Dr. Kelvin Chan, Medical Oncologist, Associate Scientist, Odette Cancer Centre, Sunnybrook Health Sciences Centre
and
Dr. Eleanor Pullenayegum, Senior Scientist, Hospital for Sick Children

Title: Underestimation of Variance of Predicted Mean Health Utilities Derived from Multi-Attribute Utility Instruments: The Use of Multiple Imputation as a Potential Solution.

Abstract: Parameter uncertainty in value sets (of health utilities) of multi-attribute utility-based instruments (MAUIs) has received little attention previously. This false precision leads to underestimation of the uncertainty of the results of cost-effectiveness analyses. This may result in drug funding decisions based on invalid estimates of cost-effectiveness of cancer drugs. In this presentation, we will illustrate the presence and magnitude of this parameter uncertainty by fitting a Bayesian model with random effects for respondents and health states to the data from the original US EQ-5D-3L valuation study, thereby estimating the uncertainty in the EQ-5D-3L scoring algorithm. We will then demonstrate the use of multiple imputation as a method to account for this uncertainty of MAUI scoring algorithms. We will apply these methods to the EQ-5D-3L data from the Commonwealth Fund (CWF) Survey for Sick Adults (n=3958), comparing the standard error of the estimated mean utility in the CWF population using the predictive distribution from the Bayesian mixed effect model (i.e., incorporating parameter uncertainty in the value set) with the standard error of the estimated mean utilities based on multiple imputation and the standard error using the conventional approach of using MAUI (i.e., ignoring uncertainty in the value set). This presentation will allow us to appreciate that ignoring uncertainty of the predicted health utilities derived from MAUIs could lead to substantial underestimation of the variance of mean utilities. The use of multiple imputation method can correct for this underestimation so that the results of cost-effectiveness analyses using MAUIs can report the correct degree of uncertainty.

Underestimation of Variance of Predicted Mean Health Utilities Derived from Multi-Attribute Utility Instruments: The Use of Multiple Imputation as a Potential Solution. ,

Eleanor Pullenayegum PhD ,
Biostatistician, Senior Scientist, Child Health Evaluative Science ,
Hospital of Sick Children ,
Kelvin Chan, MD FRCPC MSc (Clin Epi) MSc (Biostats) PhD ,
Staff Medical Oncologist, Sunnybrook Health Sciences Centre, Toronto, Canada ,
Co-Director, Canadian Centre for Applied Research in Cancer Control (ARCC) ,
Clinical Lead, Provincial Drug Reimbursement Programs, Cancer Care Ontario ,

SickKids® RESEARCH INSTITUTE

Sunnybrook HEALTH SCIENCES CENTRE when it matters MOST

Health utilities '

…affect you

…are reported alongside underestimates of uncertainty

We aim to explain:

- why you should care
- where the uncertainty comes from
- what to do about it

Health utilities are used to decide which treatments to reimburse '

CADTH Evidence Driven.

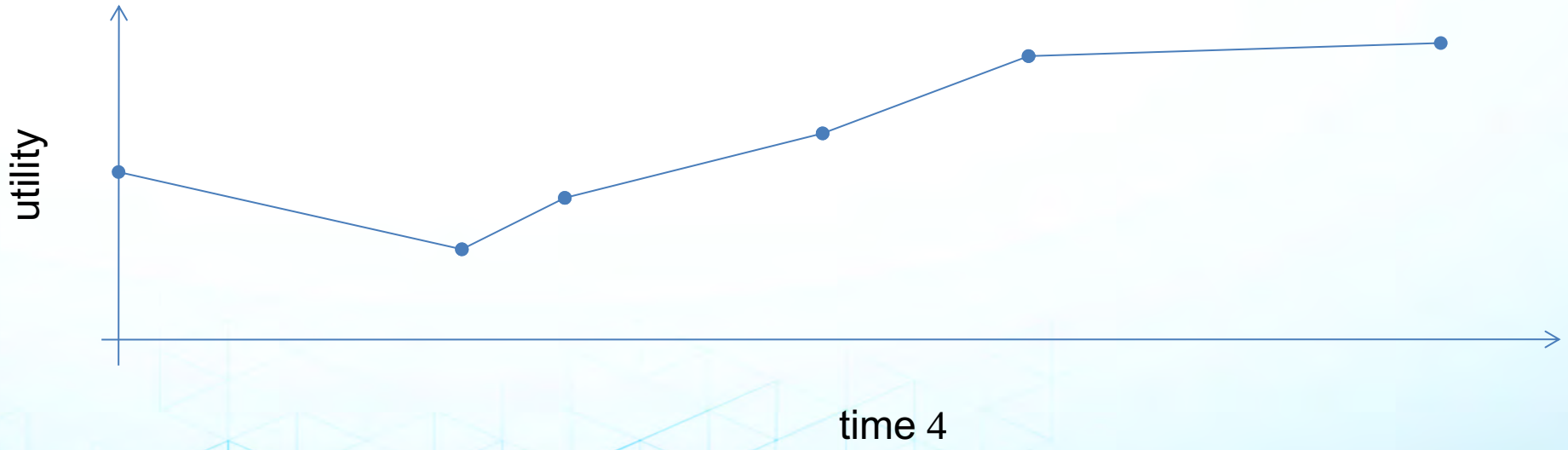
Instituts thématiques

Inserm

Institut national de la santé et de la recherche médicale

QALY = quality-adjusted life year 4

$$\text{QALYs at time t} = \int_0^t \text{utility}(s)ds$$



Why is this topic important in Oncology? $

- $ Funding of cancer drugs in Canada depends on evaluation of cost-effectiveness (as part of Health Technology Assessment)
- $ Pan-Canadian Oncology Drug Review (pCODR) at the Canadian Agency of Drug and Technology in Health (CADTH)
- $ CADTH has recently published an updated guideline for economic evaluation for HTA in Canada (2017) suggesting that the base case should be a probabilistic analysis (not deterministic)

Why is this topic important in Oncology? $

- $ The main outcome of a cost-effectiveness analysis is called incremental cost-effectiveness ratio (ICER)

$$ICER = \frac{\Delta C}{\Delta E}$$

- $ Incremental Cost = ΔC
- $ Incremental Effectiveness = ΔE

Why is this topic important in Oncology? $

$$E(ICER) = E(\frac{\Delta C}{\Delta E}) \neq \frac{E(\Delta C)}{E(\Delta E)}$$

- $ Therefore, probabilistic analysis is essential to fully account for the joint distributions of all the parameters in the model to estimate an unbiased ICER and its distribution (degree of uncertainty)

Why is this topic important in Oncology? $

$$ICER = \frac{\Delta C}{\Delta E}$$

- $ In Oncology cost-effectiveness analysis,

$$\Delta E = \Delta QALY = \Delta(utility \times survival)$$

- $ Hence, the importance of capturing the uncertainty of utility adequately
- $ Otherwise, drug funding decision may be based on invalid estimate of the distribution of ICER

Why does uncertainty matter? $

- Reimbursement decision making
  - Evidence based
  - ***Estimates*** of incremental cost and utility (ICERs, etc.)
  - Quality of estimates matters
    - ICER of $40,000 per QALY with 95% CI ($5,000 to $300,000)
    - ICER of $40,000 per QALY with 95% CI ($35,000 to $50,000)

Health Utility and QALYs '

