Optimal design when outcome values may be missing

Stefanie Biedermann

University of Southampton

Joint work with: Kim May Lee and Robin Mitra

Latest Advances in the Theory and Applications of Design and Analysis of Experiments
BIRS, Banff, Canada, 10 August 2017

1. Introduction
   - Optimal design of experiments for complete data
   - Missing data mechanisms
   - Design of experiments when responses may be missing

2. Results (Approximation, MAR scenarios)
   - Approximation
   - Simulation

3. Results (NMAR)
   - Assessing MAR designs
   - Optimal design under NMAR
   - Case study: Alzheimer’s trial

LINEAR REGRESSION MODEL

$$Y_i = f^T(x_i)\beta + \epsilon_i, \quad i = 1, \ldots, n, \quad \varepsilon_i \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$$

where

- Y_i is the ith response
- $x_i \in \mathcal{X}$ is the experimental condition under which Y_i is observed
- β is a column vector consisting of q unknown parameters
- $f(x)$ is a q-vector of linearly independent regression functions
- ε_i is ‘experimental error’ or natural variation

EXACT DESIGNS

Exact design ξ_n for sample size n:

$$\xi_n = \left\{ \begin{array}{cccc} x_1 & x_2 & \dots & x_m \\ n_1/n & n_2/n & \dots & n_m/n \end{array} \right\}; \; n_i \text{ integers}, \; \sum_{i=1}^{m} n_i = n$$

- Here, $x_1, \dots, x_m$ (where $m \leq n$) are the m different values among the n experimental conditions in the design
- $n_1, \dots, n_m$ are the corresponding replications

Approximate designs

Approximate design ξ:
A probability measure on $\mathcal{X}$, of the form

$$\xi = \begin{Bmatrix} x_1 & x_2 & \dots & x_m \\ w_1 & w_2 & \dots & w_m \end{Bmatrix}, \qquad 0 < w_i \leq 1, \sum_{i=1}^{m} w_i = 1$$

- $x_i \in \mathcal{X}$, $i = 1, \dots, m$: support points of ξ.
- w_i, $i = 1, \dots, m$: weights (proportions) corresponding to x_is.

INFORMATION MATRIX

For completely observed data, the information matrix of the linear model for design

$$\xi = \begin{Bmatrix} x_1 & \cdots & x_m \\ w_1 & \cdots & w_m \end{Bmatrix}$$

is

$$M(\xi) = \sum_{i=1}^{m} f(x_i) f^T(x_i) w_i$$

Optimality criteria

Aim: Estimate the model parameters in β with 'high precision'

- A *D*-optimal design maximises the determinant, $|M(\xi)|$, of the information matrix

  $\hookrightarrow$ minimises the volume of a confidence ellipsoid for β

- An *A*-optimal design minimises the trace of the inverse information matrix, trace$(M(\xi)^{-1})$

  $\hookrightarrow$ minimises the sum of the variances of the elements of $\hat{\beta}$

- A *c*-optimal design (with respect to a vector c) minimises $c^T M(\xi)^{-1} c$, the variance of a linear combination of the elements of $\hat{\beta}$

  $\hookrightarrow$ for estimating $c^T\beta$ most precisely

Missing data mechanisms

Let

$$\mathcal{M}_i = \begin{cases} 1, & \text{if } Y_i \text{ is missing,} \\ 0, & \text{otherwise,} \quad \text{for } i = 1, .., n. \end{cases}$$

Rubin (1976) classifies missing data mechanisms into

- missing completely at random (MCAR): $P(\mathcal{M}_i = 1) = P$
- missing at random (MAR): the probability that a response is missing depends only on observed quantities, e.g. on the design ($P(\mathcal{M}_i = 1) = P(x_i)$)
- not missing at random (NMAR): the probability that a response is missing depends on unobserved quantities, e.g. on the value of the missing response ($P(\mathcal{M}_i = 1|y_i) = P(x_i, y_i)$)

