Design replication in partial-profile choice experiments

Heiko Großmann
OvGU Magdeburg, DE

mODa 13
10-14 July 2023, Southampton, UK

Outline

- Motivation
- Technical background
- Design replication
- Simulation study

Motivating example

Question 1

Which of the two explanations below would help you most to understand and manage your own osteoarthritis?

Explanation 1A	**Explanation 1B**
Osteoarthritis is the most common form of arthritis in the UK	You have osteoarthritis, a condition which can affect the whole joint and surrounding muscles
Pain, stiffness, and limitation in full movement of the joint are typical	Osteoarthritis can affect your joints in different ways at different times, sometimes you may not have any difficulties but at other times you might
You can take steps to improve your osteoarthritis, by being physically active, maintaining a healthy weight and thinking positively. This can help how you feel and what you can do now, and may help to avoid the need for more treatments in future. Support is available to help you to achieve this	There is no cure for osteoarthritis but there are a number of things that can be done to ease symptoms
Even a modest weight loss can make quite a difference	Even a modest weight loss can make quite a difference

Please tick one box only

Explanation 1A ☐ Explanation 1B ☐

R3-S253-Q2017-B2C01MDP163

Question 5

Which of the two explanations below would help you most to understand and manage your own osteoarthritis?

Explanation 5A	Explanation 5B
Osteoarthritis is caused by an ongoing process of wear and the joint trying to heal itself	Osteoarthritis occurs perhaps because of severe wear and tear to the joints or a problem with the repair process, and osteoarthritis develops
It is mild in many cases; however, about 1 in 10 people aged over 65 years have a major disability due to osteoarthritis	It is mild in many cases; however, about 1 in 10 people aged over 65 years have a major disability due to osteoarthritis
Many people can manage a regular walk	It can be easier to keep moving if you build up from where you are now and put new activities to improve your osteoarthritis in to your daily routine
Keeping active and maintaining a healthy weight are best for your osteoarthritis in the long run, even though some social activities can make this difficult	Many people are afraid to exercise because they believe, mistakenly, that it'll cause further damage to their joints

Please tick one box only

Explanation 5A ☐ Explanation 5B ☐

R1-S022-Q0173-B1C10MDP170

Partial profiles

Full profiles ...

... use **all** factors

A2	A1
B1	B2
C2	C1
D1	D1

Partial profiles

Full profiles ...

... use **all** factors

A2	A1
B1	B2
C2	C1
D1	D1

Partial profiles ...

... use **some** factors

A2	A1
B1	B2
C2	C1
D1	D1

Partial profiles

Full profiles …

… use **all** factors

A2	A1
B1	B2
C2	C1
D1	D1

Partial profiles …

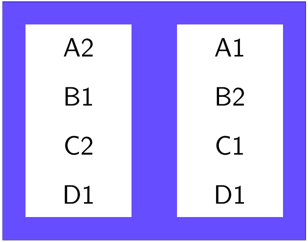
… use **some** factors

A2	A1
C2	C1
D1	D1

Partial profiles

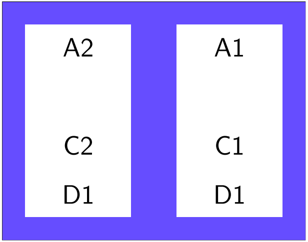
Full profiles ...

... use **all** factors



Partial profiles ...

