UNIVERSITY OF LEICESTER

Exploring results from simulation studies interactively

A Gasparini[1] IR White[2] T Morris[2] MJ Crowther[1]

[1] Department of Health Sciences, University of Leicester, Leicester, United Kingdom

[2] MRC Clinical Trials Unit at UCL, London, United Kingdom

Students' Day, 38th Annual Conference of the International Society for Clinical Biostatistics

Key messages and questions

Key messages:

- simulation studies are being increasingly used
- dissemination of results is key
- interactive tools can effectively supplement reporting of simulation studies

Questions:

- What is your experience with presenting results from simulation studies?

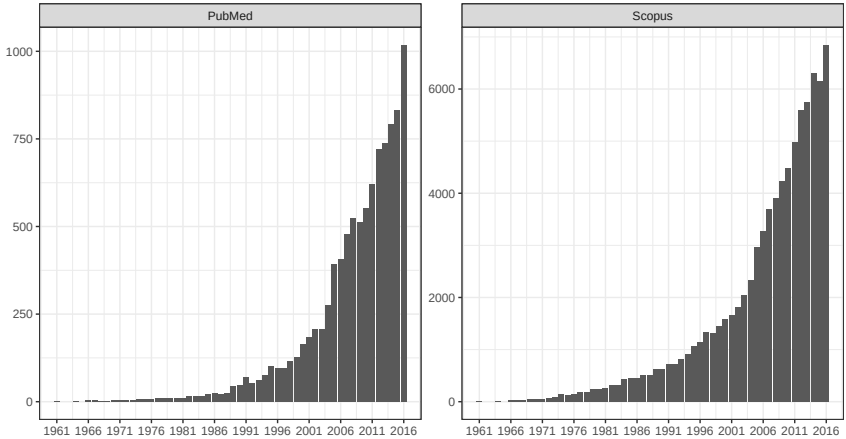
About me

- Currently: first-year PhD student at the University of Leicester, Leicester, United Kingdom
- Previously: BSc in Statistics and Computing Technologies from University of Padua, Italy; MSc in Biostatistics and Experimental Statistics from University of Milano-Bicocca, Italy
- PhD project:
  1. joint modelling of longitudinal and survival data
  2. survival models with random effects
  3. application to health records data and cardiovascular epidemiology

Simulation studies are useful...

- Evaluating new statistical methods
- Evaluate large sample approximations
- Comparing the performance of different methods/models
- Assessing the impact of violating assumptions
- You name it!

...and common!


PubMed
Scopus
1000
750
500
250
0
6000
4000
2000
0
1961 1966 1971 1976 1981 1986 1991 1996 2001 2006 2011 2016
Year
Query: 'simulation study'

Dissemination is key

1. Can drive practitioners and applied statisticians to methods that have been shown to perform well in their practical settings
2. Can guide researchers to develop new methods in a promising direction
3. Can provide insights into less established methods

With great power comes great responsibility

- Increased availability of powerful computational tools surely contributed to the rise in popularity of simulation studies
- Adding multiple data-generating mechanisms or methods to compare is cheaper than ever (computationally speaking)
- Things can get out of control quickly[1]:

	Min	Max
Number of methods evaluated	1	18
Number of estimands	1	32
Number of factors varied across DGMs	1	324
Number of DGMs	1	6×10^{11}

[1] Source: simulation studies published by Statistics in Medicine in 2015, unpublished data from the course *Using simulation studies to evaluate statistical methods* (White IR, Morris T and Crowther MJ)

My experience

- Simulation study on the impact of misspecification in survival models with shared frailties
- Fully factorial design
- A priori factors that may affect the results:
  1. baseline hazard function
  2. sample size (number of clusters, number of individuals per cluster)
  3. variance of the frailty term
  4. distribution of the frailty term
- Massive number of simulation scenarios to summarise: how?