ESTIMATION

There are various methods to analyse incomplete data sets

- Under MCAR and MAR, complete case analysis is a valid method, and leads to unbiased estimates (Little, 1992)
- Complete case analysis is popular with data analysts due to its simplicity

  ↪ In what follows, we will assume the data will be analysed using only the complete cases

ESTIMATION

- Under NMAR, all methods of analysis will result in biased estimates
- Problem: NMAR is untestable

Missing data mechanisms and design

- Imhof, Song and Wong (2002) propose to use the expected information matrix, $E[M(\xi, \mathcal{M})]$, for finding optimal designs, where $\mathcal{M} = (\mathcal{M}_1, \ldots, \mathcal{M}_n)$ and

$$\begin{aligned} E[M(\xi, \mathcal{M})] &= \sum_{i=1}^{m} w_i f(x_i) f^T(x_i)[1 - E[\mathcal{M}_i]] \\ &= \sum_{i=1}^{m} w_i f(x_i) f^T(x_i)[1 - P(\mathcal{M}_i = 1)]. \end{aligned}$$

Missing data mechanisms and design

- Under MCAR, $P(\mathcal{M}_i = 1) = P$, a constant, so optimal designs found assuming all responses will be observed, will still be optimal in this scenario:

$$E[M(\xi, \mathcal{M})] = \sum_{i=1}^{m} w_i f(x_i) f^T(x_i)[1 - P]$$

Missing data mechanisms and design

- Under MAR, $P(\mathcal{M}_i = 1) = P(x_i)$ is a function of x_i, so this approach simply introduces a weighting into the information matrix

$$E[M(\xi, \mathcal{M})] = \sum_{i=1}^{m} w_i f(x_i) f^T(x_i)[1 - P(x_i)]$$

- This scenario is equivalent to design for heteroscedastic linear regression
- While the optimal designs will change, the entire optimal design theory still holds

Missing data mechanisms and design

- However, there is no guidance available on how to deal with NMAR scenarios

Open Problems

Two lines of investigation:

- Optimal design of experiments for NMAR scenarios?
- Is $(E[M(\xi, \mathcal{M})])^{-1}$ a sufficiently close approximation to the covariance matrix?

1. Introduction
   - Optimal design of experiments for complete data
   - Missing data mechanisms
   - Design of experiments when responses may be missing
2. Results (Approximation, MAR scenarios)
   - Approximation
   - Simulation
3. Results (NMAR)
   - Assessing MAR designs
   - Optimal design under NMAR
   - Case study: Alzheimer’s trial

APPROXIMATION

- If all responses are available,

$$M(\xi)^{-1} \propto var(\hat{\beta})$$

- If responses may be missing, $var(\hat{\beta})$ does not exist
- What are we trying to approximate/optimise, and how does $E[M(\xi, \mathcal{M})]$ fit in?

Approximation

- For an exact design ξ, let $\mathcal{C}_\xi$ be the set of values of $\mathcal{M}$ such that $M(\xi, \mathcal{M})$ is non-singular
- Assume that ξ is such that the probability $v_\xi = P(\mathcal{M} \notin \mathcal{C}_\xi)$ is "sufficiently small"
- We can write the 'observed' covariance matrix as $var(\hat{\beta}|\mathcal{M} = \mu)$ where μ is the observed outcome of the vector of missingness indicators $\mathcal{M}$
- Note that this expression will exist if and only if $\mu \in \mathcal{C}_\xi$
- Since v_ξ is close to zero, we will consider only those values where $\mu \in \mathcal{C}_\xi$ to approximate the 'observed' covariance matrix (when it exists) in what follows

Approximation

- At the planning stage of the experiment, the observed value of μ is not known, and *var*$(\hat{\beta}|\mathcal{M})$ (where $\mathcal{M} \in \mathcal{C}_\xi$) is a random variable
- To approximate the 'observed' covariance matrix we take the expectation of *var*$(\hat{\beta}|\mathcal{M})$ with respect to the distribution of $\mathcal{M}$, constrained to $\mathcal{M} \in \mathcal{C}_\xi$,

$$E_{\mathcal{M}|\mathcal{M}\in\mathcal{C}_\xi}(var(\hat{\beta}|\mathcal{M})) = E_{\mathcal{M}|\mathcal{M}\in\mathcal{C}_\xi}\{[M(\xi,\mathcal{M})^{-1}]\}$$