... use **some** factors



Profile strength S: Number of factors shown

Comparison depth d: Number of shown factors where alternatives differ

MNL model

MNL model

K factors with levels $1, \ldots, v_k$, $k = 1, \ldots, K$, of interest
and extra level 0 to indicate factor is not shown

MNL model

K factors with levels $1, \ldots, v_k$, $k = 1, \ldots, K$, of interest
and extra level 0 to indicate factor is not shown

Pairs $(\mathbf{s}, \mathbf{t})$ of partial profiles where $\mathbf{s} = (s_1, \ldots, s_K)$, $\mathbf{s} = (t_1, \ldots, t_K)$
with profile strength S and comparison depth d

MNL model

K factors with levels $1, \ldots, v_k$, $k = 1, \ldots, K$, of interest
and extra level 0 to indicate factor is not shown

Pairs $(\mathbf{s}, \mathbf{t})$ of partial profiles where $\mathbf{s} = (s_1, \ldots, s_K)$, $\mathbf{s} = (t_1, \ldots, t_K)$
with profile strength S and comparison depth d

MNL probability of choosing $\mathbf{s}$ from pair $(\mathbf{s}, \mathbf{t})$:

$$\boxed{P(\mathbf{s}; (\mathbf{s}, \mathbf{t})) = \frac{\exp[\mathbf{f}^\top(\mathbf{s})\boldsymbol{\beta}]}{\exp[\mathbf{f}^\top(\mathbf{s})\boldsymbol{\beta}] + \exp[\mathbf{f}^\top(\mathbf{t})\boldsymbol{\beta}]}}$$

Design efficiency

Exact partial-profile designs ξ_N: pairs $(\mathbf{s}_1, \mathbf{t}_1), \ldots, (\mathbf{s}_N, \mathbf{t}_N)$

Fisher information matrix of ξ_N in MNL Model: $\mathbf{M}_{\xi_N, \boldsymbol{\beta}}$

Design efficiency

Exact partial-profile designs ξ_N: pairs $(\mathbf{s}_1, \mathbf{t}_1), \ldots, (\mathbf{s}_N, \mathbf{t}_N)$

Fisher information matrix of ξ_N in MNL Model: $\mathbf{M}_{\xi_N, \boldsymbol{\beta}}$

Make indifference assumption $\boldsymbol{\beta} = \mathbf{0}$.

Design efficiency

Exact partial-profile designs ξ_N: pairs $(\mathbf{s}_1, \mathbf{t}_1), \ldots, (\mathbf{s}_N, \mathbf{t}_N)$

Fisher information matrix of ξ_N in MNL Model: $\mathbf{M}_{\xi_N, \boldsymbol{\beta}}$

Make indifference assumption $\boldsymbol{\beta} = \mathbf{0}$. Then

- $\mathbf{M}_{\xi_N, \boldsymbol{\beta}} = \mathbf{M}_{\xi_N, \mathbf{0}} = \frac{1}{4}\mathbf{X}^\top \mathbf{X}$

  where $\mathbf{X}$ has rows $\mathbf{f}^\top(\mathbf{s}_n) - \mathbf{f}^\top(\mathbf{t}_n)$, $n = 1, \ldots, N$

Design efficiency

Exact partial-profile designs ξ_N: pairs $(\mathbf{s}_1, \mathbf{t}_1), \ldots, (\mathbf{s}_N, \mathbf{t}_N)$

Fisher information matrix of ξ_N in MNL Model: $\mathbf{M}_{\xi_N, \boldsymbol{\beta}}$

Make indifference assumption $\boldsymbol{\beta} = \mathbf{0}$. Then

- $\mathbf{M}_{\xi_N, \boldsymbol{\beta}} = \mathbf{M}_{\xi_N, \mathbf{0}} = \frac{1}{4}\mathbf{X}^\top \mathbf{X}$

  where $\mathbf{X}$ has rows $\mathbf{f}^\top(\mathbf{s}_n) - \mathbf{f}^\top(\mathbf{t}_n)$, $n = 1, \ldots, N$

- D-efficiency of ξ

$$\mathrm{eff}_D(\xi_N) = 100 \times \left(\frac{\det(\mathbf{M}_{\xi_N, \mathbf{0}}/N)}{D_{\mathsf{opt}}} \right)^{1/p}$$

  where D_{opt} is determinant of information matrix of D-optimal approximate design

Design efficiency

Exact partial-profile designs ξ_N: pairs $(\mathbf{s}_1, \mathbf{t}_1), \ldots, (\mathbf{s}_N, \mathbf{t}_N)$

