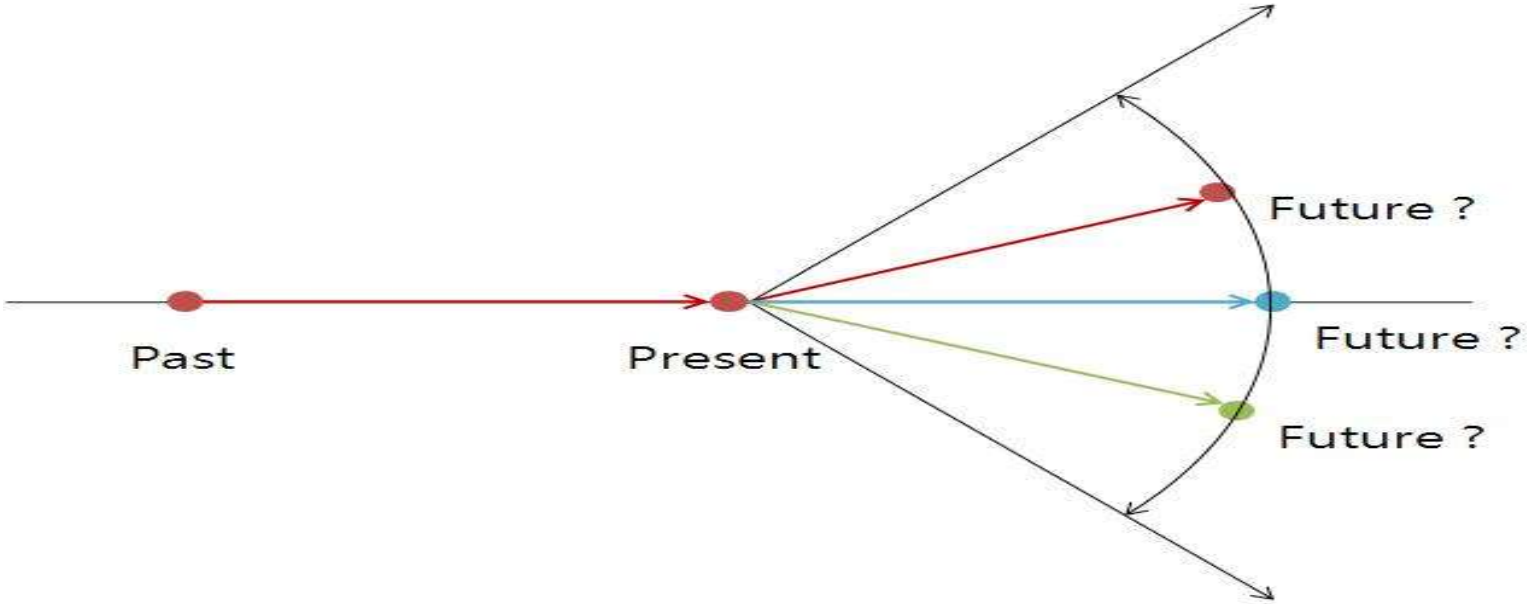
MONASH University

VICBiostat

School of Public Health and Preventive Medicine

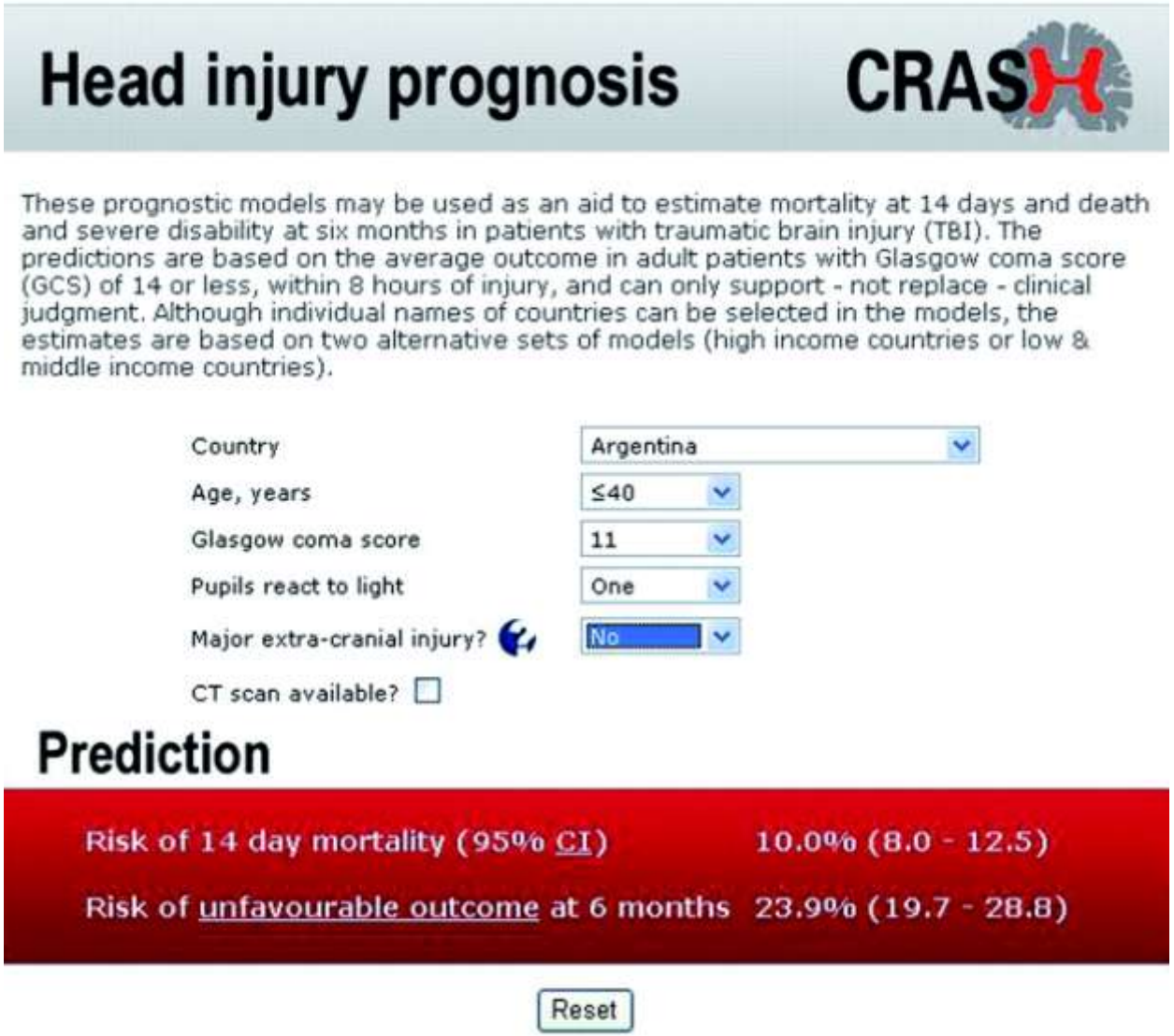
Missing data and risk prediction models

Rory Wolfe
Roman Ahmed, Masoumeh Sanagou
28th June 2012



Talk structure

- Overview of risk prediction modelling
- Motivating example – pneumonia after cardiac surgery
- Methodological work
  - Simulation study 1; missing data arises according to “missing at random” mechanism
  - Simulation study 2; “missing as normal”


Head injury prognosis
CRASH
These prognostic models may be used as an aid to estimate mortality at 14 days and death and severe disability at six months in patients with traumatic brain injury (TBI). The predictions are based on the average outcome in adult patients with Glasgow coma score (GCS) of 14 or less, within 8 hours of injury, and can only support - not replace - clinical judgment. Although individual names of countries can be selected in the models, the estimates are based on two alternative sets of models (high income countries or low & middle income countries).
Country
Argentina
Age, years
≤40
Glasgow coma score
11
Pupils react to light
One
Major extra-cranial injury?
No
CT scan available?
Prediction
Risk of 14 day mortality (95% CI) 10.0% (8.0 - 12.5)
Risk of unfavourable outcome at 6 months 23.9% (19.7 - 28.8)
Reset