ADMEP framework

A framework for harmonising reporting of simulation studies[2]:

- Aim(s)
- Data-generating mechanism(s)
- Method(s)
- Estimand(s)
- Performance measure(s)

Think of what you want to learn, and how: focusing on these aspect beforehand will make designing and reporting simulation studies easier.

[2] White IR, Morris T and Crowther MJ, unpublished

Enter SiReX

- Interactive tools can supplement the ADMEP framework very effectively
- SiReX [saɪɘ(ɹ)-ɛks], *Simulation Results eXplorer*
- Developed using R and Shiny
- Workflow:
  1. Upload your results
  2. Summary statistics are computed automatically
  3. Select a DGM and summary tables and plots are updated automatically
  4. Export summary statistics, tables, and plots for later use

Demo

https://goo.gl/iGU2Xc

Key messages and questions

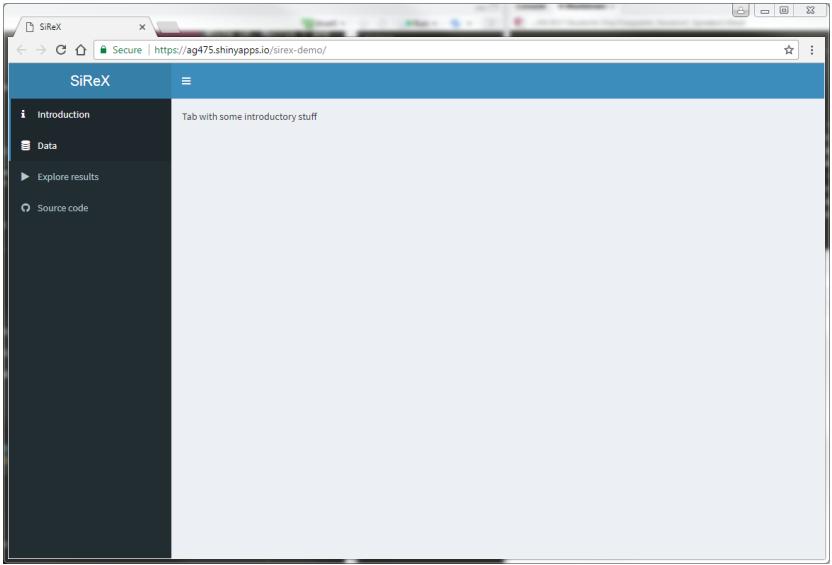
Key messages:

- simulation studies are being increasingly used
- dissemination of results is key
- interactive tools can effectively supplement reporting of simulation studies

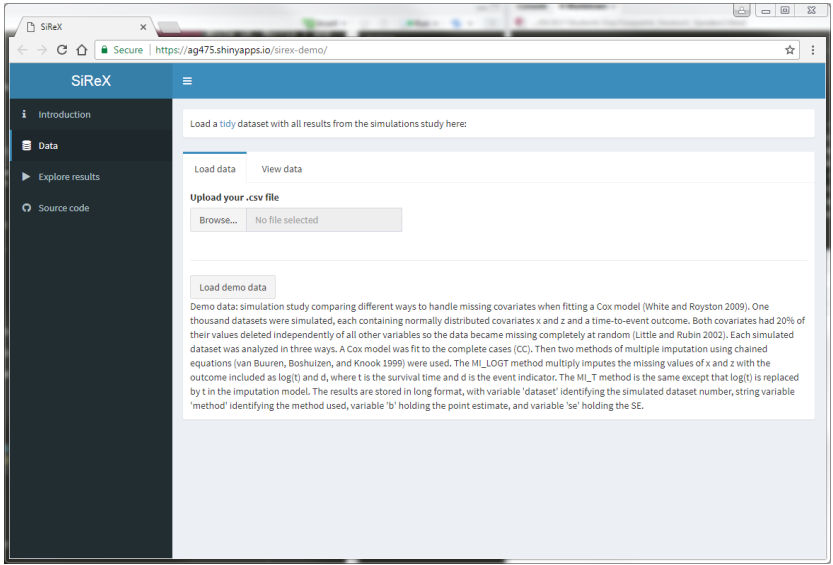
Questions:

- What is your experience with presenting results from simulation studies?