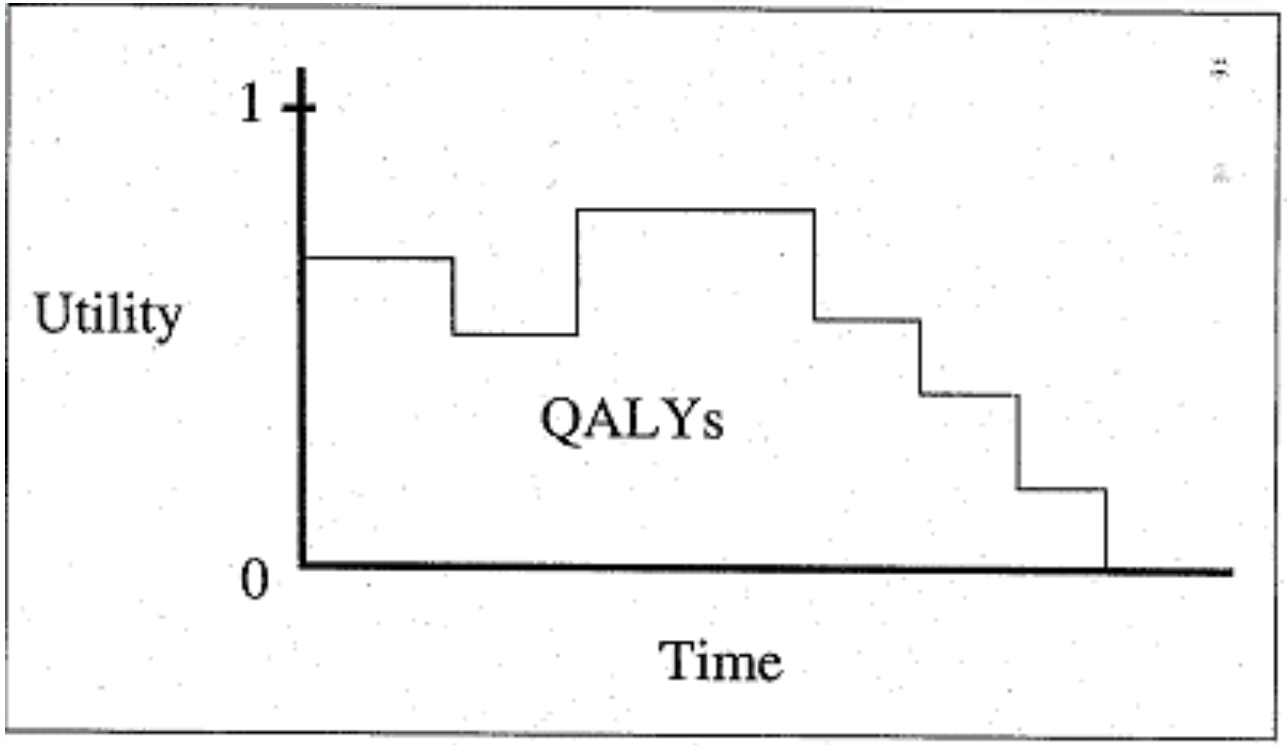
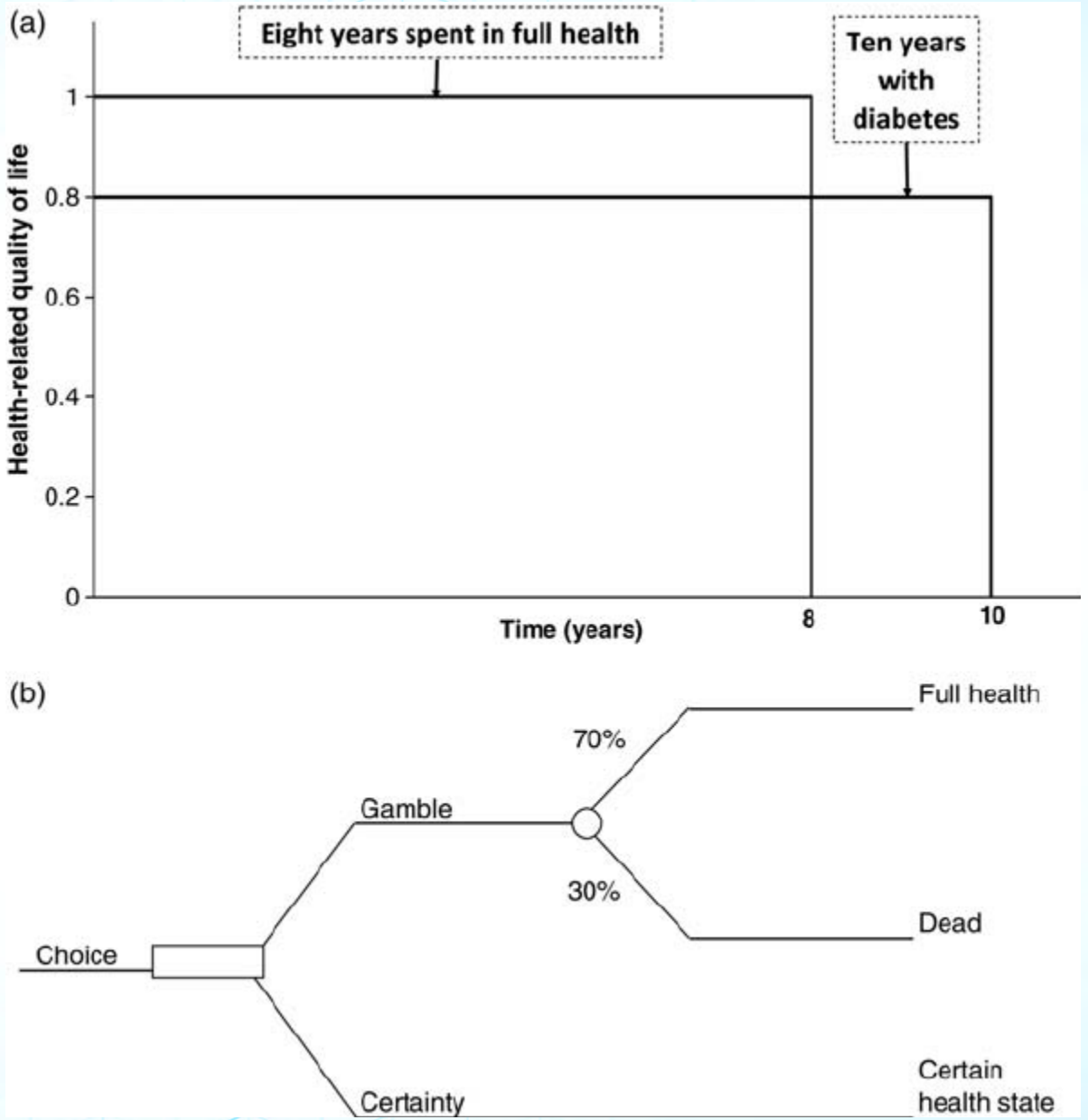


Figure: Determining Quality-Adjusted Survival—Length of life (time) is plotted against quality of life (utility). The area under the curve represents quality-adjusted survival measured in quality-adjusted life years (QALYs).

Direct Measures '

(a) Time Trade-Off (TTO)

(b) Standard Gamble (SG)



MAUI – Example EQ-5D %

By placing a tick in one box in each group, please indicate which statements best describe your health today.

Mobility
- I have no problems in walking about
- I have some problems in walking about
- I am confined to bed

Self-Care
- I have no problems with self-care
- I have some problems washing or dressing myself
- I am unable to wash or dress myself

Usual Activities *(e.g. work, study, housework, family or leisure activities)*
- I have no problems with performing my usual activities
- I have some problems with performing my usual activities
- I am unable to perform my usual activities

Pain/Discomfort
- I have no pain or discomfort
- I have moderate pain or discomfort
- I have extreme pain or discomfort

Anxiety/Depression
- I am not anxious or depressed
- I am moderately anxious or depressed
- I am extremely anxious or depressed

Levels of perceived problems are coded as follows:

Level 1 is coded as a '1'

Level 2 is coded as a '2'

Level 3 is coded as a '3'

NB: There should be only one response for each dimension.

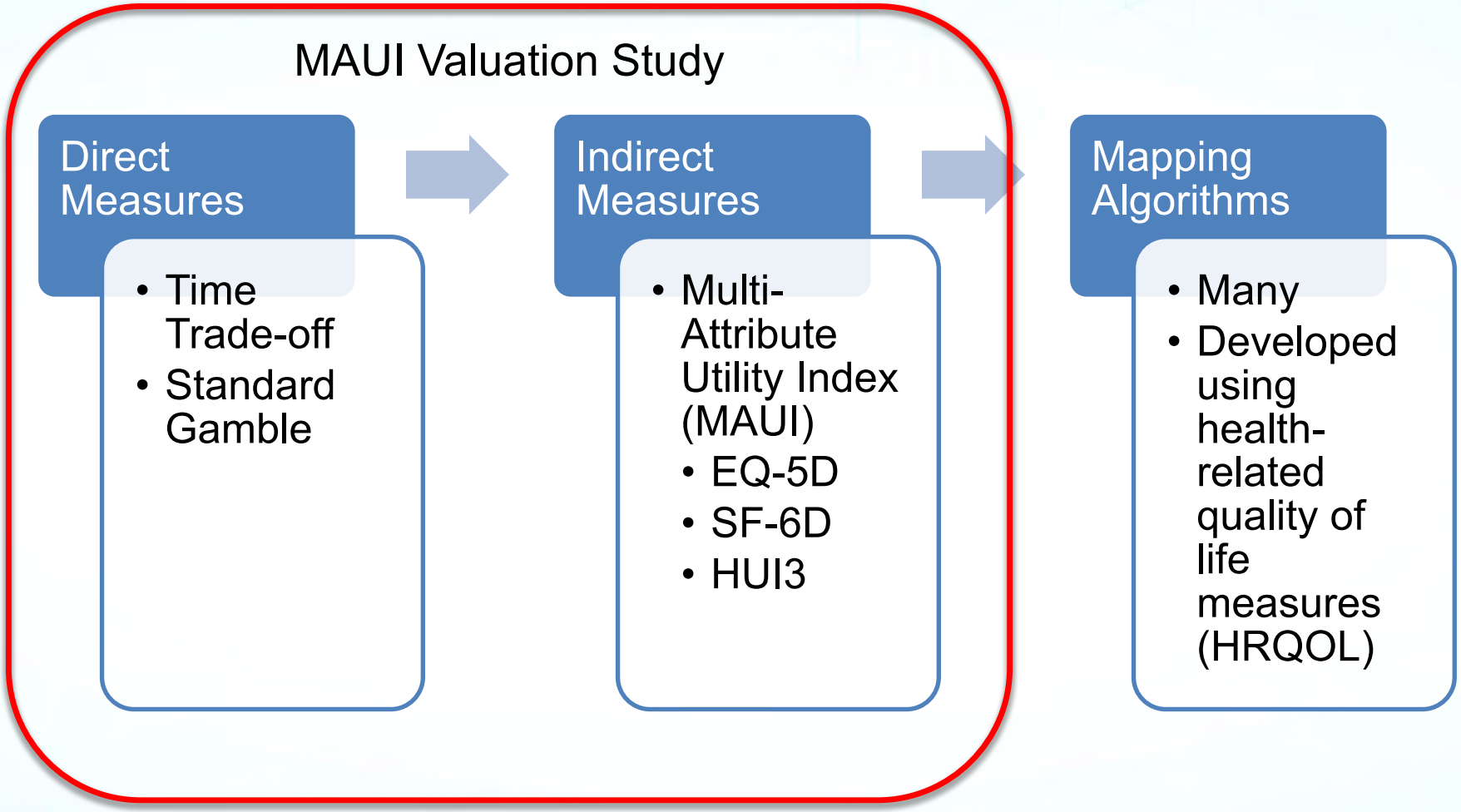
This example identifies the state 11232.