Purpose of disease risk prediction

1 Aid to clinical treatment decision-making

2 Education tool for clinicians and patients

3 Adjustment for case mix in performance audits

4 Identification of patients for research studies

5 Usually intended to complement judgment of health care professionals

Risk prediction applications

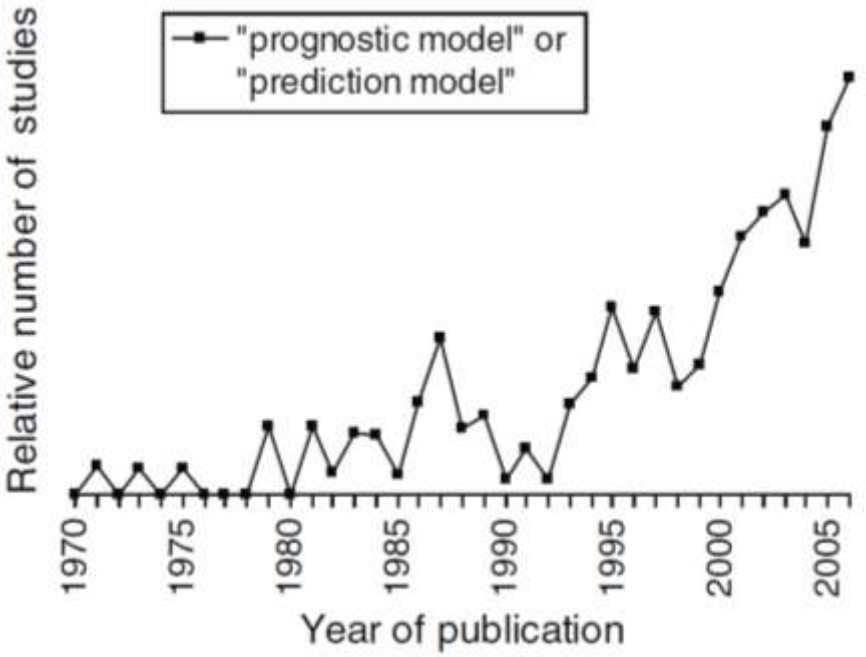
- NHF & CSANZ 2005 guideline on cost-effective lipid management:
  - Lipid-modifying therapy indicated for primary prevention in patients with absolute risk 15%+ of CVD event in next 5 years according to 1991 Framingham equation (e.g. NZ CVD absolute risk calculator)
- Breast Cancer Risk prediction calculator:
  - enables women to assess their risk of breast cancer from age, family history of breast cancer, use of the contraceptive pill, hormone replacement therapy etc. http://canceraustralia.nbocc.org.au/risk/yourrisk.html
- Daily Telegraph (UK):
  - "The Dr Foster Hospital Guide, published today, reveals the death rates at all NHS trusts in England for 2010-11 and shows if they are higher or lower than expected."
  - Expected, i.e. predicted mortality from prediction equations such as EuroSCORE, POSSUM, STS etc.
- Transcatheter aortic valve implantation trial; among the inclusion criteria is "Subject must have STS mortality risk score ≥3% and ≤8%" http://riskcalc.sts.org/STSWebRiskCalc273/de.aspx