- The expectation $E_{\mathcal{M}|\mathcal{M}\in\mathcal{C}_\xi}\{[M(\xi,\mathcal{M})^{-1}]\}$ is not normally available in closed form

APPROXIMATION

- We propose to apply a second order Taylor expansion to the respective elements of $M(\xi, \mathcal{M})^{-1}$, and then to take the expectation (where $\mathcal{M} \in \mathcal{C}_\xi$) of these
- In this context, the Imhof et al (2002) approach corresponds to a Taylor expansion of order zero/one, where the expectation is taken over the original distribution of $\mathcal{M}$

$\hookrightarrow$ Will there be any differences between the two approaches in practice?

Illustration

Consider the simple linear regression model:

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad i = 1, \ldots, n, \quad \epsilon_i \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$$

and a two-point design $\{x_1^*, x_2^*; n_1, n_2\}$ where $n_1 + n_2 = n$. Then,

$$M(\xi, \mathcal{M})^{-1} = \frac{1}{\left(x_1^* - x_2^*\right)^2 Z_1 Z_2} \begin{pmatrix} x_1^{*2} Z_1 + x_2^{*2} Z_2 & -x_1^* Z_1 - x_2^* Z_2 \\ -x_1^* Z_1 - x_2^* Z_2 & Z_1 + Z_2 \end{pmatrix},$$

where $Z_1 = \sum_{i=1}^{n_1} (1 - \mathcal{M}_i)$ and $Z_2 = \sum_{i=n_1+1}^{n} (1 - \mathcal{M}_i)$

Illustration

- $Z_j \sim Bin(n_j,\ 1 - P(x_j^*)), j = 1, 2$
- $\mathcal{C}_\xi = \{\mathcal{M} \in \{0,1\}^n; Z_1 > 0, Z_2 > 0\}$
- $v_\xi = P(x_1^*)^{n_1} + P(x_2^*)^{n_2} - P(x_1^*)^{n_1} P(x_2^*)^{n_2}$
- Hence we will consider the corresponding zero truncated binomial distributions for Z_1 and Z_2, respectively

ILLUSTRATION

Taking expectation (with respect to the zero truncated binomial random variables) of a second order Taylor series expansion about $E\{Z_j\}$ yields

$$E\left(\frac{1}{Z_j}\right) \approx \frac{1}{E\{Z_j\}} + \frac{Var(Z_j)}{(E\{Z_j\})^3} = \frac{(1-P(x_j^*)^{nw_j})^2\{P(x_j^*)+nw_j(1-P(x_j^*))\}}{(nw_j)^2(1-P(x_j^*))^2}$$

for $j = 1, 2$, and we can substitute this expression into the respective optimality criterion

Simulation

Setup:

- Simple linear regression model:

$$Y_i = 1 + x_i + \epsilon_i, \quad i = 1, \ldots, n, \quad \epsilon_i \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$$

- Logistic missing data indicator:

$$P(x) = \frac{\exp(\gamma_0 + \gamma_1 x)}{1 + \exp(\gamma_0 + \gamma_1 x)}$$

- 200,000 simulation runs
- $n = 30$, $\gamma_0 = -4.572$, $\gamma_1 = 3.191$, $\mathcal{X} = [0, \infty)$

Simulation

Simulated ‘observed’ covariance matrix for two different arbitrary designs with $n_1 = n_2 = 15$ and $P(x_1) = 0.01$.