***NB: Missing values** can be coded as '9'.*

***NB: Ambiguous values** (e.g. 2 boxes are ticked for a single dimension) should be treated as missing values.*

Figure 1. EQ-5D example from http://diabetesclinicevaluation.weebly.com/uploads/9/5/6/7/9567609/6029985.jpg?633

Measuring Health Utilities *



Country-specific functional * form of scoring algorithm

US EQ-5D scoring algorithm # (established valuation study)

For subject i valuing state j: '

$E(TTO_{ij}) = \mu_j = 1 - \text{disutility}_j$'

$\text{disutility}_j = X_j\beta$

$X_j = (MO2_j, MO3_j, SC2_j, SC3_j, UA2_j, UA3_j, PD2_j, PD3_j, AD2_j, AD3_j, D1_j, I2_j^2, I3_j, I3_j^2)$,

$MO2_j = 1$ if state j has mobility at level 2, 0 o/w

$MO3_j = 1$ if state j has mobility at level 3, 0 o/w

D1 = # of movements away from full health beyond the first,

I2 = # of dimensions at level 2 beyond the first, '

I3 = # of dimensions at level 3 beyond the first. '

US EQ-5D-3L: Health state (1,2,3,2,1)

Variable	Coeff.	
M2	0.146	
M3	0.558	
S2	0.175	
S3	0.471	
U2	0.140	
U3	0.374	
P2	0.173	
P3	0.537	
A2	0.156	
A3	0.450	
D1	−0.140	x2
I2-squared	0.011	
I3	−0.122	
I3-squared	−0.015	

Utility(1,2,3,2,1)=1-(0.175+0.374+0.173-2*0.140+0.011)
=0.547

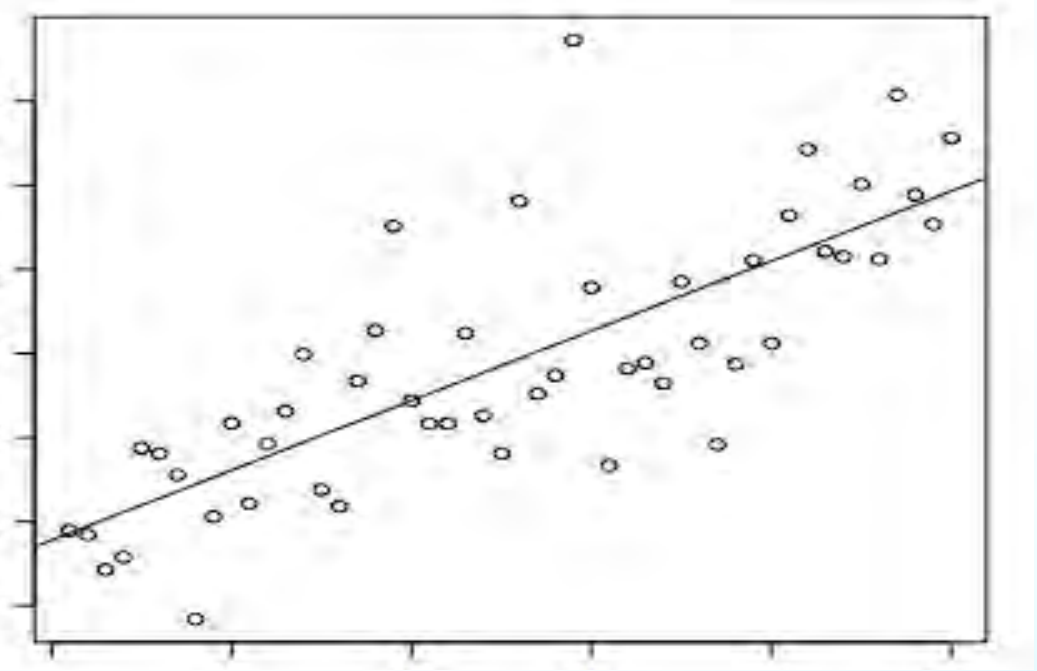
What's the issue? '

- Scoring algorithms yield predictions of population mean utility '
- Predictions are subject to uncertainty
- MAUIs elicit utilities subject to uncertainty
- What impact on uncertainty of estimates of
  - Population mean utility
  - Incremental mean utility?

Predictions .

- Regression modelling
- Uncertainty in line itself
- Points do not lie perfectly on the line .

$$E(TTO_{ij}) = \mu_j = 1 - X_j\beta$$



EQ-5D-3L US valuation study

- 3773 respondents
- Each valued 10 health states using TTO
- 42 health states valued in total
- Modelled mean utility for each health state as a function of health state attributes
- Predicted mean utilities for all 243 health states
- Observed MSE vs theoretical MSE assuming no model mis-fit (
- Bayesian analysis to yield ***predictive distribution*** for each mean utility

Quantifying prediction precision *

- Compute out-of-sample prediction errors
- Omit health state j from analysis and compute expected value of observed minus predicted mean

$$Y_{ij} = 1 - X_j\beta + b_i + \varepsilon_{ij} \quad \text{with } b_i \sim N(0, \sigma_b^2),\ \varepsilon_{ij} \sim N(0, \sigma_e^2),$$

$$E(\bar{Y}_j - 1 + X_j\hat{\beta}_{(j)})^2$$

$$= E(\bar{Y}_j - \mu_j)^2 + E(X_j\beta - X_j\hat{\beta}_{(j)})^2 + 2\,\mathrm{cov}(X_j\hat{\beta}_{(j)}, \bar{Y}_j)$$

$$= \mathrm{var}(\bar{Y}_j) + X_j\,\mathrm{var}(\hat{\beta}_{(j)})X_j' + 2\,\mathrm{cov}(X_j\hat{\beta}_{(j)}, \bar{Y}_j)$$

We see larger MSEs than we should

$$E(\bar{Y}_j - 1 + X_j\hat{\beta}_{(j)})^2$$

$$= E(\bar{Y}_j - \mu_j)^2 + E(X_j\beta - X_j\hat{\beta}_{(j)})^2 + 2\,\mathrm{cov}(X_j\hat{\beta}_{(j)}, \bar{Y}_j)$$

$$= \mathrm{var}(\bar{Y}_j) + X_j\,\mathrm{var}(\hat{\beta}_{(j)})X_j' + 2\,\mathrm{cov}(X_j\hat{\beta}_{(j)}, \bar{Y}_j)$$

Source	Contribution	Cumulative sum
Sampling variance in observed means	0.00018	0.00018
Uncertainty in estimated regression coefficients	0.00011	0.00029
Covariance between observed & fitted mean	-0.0000006	0.00029
Observed MSE on cross-validation		0.00178
Uncertainty due to model mis-specification	0.00178-0.00029	0.00149

Regression Model with model mis-fit

$$\mu_j = 1 - X_j\beta + \delta_j, \quad \delta_j \sim N(0, \sigma_d^2)$$

$$Y_{ij} = \mu_j + b_i + \varepsilon_{ij} \quad \text{with } b_i \sim N(0, \sigma_b^2),\ \varepsilon_{ij} \sim N(0, \sigma_e^2),$$