The methodology of risk prediction

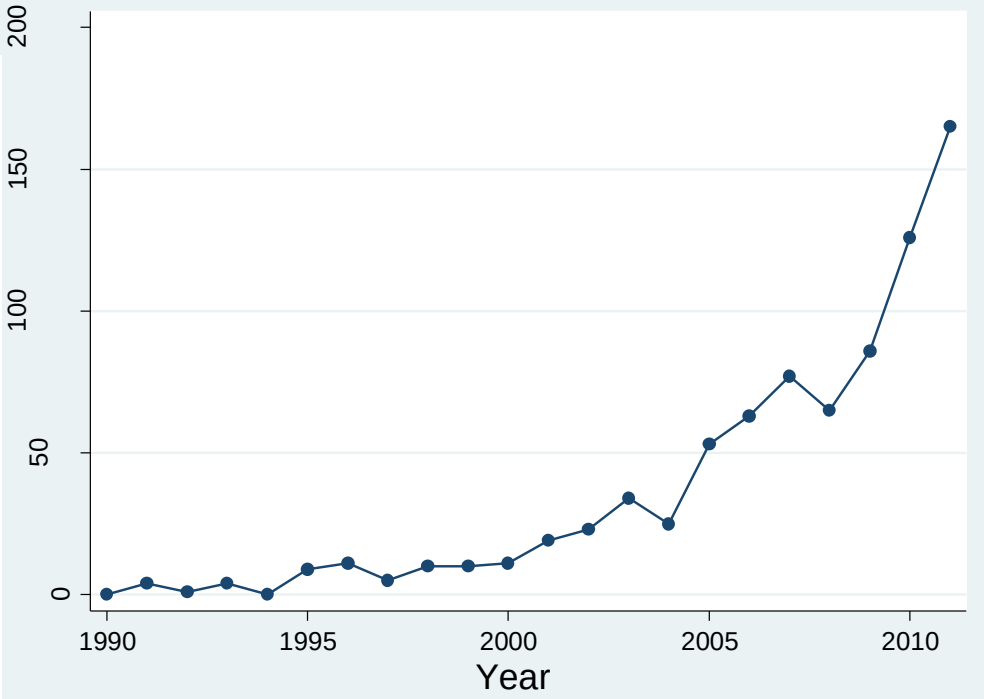
- Develop a new model
  - Define patients of interest
  - Identify data set of risk factors and outcome of interest
    - possible problem of missing data
  - Apply logistic regression, Cox PH reg, neural network, etc
  - Assess internal validity / performance
    - optimism-corrected by bootstrap resampling of modelling procedure
    - N-fold cross-validation
    - Split-sample approach
- Perform external validation – model performance in a new context
  - Identify new data set
    - possible problem of missing data and missing risk factors
  - Apply new model and quantify performance
  - Possibly update model

Steyerberg “Clinical Prediction Models”

Fig. 1.1 Studies in PubMed with the terms “prognostic model” or “prediction model” in the title, as a fraction of total number of studies in PubMed (676,000 in 2005)



Ahmed 2012: (“validation” AND ("prognostic model" OR "prediction model" OR “risk prediction model”)

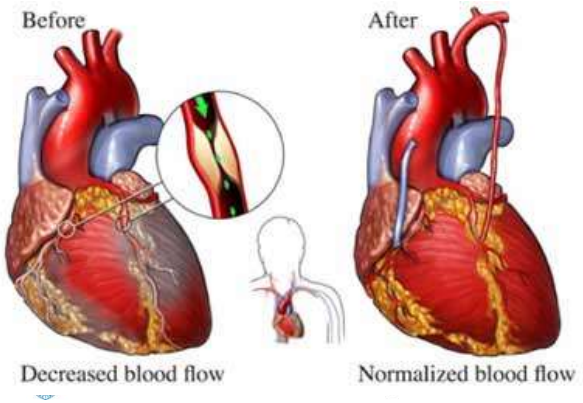


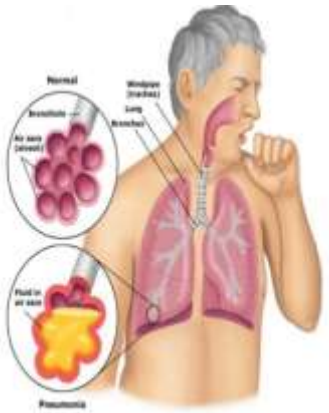
Cautionary tale – the QRISK story

- July 2007 - BMJ published a new CVD risk prediction tool, QRISK, for primary care in UK
  - electronic database representative of primary care; 10 million patients over a 17 year period from 529 general practices
  - Large missing data problem, e.g. ~70% missing TC/HDL levels
- Modelling method employed multiple imputation
- Developed tool omitted TC/HDL
- July/August 2007 - Letters to editor
  - “effect for TC/HDL completely inconsistent with previous studies”
  - Use of MI flawed due to omission of disease outcome from the imputation model
- 2008 – BMJ publish QRISK2 which is used for online calculator

Case study: External validation of a model for pneumonia after CABG surgery

- Pneumonia is a leading cause of increased LOS in ICU and increased mortality rates
- Leads to elevated health care costs among patients who have undergone cardiac surgery
- North American study used 17,143 CABG patients to develop a new model to predict nosocomial pneumonia after surgery (Kinlin 2010, *CID*)
- External validation of this model is lacking





North American study

- used logistic regression analysis (backwards-elimination algorithm)
- to identify 13 predictive factors (out of 35 considered) for risk of pneumonia after CABG surgery

Dealing with missing values: single imputation

When risk factor information was missing, it was assumed that the patient was in the low-risk category

Internal validation: split-sample

Patients were randomly assigned to either the derivation (n=8,572) or validation (n=8571) subsets.