Fisher information matrix of ξ_N in MNL Model: $\mathbf{M}_{\xi_N, \boldsymbol{\beta}}$

Make indifference assumption $\boldsymbol{\beta} = \mathbf{0}$. Then

- $\mathbf{M}_{\xi_N, \boldsymbol{\beta}} = \mathbf{M}_{\xi_N, \mathbf{0}} = \frac{1}{4}\mathbf{X}^\top \mathbf{X}$

  where $\mathbf{X}$ has rows $\mathbf{f}^\top(\mathbf{s}_n) - \mathbf{f}^\top(\mathbf{t}_n)$, $n = 1, \ldots, N$
- D-efficiency of ξ

$$\mathrm{eff}_D(\xi_N) = 100 \times \left(\frac{\det(\mathbf{M}_{\xi_N, \mathbf{0}}/N)}{D_{\text{opt}}} \right)^{1/p}$$

  where D_{opt} is determinant of information matrix of D-optimal approximate design
- D_{opt} for ME and ME + 2FI models: GGHS

The question of replication

Given

- maximum number of respondents T
- q choice questions per respondent
- exact base design ξ_N of size $N < qT$

How to replicate ξ_N ?

Approaches to replication

Approaches to replication

- Simple repetition of base design ξ_N

Approaches to replication

- Simple repetition of base design ξ_N
  - R copies of ξ_N such that $RN \approx qT$
  - split copies into distinct sets of q questions

Approaches to replication

- Simple repetition of base design ξ_N
  - R copies of ξ_N such that $RN \approx qT$
  - split copies into distinct sets of q questions
- Repetition of ξ_N with factor permutation

Approaches to replication

- Simple repetition of base design ξ_N
  - R copies of ξ_N such that $RN \approx qT$
  - split copies into distinct sets of q questions
- Repetition of ξ_N with factor permutation
  - Assume: Groups of factors with same number of levels,
    e.g. $u_1^{K_1} \times u_2^{K_2}$ where $K_1 + K_2 = K$

Approaches to replication

- Simple repetition of base design ξ_N
  - R copies of ξ_N such that $RN \approx qT$
  - split copies into distinct sets of q questions

- Repetition of ξ_N with factor permutation
  - Assume: Groups of factors with same number of levels, e.g. $u_1^{K_1} \times u_2^{K_2}$ where $K_1 + K_2 = K$

  - R copies of ξ_N such that $RN \approx qT$

Approaches to replication

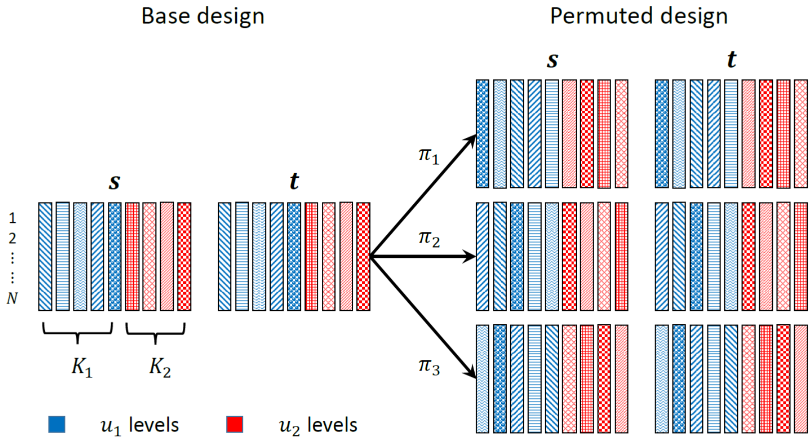
- Simple repetition of base design ξ_N
  - R copies of ξ_N such that $RN \approx qT$
  - split copies into distinct sets of q questions