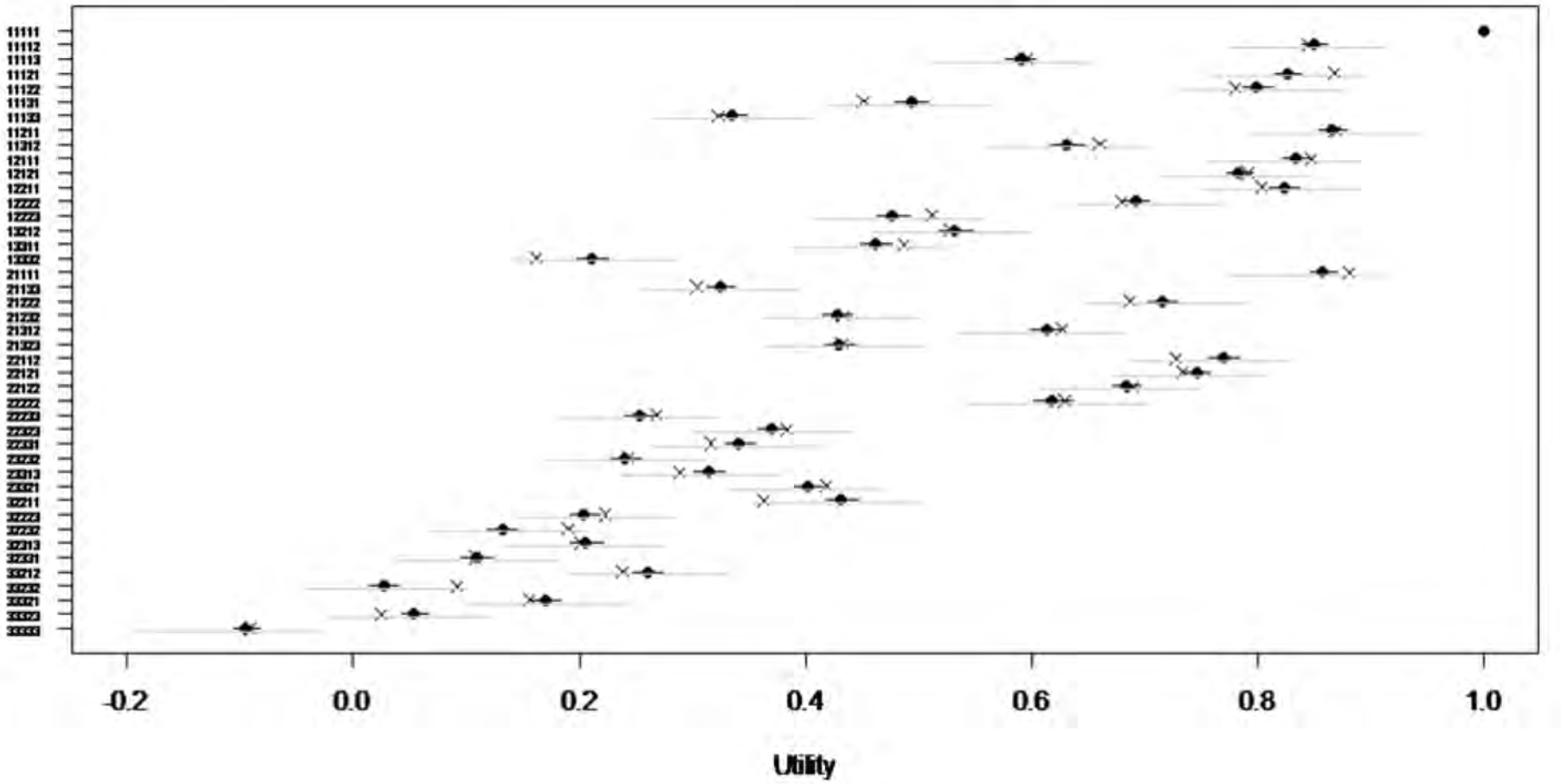
Can think of δ as

- model mis-fit term with a Gaussian prior
- or can conceptualise as a random effect

Vague priors for δ, β, σ

Get predictive distribution of μ_j

How much uncertainty?



Black dots – predicted means
x's – observed means

Black lines – 95% CI ignoring model mis-fit
Grey lines – 95% CI accounting for model mis-fit

How much uncertainty? $

Health states not included in valuation study

- Mean 95% CI width: 0.152
- Range in 95% CI width: 0.142-0.169

Minimum important difference for the EQ-5D-3L $

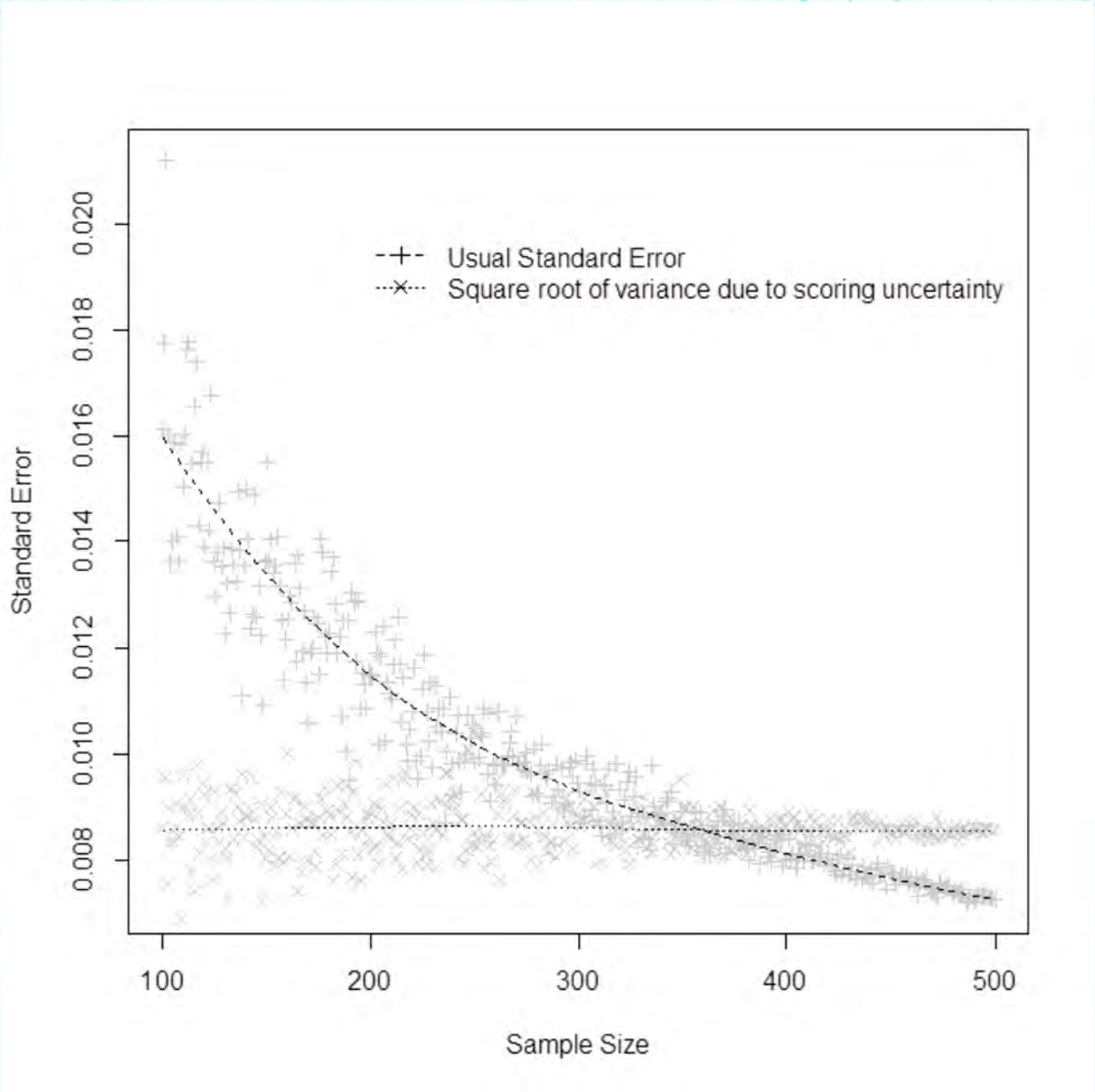
- 0.03 to 0.08

Does it matter? %

- Level of uncertainty may be important
- MAUIs used in HTA to calculate
  - Population mean utility
  - Incremental QALYs (difference in QALYs between groups)
- Use simulation to estimate impact of uncertainty in the scoring system on uncertainty in (incremental) mean utilities.