Objective of our analysis

To undertake the first external validation study of this model using the multi-centre Australian and New Zealand Society of Cardiac and Thoracic Surgeons (ANZSCTS) registry, n>23,000.

Similarities & differences

- Patients undergoing isolated CABG surgery or valve+CABG procedures.
- Similar demographics: >70% male mean age ~67-68y

Outcome definition varied → ANZSCTS 5%, Kinlin 2%

Yet mortality ANZSCTS 2% v Kinlin 4%

Four risk factors unavailable in ANZSCTS

Missing values assumed to be in low risk category (also performed complete case analysis)

Model risk factor	Point score	Kinlin %	ANZSCTS %
Mechanical ventilation>1 day	11	21	10
Underweight	7	2	1
PTCA during hospitalization	5	4	1
Prior CABG with IM artery graft	5	3	NA
Smoking history	4	33	66
Cancer history	4	9	NA
Vancomycin administration pre-op	4	7	NA
Creatinine >1.2 mg/dl	3	25	27
Admitted from non-residential setting	3	23	NA
COPD	3	15	3
Emergent surgery	3	8	4
CCS class ≥3	2	42	47
Intra-op blood transfusion	2	39	41

Results

		AUC (95% CI)	H-L P-value	Comment
Model development	Kinlin et al	0.78 (0.75,0.82)	0.7	Good discrimination, well calibrated
Internal validation	Kinlin et al	0.75 (0.71,0.78)	0.2	
External validation	Missing as low risk	0.69 (0.68,0.71)	<0.001	Acceptable discrimination? Unsatisfactory calibration?
External validation	Complete case	0.69 (0.68,0.71)	<0.001	
Updated model		0.71	0.5	

Prediction of nosocomial pneumonia; Published and Updated prediction tools

Characteristic	Published coefficient	Updated coefficient	Point value
Mechanical ventilation>1 day	1.75	1.47	11
Underweight	1.06	0.11	7
PTCA during hospitalization	0.77	0.05	5
Smoking history	0.58	0.44	4
COPD	0.54	0.06	3
Creatinine >1.2 mg/dl	0.46	0.09	3
Emergent surgery	0.44	0.18	3
Intra-op blood transfusion	0.36	0.69	2
CCS class ≥3	0.32	0.15	2
Prior CABG with internal mammary artery graft	0.82	Not available	5
Vancomycin administration pre-op	0.64	Not available	4
Cancer history	0.59	Not available	4
Admitted from non-residential setting	0.41	Not available	3

Conclusion

- First external validation of a prediction model derived from American-Canadian patients

- Demonstrated discrimination ability of borderline acceptability and poor calibration for the prediction model
  - performance improved by model updating

- Reinforces the importance of external validation of predictive models.
  - Highlights the challenge of missing risk factor values and missing risk factors

- We will use the updated model as a patient-level adjustment when exploring hospital-level risk factors for pneumonia

Missing data in risk prediction:

Multiple imputation versus complete case?

Is missing as normal ever a sensible strategy?

Justifying methodological approaches

- Missing data on risk factors sometimes assumed to be
  - Missing at random (MAR)
    - By default, according to variables in model (complete case analysis)
    - According to a specific imputation model
  - “missing as normal” (MAN)
- Question 1: If true missing data mechanism is MAR then which approach to data analysis is preferable?
  - complete case (CC), multiple imputation (MI) or MAN?
  - Evidence exists to support use of MI in this context but
    - Evidence base is incomplete
    - Unclear the degree of advantage
    - Little study of whether an advantage exists when validating models
- Question 2: What if missing data mechanism is close to MAN?

Existing evidence: CC v. MAN v. MI

- Main problem with CC is reduced precision in estimation of model coefficients
  - Possible bias if MAR requires variable other than those in model
  - Downward bias in discrimination?
- Single imputation (e.g. replace with “low risk” / “normal”, MAN approach)
  - Almost certain bias in model coefficients if MAR missingness
  - Over-estimates precision of model coefficients
  - Downward bias in discrimination?

Multiple imputation

- Use observed relationships among variables included in an imputation model to generate a plausible value to replace each missing value
  - Repeat m times, i.e. m completed datasets created
  - Analyse each dataset →m results; combine using Rubin's rules
- Possibly biased estimates if true missingness process doesn't match MAR of imputation model
- Appropriate precision in model estimates
- Congeniality of imputation and analysis model is essential

Motivation for simulation design

- Reflect four examples
  - prediction of incident diabetes in older Americans (Kanaya2005) and in Thailand (Aekplakorn2006)
  - two clinical patient management examples; prediction of fistula maturation failure in preparation for hemodialysis (Lok2006) and prediction of 6-month mortality among older patients with heart failure (Huynh2008).
- Examples share the following characteristics
  - small sample size for model development (n of 1549, 2667, 422, 282)…
  - …and validation (n between 445 and 2503)
  - small number of risk factors in developed model (number of OR = 6, 9, 4, 4)
  - mix of demographic and clinical or laboratory measures among the risk factors often with prevalence close to 50%
  - reasonable discrimination (AUC = 0.73, 0.75, 0.76, 0.80 in development data).
- For simplicity we narrow our focus to missing data in one predictor
  - E.g. fasting glucose concentration in a model for incident diabetes that also contains demographic (age, sex) and clinical (BMI) measures.
- Restrict investigation of internal validity to split sample approach