- Repetition of ξ_N with factor permutation
  - Assume: Groups of factors with same number of levels, e.g. $u_1^{K_1} \times u_2^{K_2}$ where $K_1 + K_2 = K$
  - R copies of ξ_N such that $RN \approx qT$
  - for each copy: randomly permute factors in same group

Approaches to replication

- Simple repetition of base design ξ_N
  - R copies of ξ_N such that $RN \approx qT$
  - split copies into distinct sets of q questions
- Repetition of ξ_N with factor permutation
  - Assume: Groups of factors with same number of levels,
    e.g. $u_1^{K_1} \times u_2^{K_2}$ where $K_1 + K_2 = K$
  - R copies of ξ_N such that $RN \approx qT$
  - for each copy: randomly permute factors in same group
  - split copies into distinct sets of q questions

Illustration of factor permutation



Properties of factor permutation

Consider exact base design ξ_N and design ξ_{RN} consisting of R factor-permuted replicates of ξ_N for ME or ME + 2FI model

Properties of factor permutation

Consider exact base design ξ_N and design ξ_{RN} consisting of R factor-permuted replicates of ξ_N for ME or ME + 2FI model

- $\mathrm{eff}_D(\xi_{RN}) \geq \mathrm{eff}_D(\xi_N)$

  Justification: equivariance of $\mathbf{f}$, convexity and invariance under orthogonal transformations of (appropriate) criterion function for D-optimality

Properties of factor permutation

Consider exact base design ξ_N and design ξ_{RN} consisting of R factor-permuted replicates of ξ_N for ME or ME + 2FI model

- $\text{eff}_D(\xi_{RN}) \geq \text{eff}_D(\xi_N)$

  Justification: equivariance of $\mathbf{f}$, convexity and invariance under orthogonal transformations of (appropriate) criterion function for D-optimality

- $\mathbf{M}_{\xi_N,\mathbf{0}}$ block diagonal $\Rightarrow$ $\mathbf{M}_{\xi_{RN},\mathbf{0}}$ block diagonal

Example for simulation study

Entscheidung 1 von 10:

	Operation A	Operation B
Wahrscheinlichkeit für dauerhafte Inkontinenz	0 von 100 Personen	1 von 100 Personen
Häufigkeit des Wasserlassens in der Nacht	**1** Mal pro Nacht	**3** Mal pro Nacht
Dringlichkeit des Wasserlassens	**Sofort** auf Toilette müssen	**30** Minuten einhalten können
Dauer des Wasserlassens	**6** Minuten	**3** Minuten
Wahrscheinlichkeit einer erneuten Operation	20 von 100 Personen	10 von 100 Personen
Würden Sie Operation A oder B wählen?	◌	◌

Entscheidung 2 von 10:

	Operation A	Operation B
Wahrscheinlichkeit für dauerhafte Inkontinenz	0 von 100 Personen	5 von 100 Personen
Häufigkeit des Wasserlassens in der Nacht	**3** Mal pro Nacht	**5** Mal pro Nacht
Veränderung der Erektionsfähigkeit	Nimmt sehr gering ab	Unverändert
Wahrscheinlichkeit einer erneuten Operation	10 von 100 Personen	0 von 100 Personen
Wahrscheinlichkeit für eine Funktionsstörung des Samenergusses	65 von 100 Personen	5 von 100 Personen
Würden Sie Operation A oder B wählen?	◌	◌

Simulation study

Simulation study

- $3^5 \times 4^4$ main-effects model
  - $p = 22$ parameters
  - $S = d = 5$ factors with distinct levels per question

Simulation study

- $3^5 \times 4^4$ main-effects model
  - $p = 22$ parameters
  - $S = d = 5$ factors with distinct levels per question
- Survey
  - Maximum sample size: 200 respondents
  - Expected response rate: 25%
  - 8 choice pairs per person