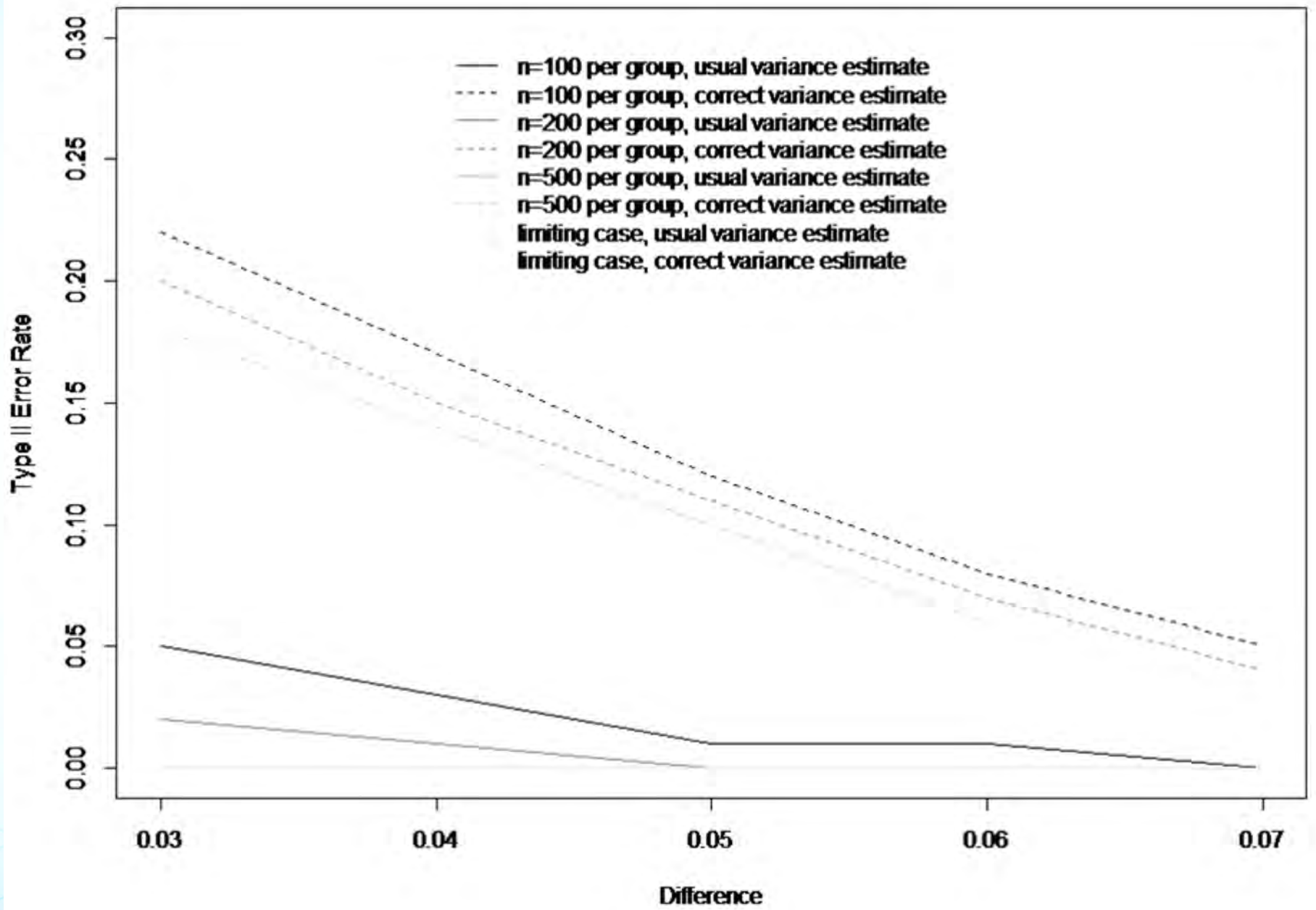
Population mean utility *

- Each respondent fills out the EQ-5D-3L
- Scoring algorithm -> utility for each respondent
- Target of inference: population mean utility
- Sources of uncertainty
  - Sampling variation in health states
  - Uncertainty in scoring algorithm
- Sample of 500 adults from the US general population *
- Take random subsamples of varying sizes

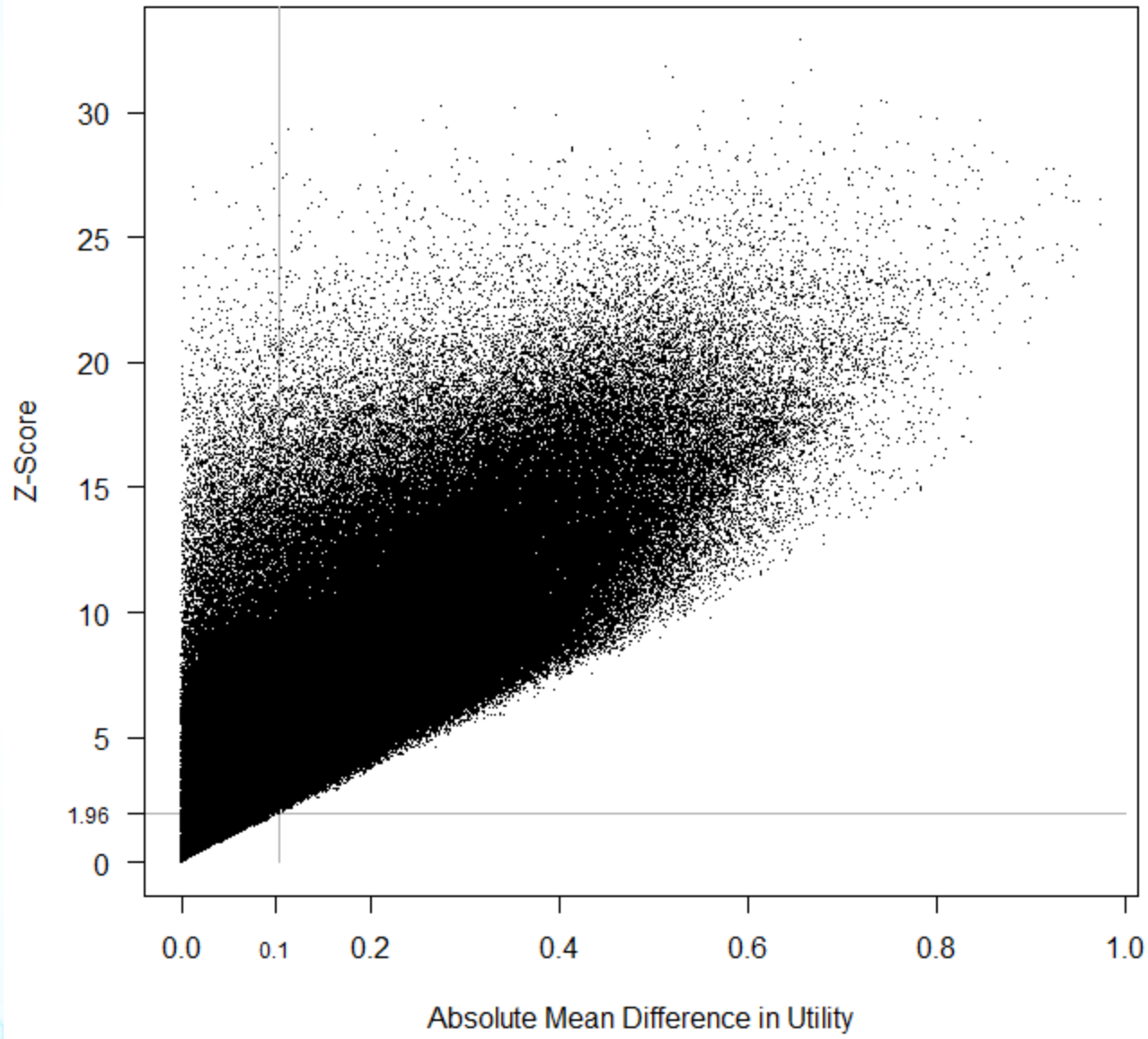

Usual Standard Error
Square root of variance due to scoring uncertainty
Standard Error
0.008
0.010
0.012
0.014
0.016
0.018
0.020
100
200
300
400
500
Sample Size

Incremental mean utility

- Simulate data from an RCT
  - Simulate health state distributions
  - Simulate health states for each person in each arm of the trial
  - Scoring algorithm -> utilities
  - Mean utility per arm
  - Difference -> incremental mean utility


n=100 per group, usual variance estimate
n=100 per group, correct variance estimate
n=200 per group, usual variance estimate
n=200 per group, correct variance estimate
n=500 per group, usual variance estimate
n=500 per group, correct variance estimate
limiting case, usual variance estimate
limiting case, correct variance estimate
Type II Error Rate
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.03
0.04
0.05
0.06
0.07
Difference

For an arbitrarily large study, accounting for uncertainty in the scoring algorithm $



Findings so far… '

- Uncertainty in scoring algorithm is substantial
- Should be accounted for
  - Bayesian methods
  - Multiple imputation (Dr Kelvin Chan)
- Problem not unique to the EQ-5D
  - E.g. SF-6D utilities estimated subject to std error of 0.06
- Current practice gives decision makers a false level of certainty. '

What is the problem?

- Health utilities in the value sets of MAUI are subject to uncertainty
  - E.g. 95% CI prediction error
    - ± 0.0754 for EQ-5D
    - ± 0.1655 for SF-6D[1]

- Minimal clinically important differences
  - 0.05 – 0.08 for EQ-5D[2]
  - 0.01 – 0.09 for SF-6D[3]

[1]Kharroubi S, O'Hagan A, Brazier JE. Estimating utilities from individual health preference data: A nonparametric bayesian method. Applied Statistics. 2005;54(5):879-895.
[2]Le QA, Doctor JN, Zoellner LA, Feeny NC. Minimal clinically important differences for the EQ-5D and QWB-SA in post-traumatic stress disorder (PTSD): Results from a doubly randomized preference trial (DRPT). Health Qual Life Outcomes. 2013;11:59-7525-11-59.
[3]Walters SJ, Brazier JE. Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. Qual Life Res. 2005;14(6):1523-1532.

What is the problem?

$$\mu_j = 1 - X_j\beta + \delta_j$$

Traditionally, utilities from MAUI were treated as known with certainty (rather than estimated with uncertainty)

Ignore δ_j $\quad\quad$ $\hat{\mu}_j = 1 - X_j\beta$

[4]Pullenayegum EM, Chan KKW, Feng X. EQ-5D health utilities are estimated subject to considerable uncertainty. MDM 2015.

What is the problem?