Simulation study design – outcome

- 2000 datasets with n=1500 simulated using a logistic relationship between outcome and four binary risk factors
  - X1, X2 have 20% prevalence
  - X3, X4 have 50% prevalence
  - weak pairwise correlations among X1-X4
  - logit (Y=1|X1-X4) = -5.3 + 1.6X1 + 0.8X2 + 1.6X3 + 0.8X4
- Outcome prevalence = 5%
  - expect 75 events in n=1500; 19 events per model coefficient
- AUC=0.80
- Repeat entire simulation with continuous X3

Simulation study design – introducing missing data

- 0, 5, 15, 30 and 50% of values of risk factor X1 set to missing
  - Under MCAR assumption (missingness not dependent on any observed or unobserved data)
  - under the MAR assumption X1|X3:
    - P(X1 missing | X3=0) = 3p
    - P(X1 missing | X3=1) = p
- To deal with missing data when developing and validating models:
  - MAN, CC, MI

Simulation study design – validation scenarios

- In each of the 2000 simulations
- New datasets n=750 generated for validation

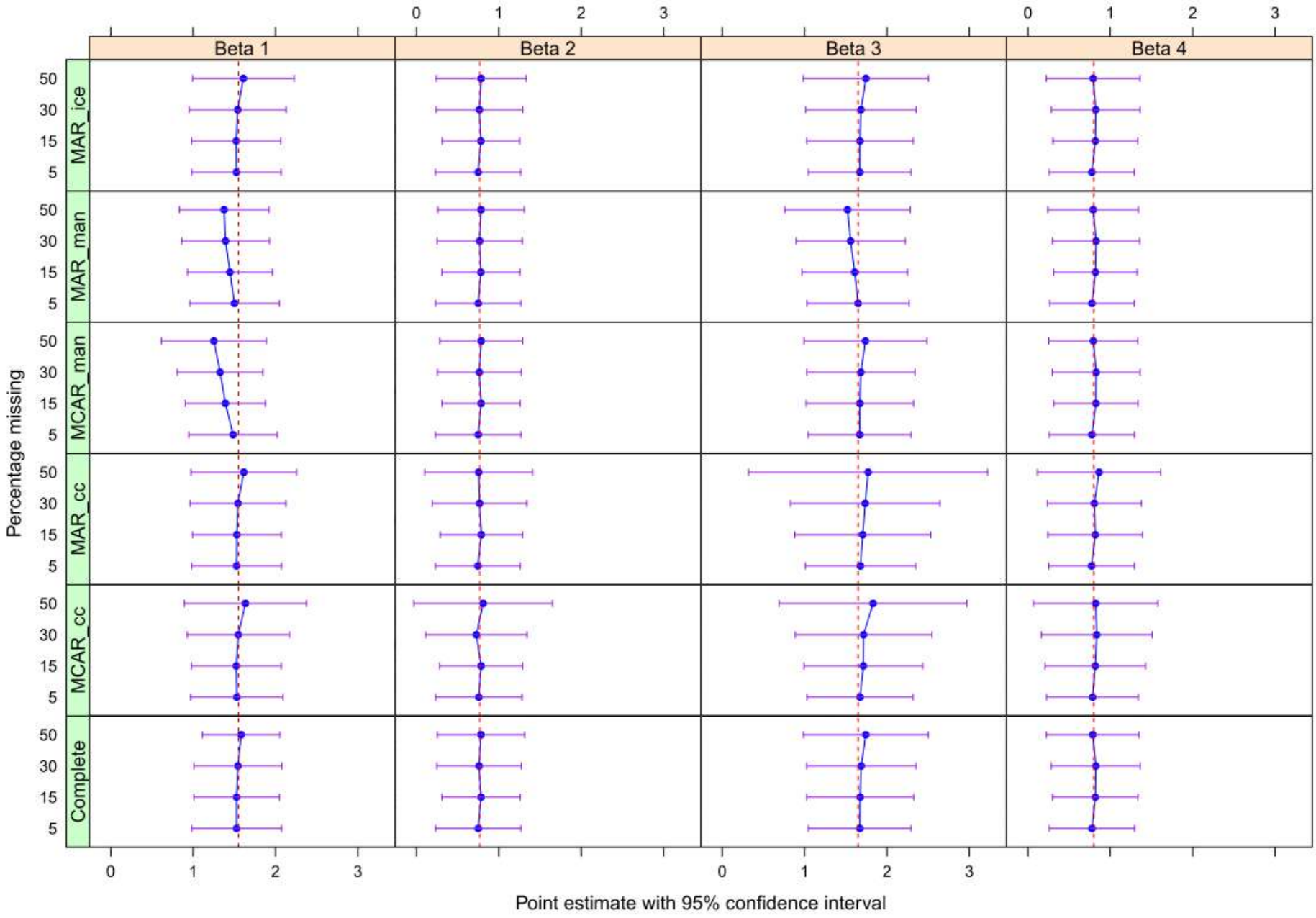
- Five different validation scenarios
  - split sample validation,
  - external validation varying outcome prevalence
  - external validation varying discrimination.