$\{x_1, x_2\}$	$\{0, 1\}$	$\{0, 1.5\}$
$[1, 1]$ element of covariance matrix	0.06740	0.06740
First order Taylor series approximation	0.06736	0.06736
Second order Taylor series approximation	0.06740	0.06740
$[2, 2]$ element of covariance matrix	0.15242	0.10375
First order Taylor series approximation	0.15078	0.09628
Second order Taylor series approximation	0.15222	0.10177
$[1, 2]$ element of covariance matrix	-0.06740	-0.04494
First order Taylor series approximation	-0.06736	-0.04490
Second order Taylor series approximation	-0.06740	-0.04493
Determinant of covariance matrix	0.00573	0.00497
First order Taylor series approximation	0.00562	0.00447
Second order Taylor series approximation	0.00572	0.00484
No. of cases failed	0	23
$P(x_2)$	0.20085	0.55342

Simulation

Optimal designs found using the two approximations, respectively. The other support point is $x_1^* = 0$ and $P(x_1^*) = 0.01$.

	$\xi^*_{A\ 2nd}$	$\xi^*_{A\ 1st}$	$\xi^*_{c\ 2nd}$	$\xi^*_{c\ 1st}$	$\xi^*_{D\ 2nd}$	$\xi^*_{D\ 1st}$
x_2^*	1.4630	1.51466	1.5497	1.60059	1.3360	1.37660
w_2	0.4664	0.4539	0.6257	0.6208	0.5110	0.5
$P(x_2^*)$	0.5241	0.5650	0.5922	0.6308	0.4234	0.4553
v_ξ	1.186 e-04	3.378 e-04	5.359 e-05	1.577 e-04	1.897 e-06	7.4897 e-06

The larger support point is smaller for the second order designs, but has higher weight

Simulation

Simulated criterion values for different designs. The numbers in the last row indicate the frequency of the cases where $M(\xi, \mathcal{M})$ becomes singular

	sample $var(\hat{\beta}_1)$	tr(sample $var(\hat{\boldsymbol{\beta}})$)	|sample $var(\hat{\boldsymbol{\beta}})$|	Failures
$\xi^*_{A\ 2nd}$	1.0690e-01	**1.6992e-01**	4.8805e-03	19
$\xi^*_{A\ 1st}$	1.0823e-01	1.7123e-01	5.0880e-03	67
$\xi^*_{c\ 2nd}$	**9.7359e-02**	1.8894e-01	5.4195e-03	16
$\xi^*_{c\ 1st}$	9.8102e-02	1.8968e-01	5.7121e-03	35
$\xi^*_{D\ 2nd}$	1.0400e-01	1.7590e-01	**4.5807e-03**	0
$\xi^*_{D\ 1st}$	1.0486e-01	1.7197e-01	4.6526e-03	2

FINDINGS

- We used several further values for the parameters γ_0 and γ_1, and different sample sizes
- For smaller sample sizes, e.g. $n = 30$, the second order approximations were slightly closer, and the corresponding optimal designs tended to generate fewer failures
- If we use the second order expansion, convexity of the criterion function is no longer guaranteed
- For sample sizes ≥ 60, there was hardly any difference between the two approximation methods
- In the next section on NMAR scenarios we will assume large enough sample sizes to use the simpler approximation

1. Introduction
   - Optimal design of experiments for complete data
   - Missing data mechanisms
   - Design of experiments when responses may be missing

2. Results (Approximation, MAR scenarios)
   - Approximation
   - Simulation

3. Results (NMAR)
   - Assessing MAR designs
   - Optimal design under NMAR
   - Case study: Alzheimer's trial

PROBLEM 1:

How well will designs found under MAR assumption perform if the true missing data mechanism is NMAR?

Simulation

Consider the simple linear regression model

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, \ldots, n, \;\; \varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$$

- For finding A- and D-optimal designs, assume

$$P(x_i) = \frac{\exp(\gamma_0 + \gamma_1 x_i)}{1 + \exp(\gamma_0 + \gamma_1 x_i)}$$