- No method to account for variance of the estimated predicted mean health state utilities from MAUI
  - E.g. subjects with the same health states will always “map” to the same health utility without variation

- Cost-utility analyses based on MAUI do not capture parameter uncertainty in quality-adjusted life years

Solutions

1. Full Bayesian analysis
- Using posterior predictive distributions of health states using original study data[1]

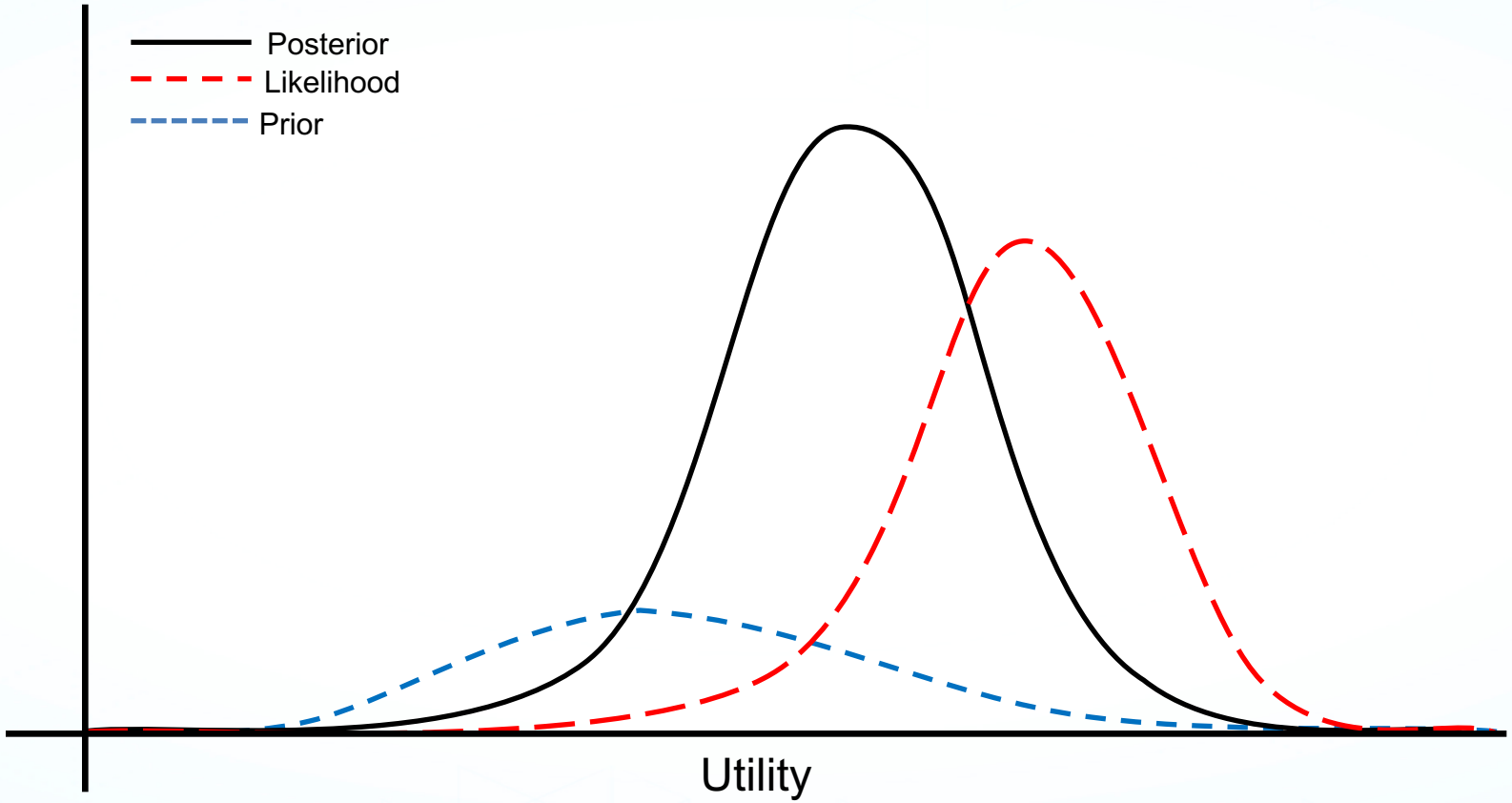
Need to specify a likelihood for the data, and acknowledge the possibility of model misspecification:

$$\mu_j = 1 - X_j\beta + \delta_j$$

If no misspecification $\delta_j = 0$ for all j

[1]Pullenayegum EM, Chan KKW, Feng X. EQ-5D health utilities are estimated subject to considerable uncertainty. MDM 2015.

Bayesian Analysis



Bayesian Analysis

- Challenges:
  - Requires implementation by the original authors
  - Lack of raw data from valuation studies

[1]Pullenayegum EM, Chan KKW, Feng X. EQ-5D health utilities are estimated subject to considerable uncertainty. MDM 2015.

Multiple Imputation

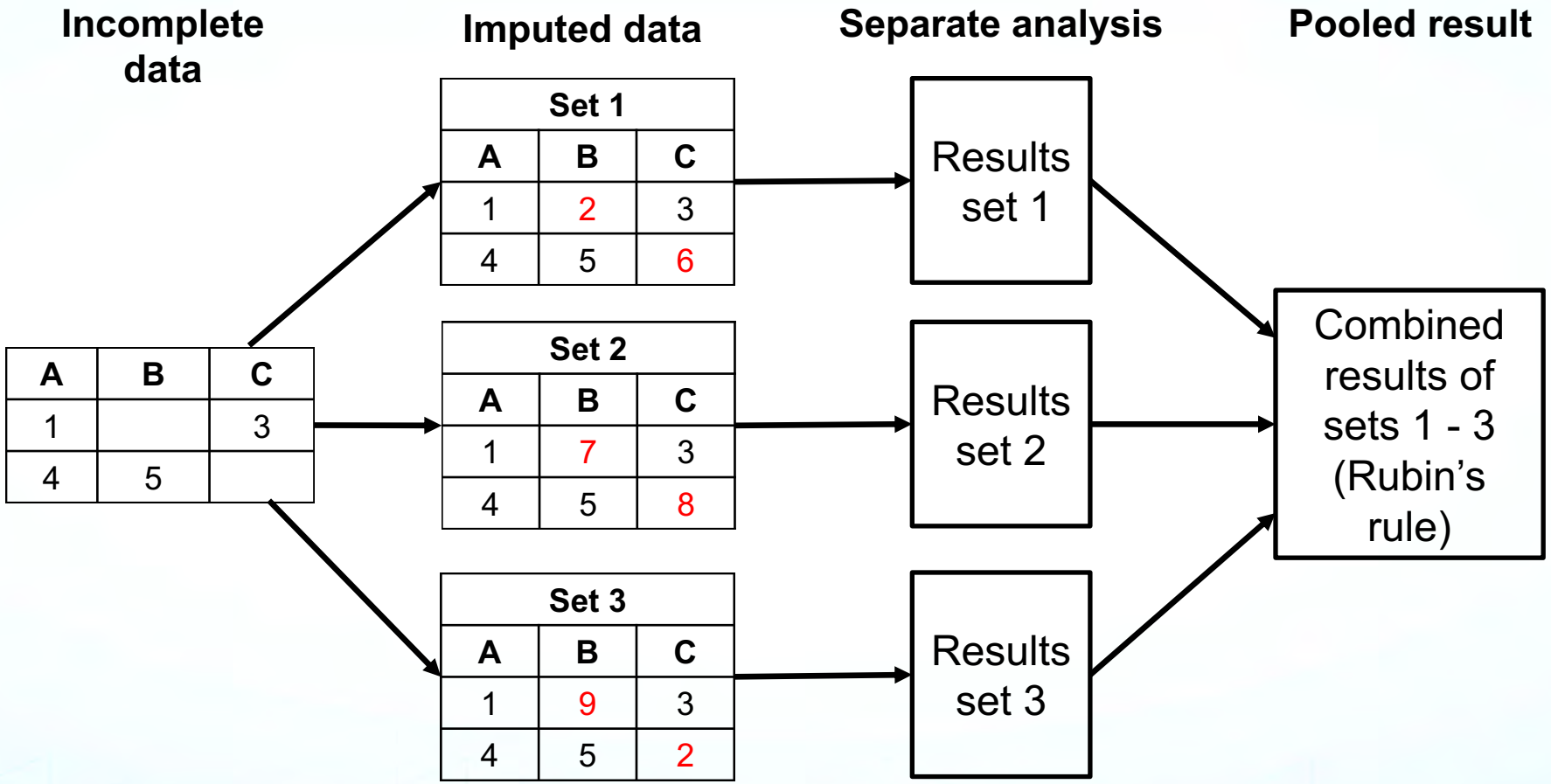
- Three phases:

1. Missing data are filled in *m* times to generate *m* complete data sets

2. The *m* complete data sets are analyzed by using standard procedures

3. The results from the *m* complete data sets are combined for the inference

Multiple Imputation



Solutions

Multiple imputation

- Approximation to Bayesian treatment of parameter uncertainty, used to handle missing data[2,3]
  - E.g. true mean utilities of each health state

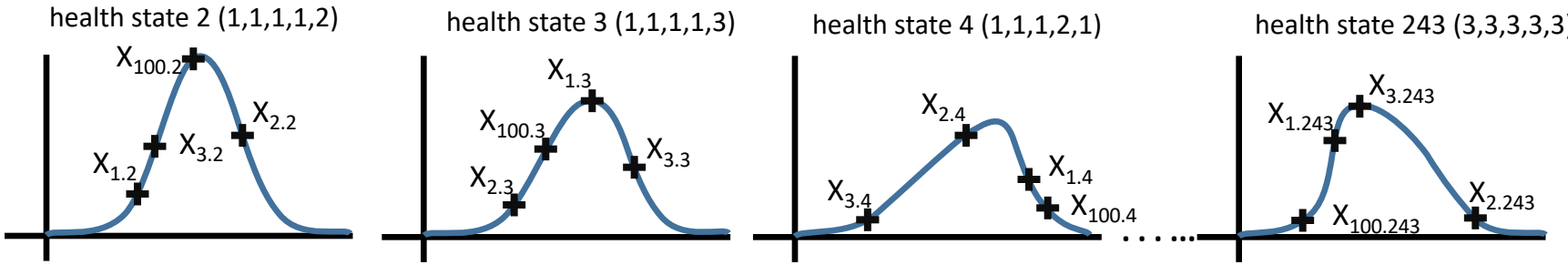
- Replaces missing value with a set of plausible values that represent the uncertainty about the right value to impute

[2]Rubin DB. Multiple imputation for non-response in surveys. John Wiley & Sons; 1987.
[3]Schafer JL. Multiple imputation: A primer. Stat Methods Med Res. 1999;8(1):3-15.

