>
<td>4.01</td>
<td>4.64</td>
</tr>
</tbody>
</table>
## Application

### Results

<table>
<thead>
<tr>
<th>Var.</th>
<th>Control</th>
<th>Treatment</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \hat{Y}_1 )</td>
<td>( \hat{Y}_2 )</td>
<td>( \hat{\Delta} )</td>
</tr>
<tr>
<td></td>
<td>( \sqrt{\hat{d}_1} )</td>
<td>( \sqrt{\hat{d}_2} )</td>
<td>( t^{(*)} )</td>
</tr>
<tr>
<td></td>
<td>( \hat{\Delta}^{(*)} )</td>
<td>( t^{(**)} )</td>
<td>( p )</td>
</tr>
<tr>
<td>Satisp</td>
<td>55.46 (0.75)</td>
<td>53.95 (2.35)</td>
<td>1.51</td>
</tr>
<tr>
<td></td>
<td>0.610 0.542</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nuis</td>
<td>2.94 (0.01)</td>
<td>3.19 (0.03)</td>
<td>-0.25</td>
</tr>
<tr>
<td></td>
<td>-6.958 0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Propv</td>
<td>16.02 (0.38)</td>
<td>14.47 (1.06)</td>
<td>1.55</td>
</tr>
<tr>
<td></td>
<td>1.377 0.169</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Violv</td>
<td>8.47 (0.34)</td>
<td>8.71 (1.00)</td>
<td>-0.23</td>
</tr>
<tr>
<td></td>
<td>-0.220 0.826</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Repv</td>
<td>35.95 (0.74)</td>
<td>34.36 (2.20)</td>
<td>1.60</td>
</tr>
<tr>
<td></td>
<td>0.492 0.689</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[ \hat{\Delta} = \hat{Y}_1 - \hat{Y}_2 \]

\[ t = \frac{\hat{\Delta}}{\sqrt{\hat{\sigma}_1 + \hat{\sigma}_2}} \]
Discussion

Embedding randomized experiments in probability samples:

- Potentially strong internal and external validity
- Useful to generalize results to large target population
- Design-based inference for a large set of designs
- Many applications in official statistics
- Software: X-tool (Blaise component)
Acknowledgement

Prof. S.E. Fienberg, Prof P. Kooiman, Dr. R. Renssen
Dr. D. Binder
Referees
Thank you for your attention