- For generating the missing data indicators, use

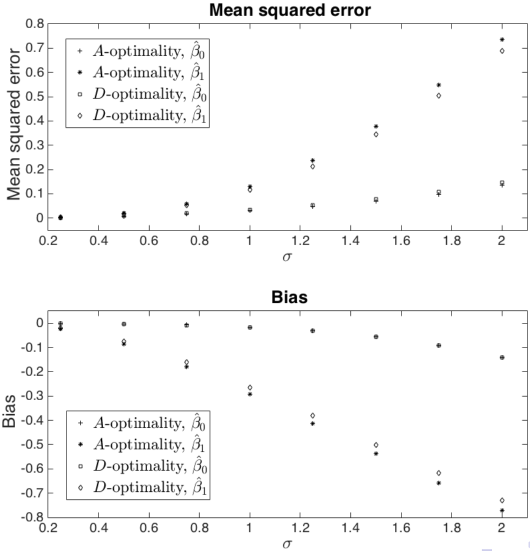
$$P(x_i, y_i) = \frac{\exp(\tilde{\gamma}_0 + \tilde{\gamma}_1 x_i + y_i)}{1 + \exp(\tilde{\gamma}_0 + \tilde{\gamma}_1 x_i + y_i)}$$

where $\tilde{\gamma}_j + \beta_j = \gamma_j$, $j = 0, 1$

SIMULATION

For the choices $(\tilde{\gamma}_0, \tilde{\gamma}_1, \beta_0, \beta_1) = (-5.572, 2.191, 1, 1)$ (and hence $(\gamma_0, \gamma_1) = (-4.572, 3.191)$), we find the *A*- and the *D*-optimal design under MAR, then generate data under NMAR and analyse these using complete case analysis (100,000 simulation runs)

SIMULATION


Mean squared error
Mean squared error
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2
σ
+ A-optimality, $\hat{\beta}_0$
* A-optimality, $\hat{\beta}_1$
□ D-optimality, $\hat{\beta}_0$
◇ D-optimality, $\hat{\beta}_1$
Bias
Bias
0
-0.1
-0.2
-0.3
-0.4
-0.5
-0.6
-0.7
-0.8
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2
σ
+ A-optimality, $\hat{\beta}_0$
* A-optimality, $\hat{\beta}_1$
□ D-optimality, $\hat{\beta}_0$
◇ D-optimality, $\hat{\beta}_1$

Simulation

As σ^2 increases, i.e. the further away we get from the MAR scenario, the larger the absolute value of the bias, and the MSE

$\hookrightarrow$ The optimal MAR designs do not perform well under NMAR

PROBLEM 2:

How do we approximate the covariance matrix under NMAR?

- Assume large sample size ($n \geq 60$) and use the Imhof, Song and Wong (2002) approach
- We have the expression $P(\mathcal{M}_i = 1)$ in the information matrix
- Recall: $P(\mathcal{M}_i = 1|y_i) = P(x_i, y_i)$ depends on the unobserved value of y_i
- $\hookrightarrow$ Use the expected value of $P(x_i, Y_i)$

Problem 2:

- If $Y_i \sim N(f^T(x_i)\beta, \sigma^2)$, then

$$P(x_i, Y_i) = \frac{\exp(\tilde{\gamma}_0 + \tilde{\gamma}_1 x_i + Y_i)}{1 + \exp(\tilde{\gamma}_0 + \tilde{\gamma}_1 x_i + Y_i)}$$

  follows a logit-normal distribution

- There is no closed form for the expectation of a logit-normal distribution, so we used the *integral* function in Matlab to evaluate it

- We tried several simpler approximations, e.g. the median, but neither of these performed well

Problem 3:

What can we do about the bias under NMAR?

Consider the mean squared error matrix rather than the covariance matrix in the optimality criterion

$$\begin{aligned} m.s.e.(\hat{\beta}) &= E[(\hat{\beta}-\beta)(\hat{\beta}-\beta)^T] \\ &= var(\hat{\beta}) + \left[E(\hat{\beta})-\beta\right]\left[E(\hat{\beta})-\beta\right]^T \end{aligned}$$

How to approximate the bias $E(\hat{\beta})-\beta$?

PROBLEM 3:

What can we do about the bias under NMAR?