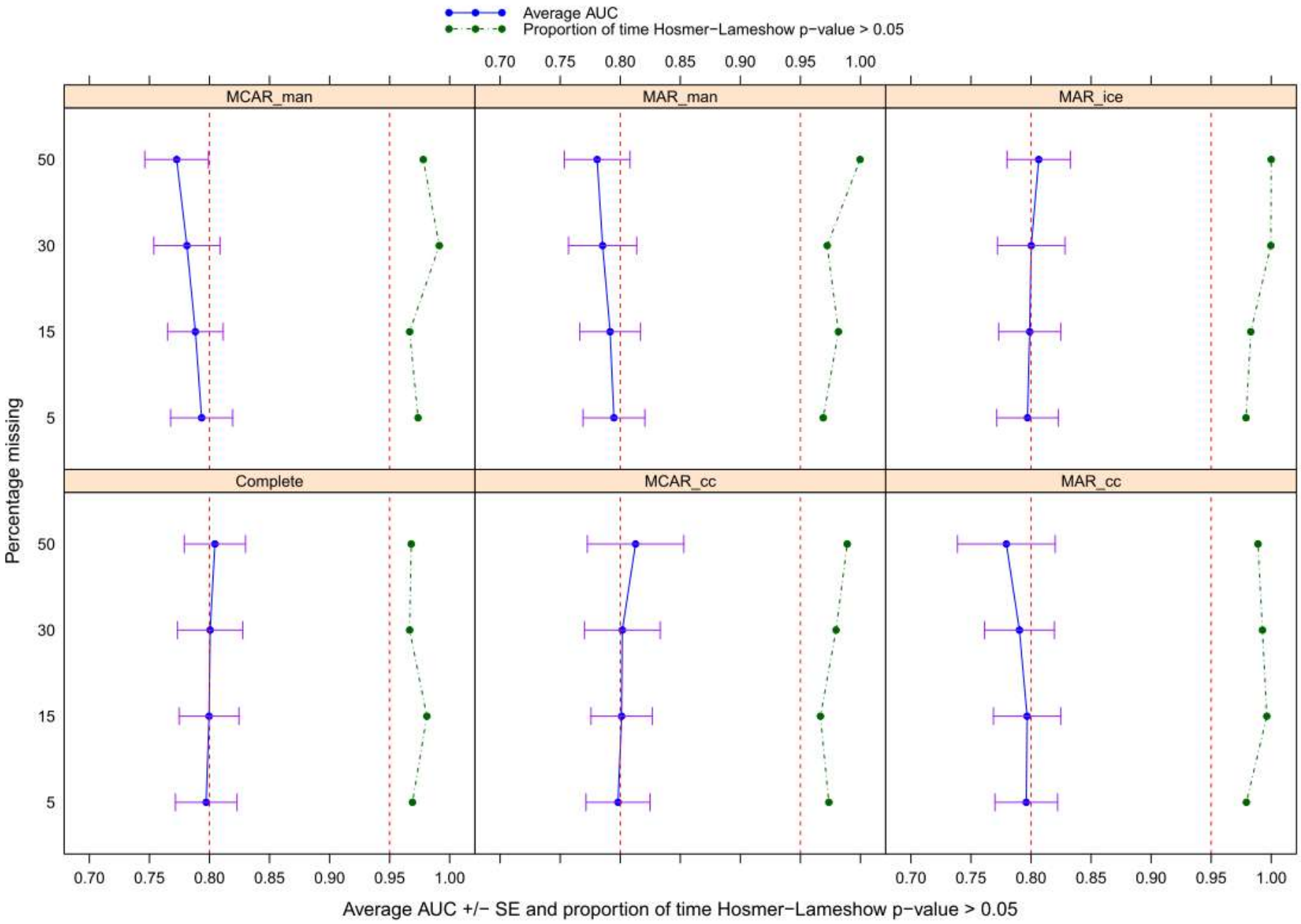
Five validation scenarios

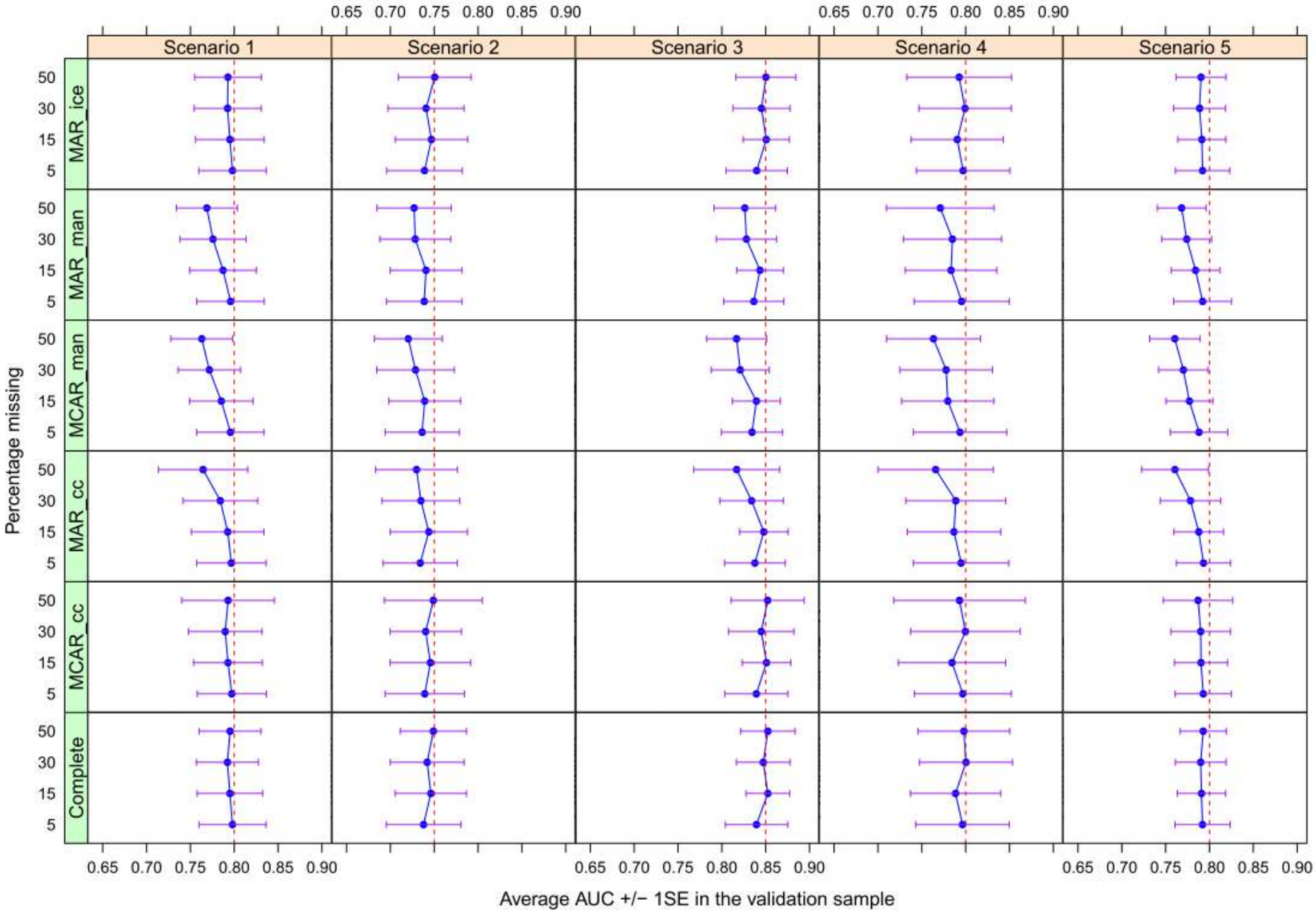
Scenario	Prevalence of outcome	AUC	Type	Comment
1	5%	0.80	Internal	Split-sample
2	5%	0.75	External	Worse true discrimination ability of X1-X4
3	5%	0.85	External	Better true discrimination
4	2.5%	0.80	External	Lower prevalence
5	10%	0.80	External	Higher prevalence

Summary across 2000 simulations

- Statistics of interest included:
  - bias and variability in estimated model parameters
  - Coverage
  - discrimination as assessed using area under the ROC curve, AUC
  - calibration
    - as assessed using Hosmer-Lemeshow goodness of fit tests in the development dataset
    - calibration-in-the-large, calibration intercept and slope in validation datasets


Beta 1
Beta 2
Beta 3
Beta 4
MAR_ice
MAR_man
MCAR_man
MAR_cc
MCAR_cc
Complete
50
30
15
5
0
1
2
3
Percentage missing
Point estimate with 95% confidence interval


Average AUC
Proportion of time Hosmer−Lameshow p−value > 0.05
MCAR_man
MAR_man
MAR_ice
Complete
MCAR_cc
MAR_cc
Percentage missing
50
30
15
5
0.70
0.75
0.80
0.85
0.90
0.95
1.00
Average AUC +/− SE and proportion of time Hosmer−Lameshow p−value > 0.05


Scenario 1
Scenario 2
Scenario 3
Scenario 4
Scenario 5
MAR_ice
MAR_man
MCAR_man
MAR_cc
MCAR_cc
Complete
Percentage missing
50
30
15
5
0.65 0.70 0.75 0.80 0.85 0.90
Average AUC +/− 1SE in the validation sample

Conclusions from simulation study 1

- Suggests that MI is preferred to CC and MAN for both model development and validation
  - In relatively small datasets (with about 19 EPV)
  - With missing data in a single predictor
  - Under specific MAR assumption (probablity of missing in X1 is 3 times higher for X3=0 than X3=1)
- When data are MAR, using MAN introduced downward bias in discrimination
  - can be viewed as worst-case scenario regarding model performance?
  - Useful for sensitivity (to missing data) analysis?

Simulation study 2

- If missing data mechanism is MAR then MAN introduces bias and overestimates precision
- But what if missing data mechanism is close to MAN?
- MAN is a non-ignorable missing data assumption (missingness depends on unobserved data)
  - has been shown previously that CC can be preferable to MI when data are missing according to a non-ignorable mechanism