EQ-5D-3L	Health States (HS)
1,1,1,1,1	1
1,1,1,1,2	2
1,1,1,1,3	3
↓	↓
3,3,3,3,1	241
3,3,3,3,2	242
3,3,3,3,3	243

1st IMPUTATION

Posterior distributions for 243 health states



100 points randomly selected from each health state to generate 100 imputed datasets

Imputed dataset 1

Health State	Utility
2	$X_{1.2}$
3	$X_{1.3}$
4	$X_{1.4}$
↓	↓
243	$X_{1.243}$

Imputed dataset 2

Health State	Utility
2	$X_{2.2}$
3	$X_{2.3}$
4	$X_{2.4}$
↓	↓
243	$X_{2.243}$

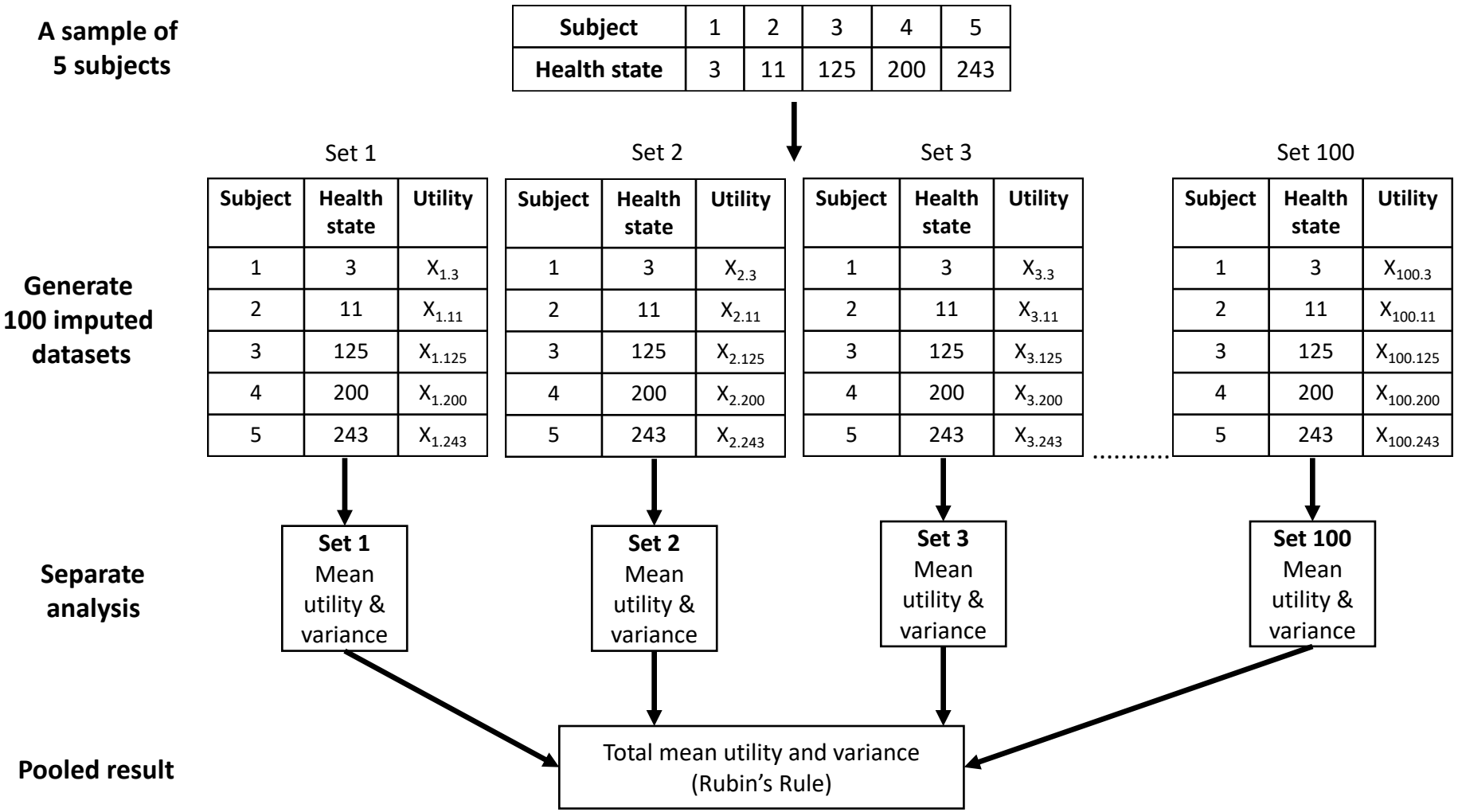
Imputed dataset 3

Health State	Utility
2	$X_{3.2}$
3	$X_{3.3}$
4	$X_{3.4}$
↓	↓
243	$X_{3.243}$

. . . .

Imputed dataset 100

Health State	Utility
2	$X_{100.2}$
3	$X_{100.3}$
4	$X_{100.4}$
↓	↓
243	$X_{100.243}$


A sample of 5 subjects
Subject 1 2 3 4 5
Health state 3 11 125 200 243
Generate 100 imputed datasets
Set 1
Set 2
Set 3
Set 100
Subject Health state Utility
1 3 X1.3
2 11 X1.11
3 125 X1.125
4 200 X1.200
5 243 X1.243
Subject Health state Utility
1 3 X2.3
2 11 X2.11
3 125 X2.125
4 200 X2.200
5 243 X2.243
Subject Health state Utility
1 3 X3.3
2 11 X3.11
3 125 X3.125
4 200 X3.200
5 243 X3.243
.....
Subject Health state Utility
1 3 X100.3
2 11 X100.11
3 125 X100.125
4 200 X100.200
5 243 X100.243
Separate analysis
Set 1 Mean utility & variance
Set 2 Mean utility & variance
Set 3 Mean utility & variance
Set 100 Mean utility & variance
Pooled result
Total mean utility and variance (Rubin's Rule)

Validation (Aim)

- Using the US EQ-5D-3L valuation study,
- Demonstrate that multiple imputation can correct underestimation of variance of mean health utilities

Methods: Full Bayesian Analysis

- Derivation set: N = 3,773 (US EQ-5D-3L)
- Application set: N =3,958 (CWF dataset)
- Derivation set used D1 model, which was fitted to Bayesian mixed effect model
  - Obtained posterior predictive distribution of the mean utility attached to each health state
  - Applied to application set to compute mean and variance