Consider the mean squared error matrix rather than the covariance matrix in the optimality criterion

$$\begin{aligned} m.s.e.(\hat{\beta}) &= E[(\hat{\beta}-\beta)(\hat{\beta}-\beta)^T] \\ &= var(\hat{\beta}) + \left[E(\hat{\beta})-\beta\right]\left[E(\hat{\beta})-\beta\right]^T \end{aligned}$$

How to approximate the bias $E(\hat{\beta})-\beta$?

APPROXIMATING THE BIAS

- The bias is likely to depend on σ^2 (see simulations) and on the design
- For a given sample size n, define a grid for values of σ^2 and the design variables $x_1, \ldots, x_m$, $n_1, \ldots, n_m$ where $\sum_{i=1}^{m} n_i = n$
- For some selected values from the grid, simulate data using the NMAR model, and estimate the parameters via complete case analysis
- Repeat a large number of times, and use the average empirical bias for each grid value as an ‘observation’ from the unknown bias function
- Fit a model to these ‘data’, e.g. a Gaussian process model, and use this predicted response surface to approximate the bias in the MSE

OPTIMAL NMAR DESIGNS

A- and *D*-optimal designs for the example, for $n = 60$, $\mathcal{X} = [0, \infty)$, and different values of σ^2. The lower support point is 0 in all cases.

		MAR	$\sigma = 1$	$\sigma = 1.5$	$\sigma = 2$
D-optimal	x_2^*	1.3766	0.9793	1.0202	1.1210
design	$w_2(n_2)$	0.5000(30)	0.3811 (23)	0.3194 (19)	0.2879 (17)
A-optimal	x_2^*	1.5147	1.0871	1.0617	1.0671
design	$w_2(n_2)$	0.4539(27)	0.4462 (27)	0.4508 (27)	0.4534 (27)

- The choice of the parameter values makes 0 the point with the lowest probability of missingness
- Incorporating the NMAR mechanism results in smaller values of the larger support point - reduces the probability of Y_i missing

Comparison of designs under NMAR

$\sigma^2 = 1$ in generating y_i and in the NMAR mechanism

	D-optimal design that assumes			
	MAR	$\sigma = 1$	$\sigma = 1.5$	$\sigma = 2$
bias of $\hat{\beta}_0$	-0.015710	-0.015657	-0.015559	-0.015525
bias of $\hat{\beta}_1$	-0.26664	-0.18472	-0.19344	-0.21511
$m.s.e.\ (\hat{\beta}_0)$	0.033581	0.027279	0.024665	0.023522
$m.s.e.\ (\hat{\beta}_1)$	0.11689	0.11449	0.12077	0.12403
$tr(m.s.e.\ (\hat{\boldsymbol{\beta}}))$	0.15047	0.14176	0.14544	0.14756
$\|m.s.e.\ (\hat{\boldsymbol{\beta}})\|$	0.0035232	**0.0025149**	0.0025445	0.0026165

	A-optimal design that assumes			
	MAR	$\sigma = 1$	$\sigma = 1.5$	$\sigma = 2$
bias of $\hat{\beta}_0$	-0.015717	-0.015717	-0.015717	-0.015717
bias of $\hat{\beta}_1$	-0.29240	-0.20739	-0.20208	-0.20313
$m.s.e.\ (\hat{\beta}_0)$	0.030604	0.030604	0.030604	0.030604
$m.s.e.\ (\hat{\beta}_1)$	0.13022	0.10697	0.10728	0.10713
$tr(m.s.e.\ (\hat{\boldsymbol{\beta}}))$	0.16083	**0.13758**	0.13788	0.13774
$\|m.s.e.\ (\hat{\boldsymbol{\beta}})\|$	0.0037448	0.0026704	0.0026408	0.0026451