Simulation study design

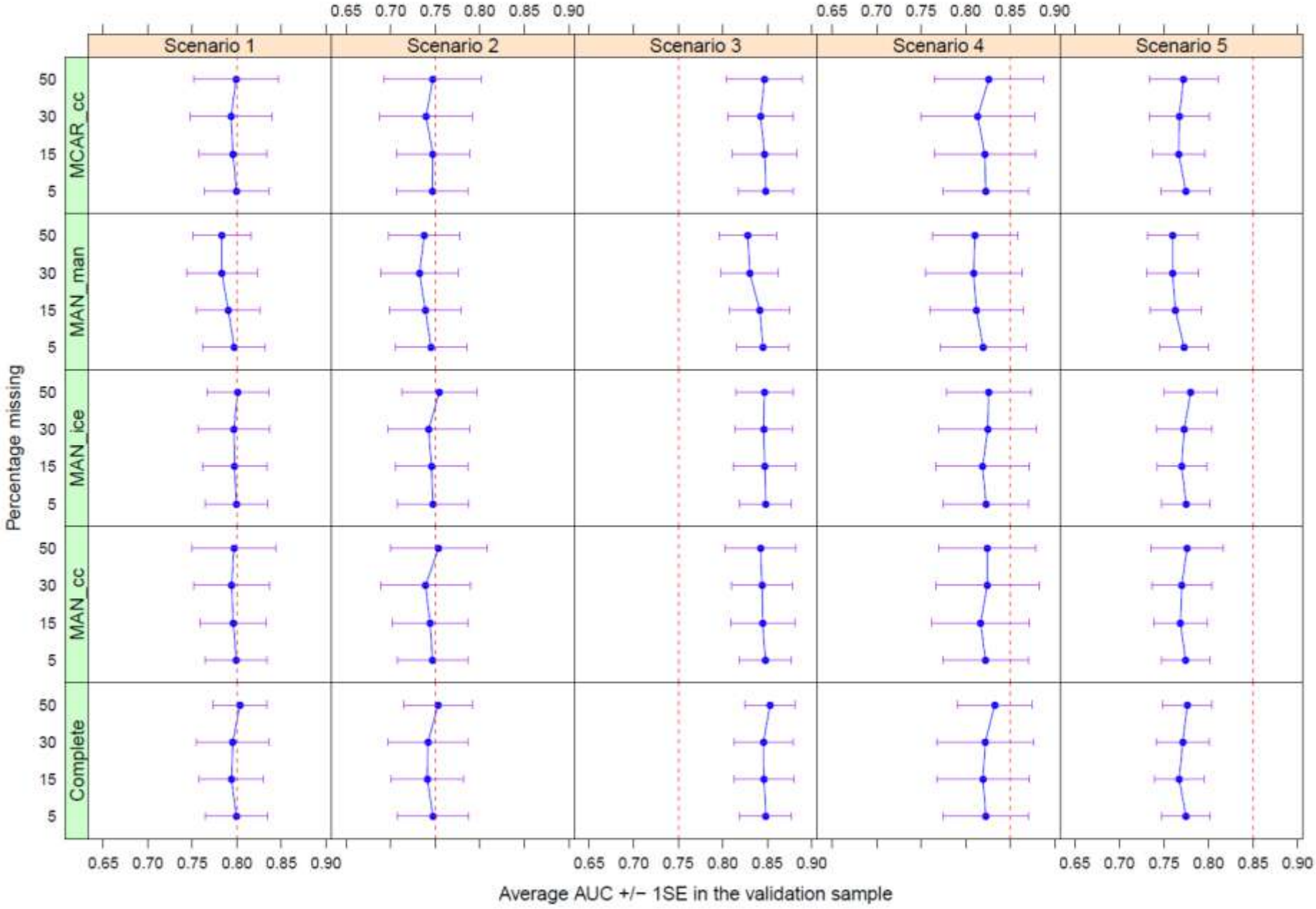
- 2000 datasets with n=1500 simulated using a logistic relationship between outcome and four binary risk factors
  - weak pairwise correlations among X1-X4.
- 5, 15, 30 and 50% of values of risk factor X1 set to missing
  - under MAR assumption
  - under “tending towards” MAN
  - under “nearly” MAN
- To deal with missing data when developing and validating models: used MAN, CC, MI
- Five different validation scenarios were investigated
  - split sample validation,
  - external validation when external dataset has higher (or lower) outcome prevalence than development data
  - external validation when an external dataset has better (or worse) true ability of the four risk factors to discriminate outcome.

Design details

logit (Y=1|X1-X4) = -5.3 + 1.6X1 + 0.8X2 + 1.6X3 + 0.8X4

- X1 20% (abnormal:normal 1:4)
- X2 20%
- X3 50%, X4 50%
- AUC 0.8

- MAN: all missing values of X1 have X1=normal
- Tending towards MAN: X1 abnormal/normal 1:9 split of missing values
- Nearly MAN: 1:19 split of missing values


Scenario 1
Scenario 2
Scenario 3
Scenario 4
Scenario 5
MCAR_cc
MAN_man
MAN_ice
MAN_cc
Complete
Percentage missing
50
30
15
5
0.65 0.70 0.75 0.80 0.85 0.90
Average AUC +/− 1SE in the validation sample

Results

- When true mechanism tended towards MAN…
  - using MAN to deal with the missing data problem led to bias in estimates of the coefficient for X1.
  - MAN under-estimated AUC in development and validation datasets.
  - CC led to bias in the assessment of calibration-in-the-large.
  - MI exhibited least bias across all model performance measures that were considered.

- Under “nearly MAN”, as the % missing in X1 increases, the same pattern of findings are still observed.

Conclusions

- Even when the true missing data mechanism is nearly MAN, using MAN as the method for dealing with missing data in development and validation of risk prediction models is not preferable to using MI.
  - However CC is not necessarily an acceptable alternative to MAN in this scenario.
- Under MAR, MI is preferable to CC and MAN
- These results bring into question the ongoing use of CC and MAN for dealing with missing data problems in the development and validation of risk prediction models.
- However, further comparisons are warranted as MAN is a non-ignorable missing data assumption and it has been shown previously that CC can be preferable to MI when data are missing according to a non-ignorable mechanism.

Acknowledgements

- Roman Ahmed
- Masoumeh Sanagou
- Gareth Ambler
- Andrew Forbes, Paul Roderick, Kevan Polkinghorne, Sara Ivory
- Karin Leder, Chris Reid & ANZSCTS registry
- NHMRC project grant 546059