Approach

1. Fit a full Bayesian model

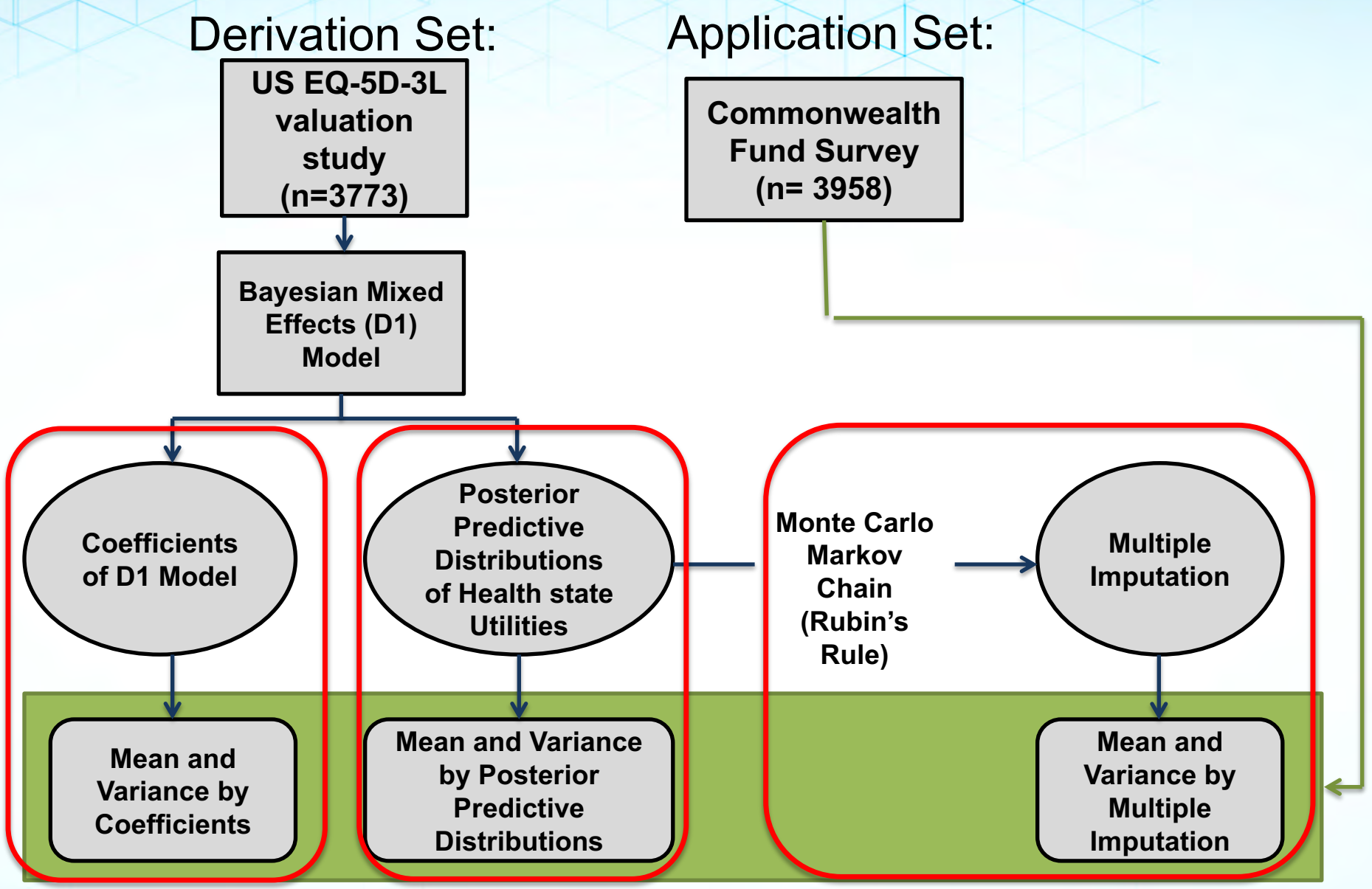
$$\mu_j = 1 - X_j\beta + \delta_j$$

μ_j : health utility of the *j*th health state
$X_j\beta$: linear predictor of the MAUI of the *j*th health state
δ_j : deviation of predicted health utility from the true utility

2. Use the joint predictive distribution to implement multiple imputation

- Simulated US EQ-5D data to examine the 95% CI coverage of health utilities from multiple imputation

	Using Covariance Matrix of US D1 Regression Model	Random sampling of joint predictive distributions (MI)
Number of Health States that have coverage >95%	2 out of 42	38 out of 42
Percentage of coverage	40%	98%



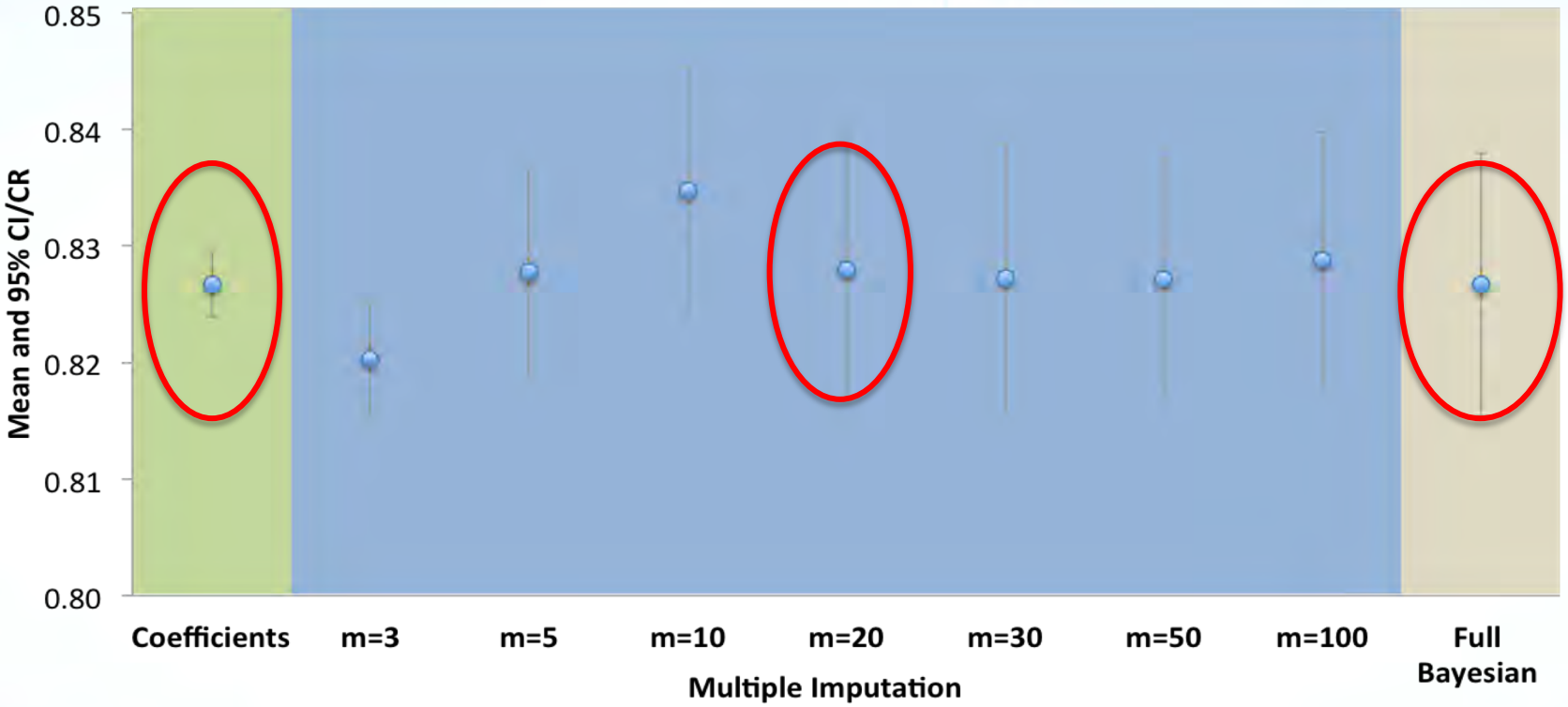
Methods to illustrate that multiple imputation can be used to correct for the underestimation of variance of mean health utilities of the sample.

Results

Comparisons of the sample mean and sample standard error of the mean health utility of the application set (N =3958) based on (i) the regression coefficients (i.e. scoring algorithm), (ii) the full Bayesian model's posterior predictive distribution and multiple imputation

	Traditional Method (based on coefficients)	Multiple Imputation	Full Bayesian Model
Mean	0.827	0.828	0.827
Variance	7.96×10^{-6}	1.28×10^{-4}	1.27×10^{-4}
SE	2.82×10^{-3}	1.13×10^{-2}	1.12×10^{-2}

Results



95% of confidence intervals (CI)/credible regions (CR) of sample mean utility of the application set

Discussion

- Multiple imputation provides “middle ground”
  - Researchers do not have to learn Bayesian methods
  - Variance and standard error reflect appropriate degree of parameter uncertainty
  - Applicable to a wide variety of analyses (e.g. regressions) where traditional MI is applicable.

Limitations

- Need original publishers of MAUI studies to create the imputed datasets to make it publicly available to apply this imputation method

Conclusions

MI is a potential method to account for the underestimation of variance of predicted health utilities

Current work

- Improving precision based on existing data
  - Use posterior distribution of δ
  - Model correlation among δ
- Improving precision based on better designs
  - How many health states to value?
  - Quantify MSE as a function health state selection & SS

Acknowledgement

- Eleanor Pullenayegum
- Feng Xie
- Andy Willan
- Wendy Lou

Extra Slides

Multiple Imputation

MCMC method:

The imputation I-step: draw values of $Y_{i(mis)}$ from a conditional distribution of $Y_{i(mis)}$ given $Yi_{(obs)}$

The posterior P-step: simulates the posterior population mean vector and covariance matrix from the complete sample estimates

Multiple Imputation

- Three phases:

1. Missing data are filled in *m* times to generate *m* complete data sets

2. The *m* complete data sets are analyzed by using standard procedures

3. The results from the *m* complete data sets are combined for the inference

Multiple Imputation

With *m* imputations, you can compute *m* different sets of the point and variance estimates for a parameter Q. Let $\hat{Q}_i$ and $\hat{U}_i$ be the point and variance estimates from the *i*th imputed data set, *i*=1, 2, ..., *m*. Then the point estimate for Q from multiple imputations is the average of the *m* complete-data estimates:

$$\overline{Q} = \frac{1}{m}\sum_{i=1}^{m}\hat{Q}_i$$

Let $\overline{U}$ be the within-imputation variance, which is the average of the *m* complete-data estimates

$$\overline{U} = \frac{1}{m}\sum_{i=1}^{m}\hat{U}_i$$

And B be the between-imputation variance

$$B = \frac{1}{m-1}\sum_{i=1}^{m}(\hat{Q}_i - \overline{Q})^2$$

Multiple Imputation

The total variance is:

$$T = \overline{U} + (1 + \frac{1}{m})B$$

Multiple imputation

- Multiple imputation with Monte Carlo Markov chain models
  - Performed using derivation set
  - Randomly drawn multiple imputed sets applied to application dataset

- Mean, variance and standard error across imputed sets calculated using Rubin's rule[10]

[10]Rubin DB. Multiple imputation for non-response in surveys. John Wiley & Sons; 1987.

Bayesian mixed effect model

- Take posterior predictive joint distributions of the mean utility to
  - Capture parameter uncertainty
  - Perform multiple imputations (imputed sets drawn randomly from the Gibbs sampler)