Comparison of designs under NMAR

$\sigma^2 = 1.5^2$ in generating y_i and in the NMAR mechanism

	D-optimal design that assumes			
	MAR	$\sigma = 1$	$\sigma = 1.5$	$\sigma = 2$
bias of $\hat{\beta}_0$	-0.054443	-0.054393	-0.054202	-0.054178
bias of $\hat{\beta}_1$	-0.50182	-0.38675	-0.39934	-0.42936
$m.s.e.\ (\hat{\beta}_0)$	0.076555	0.062639	0.056827	0.054331
$m.s.e.\ (\hat{\beta}_1)$	0.34630	0.32185	0.33703	0.34929
$tr(m.s.e.\ (\hat{\boldsymbol{\beta}}))$	0.42285	0.38449	0.39386	0.40362
$\lvert m.s.e.\ (\hat{\boldsymbol{\beta}})\rvert$	0.025828	0.018580	**0.018181**	0.018456

	A-optimal design that assumes			
	MAR	$\sigma = 1$	$\sigma = 1.5$	$\sigma = 2$
bias of $\hat{\beta}_0$	-0.054465	-0.054465	-0.054465	-0.054465
bias of $\hat{\beta}_1$	-0.53864	-0.41838	-0.41095	-0.41264
$m.s.e.\ (\hat{\beta}_0)$	0.070012	0.070012	0.070012	0.070012
$m.s.e.\ (\hat{\beta}_1)$	0.37910	0.31198	0.31145	0.31162
$tr(m.s.e.\ (\hat{\boldsymbol{\beta}}))$	0.44912	0.38199	**0.38146**	0.38163
$\lvert m.s.e.\ (\hat{\boldsymbol{\beta}})\rvert$	0.026319	0.020325	0.020139	0.020183

Case study: Alzheimer's trial

- Howard et al. (2012) describe a trial with originally 72 patients in each of two groups (active treatment/placebo)
- They fit a simple linear regression model to the response 'change of SMMSE score from baseline (after 52 weeks)'
- After 52 weeks, only 26 patients in the placebo group and 49 patients in the treatment group come back for their tests
- We fit an NMAR model to the data and use the estimates to redesign the trial for $n = 144$

CASE STUDY: ALZHEIMER'S TRIAL

- We find the *A*-optimal design to be: 95 patients in the placebo group and 49 in the treatment group. (The support points are fixed here, $x_1 = 0$ and $x_2 = 1$, as there are only two groups.)
- Simulations show:

n_2	52	51	50	49	72
$tr(m.s.e.\ (\hat{\beta}))(\times 10^{-4})$	3.2950	3.2927	3.2934	**3.2919**	3.6155

- There is about a 9% $[(1 - 3.2919/3.6155) \times 100\%]$ efficiency loss if we use the equal sample size design instead of the optimal design.

CONCLUSION AND FUTURE WORK:

- This is the first approach to mitigate the problems caused by NMAR missingness through designed experiments
- The designs are locally optimal, so robustness with respect to parameter values and the form of the NMAR mechanism needs to be assessed
- We could try to make the designs more robust to parameter misspecifications by using prior distributions
- Better approximations for the bias function should be investigated. (Here we used second order response surfaces, but consider Gaussian processes for future work)
- Choice of grid values for simulating the bias function?
- Extensions to nonlinear and generalised linear models

Thank you!

REFERENCES:

- Howard, R., McShane, R., Lindsay, J., Ritchie, C., Baldwin, A., Barber, R., ... and Phillips, P. (2012). Donepezil and memantine for moderate-to-severe Alzheimer's disease. *New England Journal of Medicine* **366(10)**, 893-903.
- Imhof, L. A and Song, D. and Wong, W. K. (2002). Optimal design of experiments with possibly failing trials. *Statistica Sinica* **12**, 1145-1155.
- Lee, K.M., Biedermann, S. and Mitra, R. (2017). Optimal design for experiments with possibly incomplete observations. *Statistica Sinica*, in press.
- Lee, K.M., Mitra, R. and Biedermann, S. (2017). Optimal design when outcome values are not missing at random. *Statistica Sinica*, in press.
- Little, R. J. A. (1992). Regression with missing X's: a review. *Journal of the American Statistical Association* **87**, 1227-1237.
- Rubin, D.B (1976). Inference and missing data. *Biometrika* **63**, 581-592.