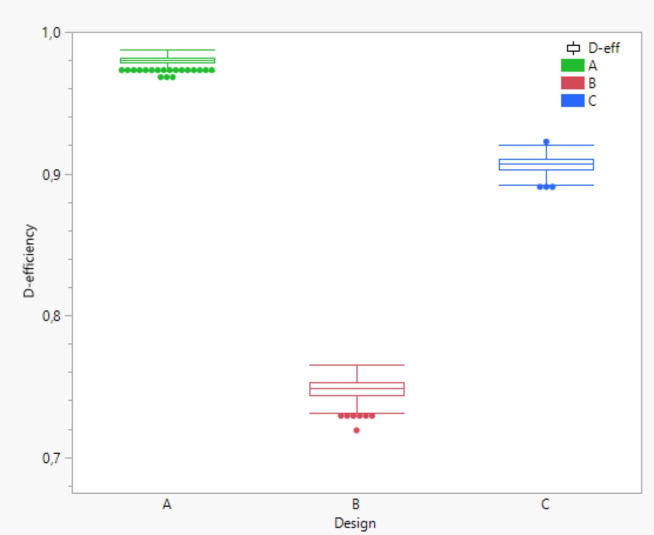
Simulation study

- $3^5 \times 4^4$ main-effects model
  - $p = 22$ parameters
  - $S = d = 5$ factors with distinct levels per question
- Survey
  - Maximum sample size: 200 respondents
  - Expected response rate: 25%
  - 8 choice pairs per person
- Base designs and replication
  - **A** D-opt. (GGS09, $N = 528$ pairs, D-eff: 100%), 3 permuted reps
  - **B** D-eff. (JMP, $N = 44$ pairs, D-eff: $\approx 77\%$), 36 non-permuted reps
  - **C** B, 36 permuted reps

Simulation study

- $3^5 \times 4^4$ main-effects model
  - $p = 22$ parameters
  - $S = d = 5$ factors with distinct levels per question
- Survey
  - Maximum sample size: 200 respondents
  - Expected response rate: 25%
  - 8 choice pairs per person
- Base designs and replication
  - **A** D-opt. (GGS09, $N = 528$ pairs, D-eff: 100%), 3 permuted reps
  - **B** D-eff. (JMP, $N = 44$ pairs, D-eff: $\approx 77\%$), 36 non-permuted reps
  - **C** B, 36 permuted reps
- Simulation of 1000 surveys
  - Per survey: $200/4 = 50$ respondents with 8 pairs each
  - For each of A, B, C and each survey: realized design with 400 pairs
  - Compare D-efficiencies of 1000 realized designs

Results



Conclusions

- When applicable, replication with factor permutation preferable to simple repetition of base design
- Factor permutation increases D-efficiency
- Permutation approach works well with both analytical and algorithmic D-optimal or D-efficient base designs
- Suggestion for algorithmic designs: Use larger number of choice sets than "usual"

References

Graßhoff, U., Großmann, H., Holling, H. & Schwabe, R. (2003). Optimal paired comparison designs for first-order interactions. *Statistics*, 37, 373–386.

Graßhoff, U. Großmann, H., Holling, H., & Schwabe, R. (2004). Optimal designs for main effects in linear paired comparison models. *J. Statist. Plann. Inference*, 126, 361–376.

Großmann, H. (2017). Partial-profile choice designs for estimating main effects and interactions of two-level attributes from paired comparison data. *Journal of Statistical Theory and Practice*, 11, 236–253.

Großmann, H., Graßhoff, U. & Schwabe, R. (2009). Approximate and exact optimal designs for paired comparisons of partial profiles when there are two groups of factors. *J. Statist. Plann. Inference*, 139, 1171–1179.

Großmann, H., Holling, H., Graßhoff, U. & Schwabe, R. (2006). Optimal designs for asymmetric linear paired comparisons with a profile strength constraint. *Metrika*, 64, 109–119.

Luyten, J., Kessels, R., Goos, P. & Beutels, P. (2015). Public preferences for prioritizing preventive and curative health care interventions: a discrete choice experiment. *Value in Health*, 18, 224–233.