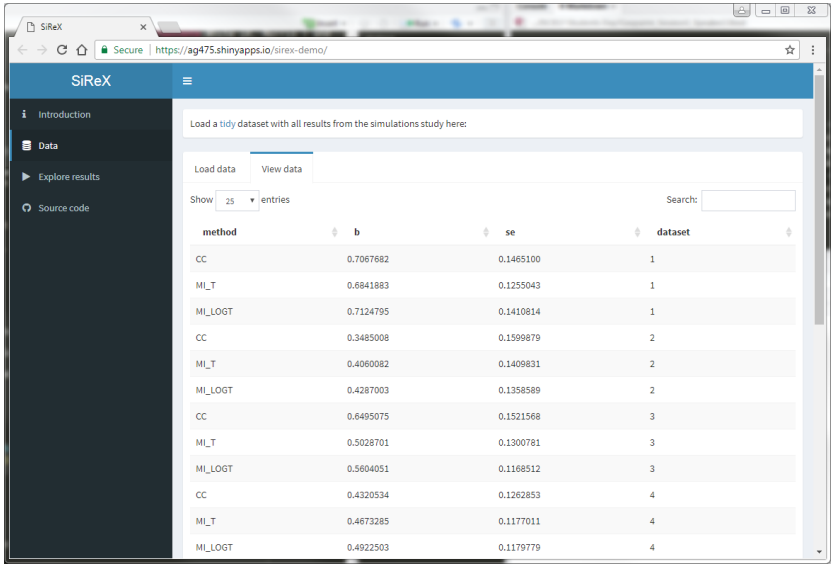
Demo: landing page



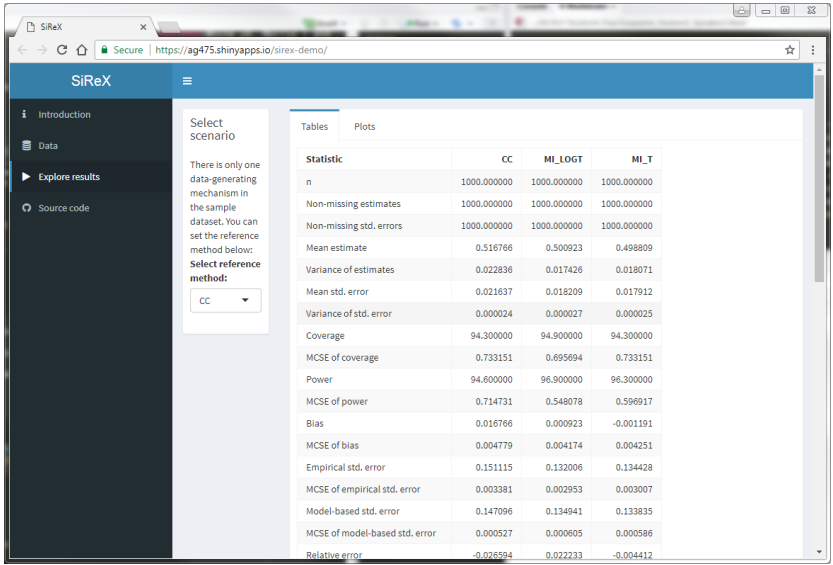
Demo: load data



Demo: landing page

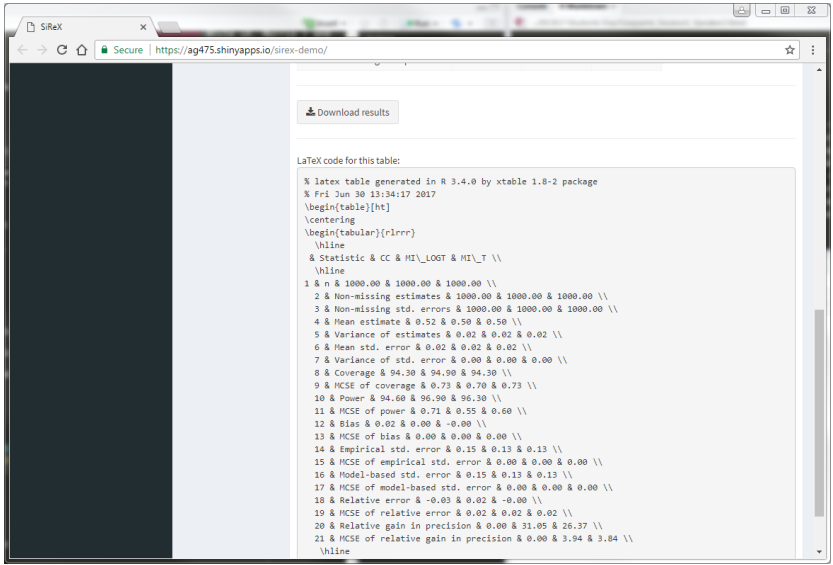


Demo: table of summary results

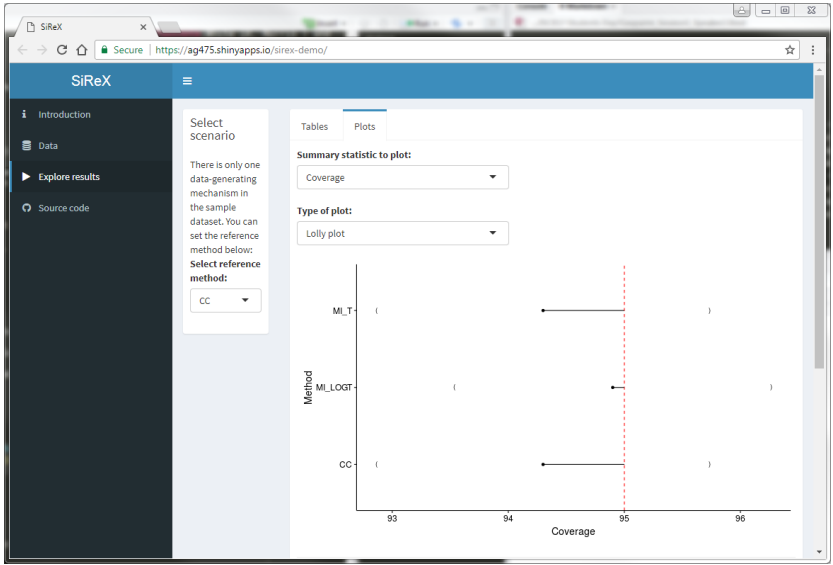


Statistic	CC	MI_LOGT	MI_T
n	1000.000000	1000.000000	1000.000000
Non-missing estimates	1000.000000	1000.000000	1000.000000
Non-missing std. errors	1000.000000	1000.000000	1000.000000
Mean estimate	0.516766	0.500923	0.498809
Variance of estimates	0.022836	0.017426	0.018071
Mean std. error	0.021637	0.018209	0.017912
Variance of std. error	0.000024	0.000027	0.000025
Coverage	94.300000	94.900000	94.300000
MCSE of coverage	0.733151	0.695694	0.733151
Power	94.600000	96.900000	96.300000
MCSE of power	0.714731	0.548078	0.596917
Bias	0.016766	0.000923	-0.001191
MCSE of bias	0.004779	0.004174	0.004251
Empirical std. error	0.151115	0.132006	0.134428
MCSE of empirical std. error	0.003381	0.002953	0.003007
Model-based std. error	0.147096	0.134941	0.133835
MCSE of model-based std. error	0.000527	0.000605	0.000586
Relative error	-0.026594	0.022233	-0.004412

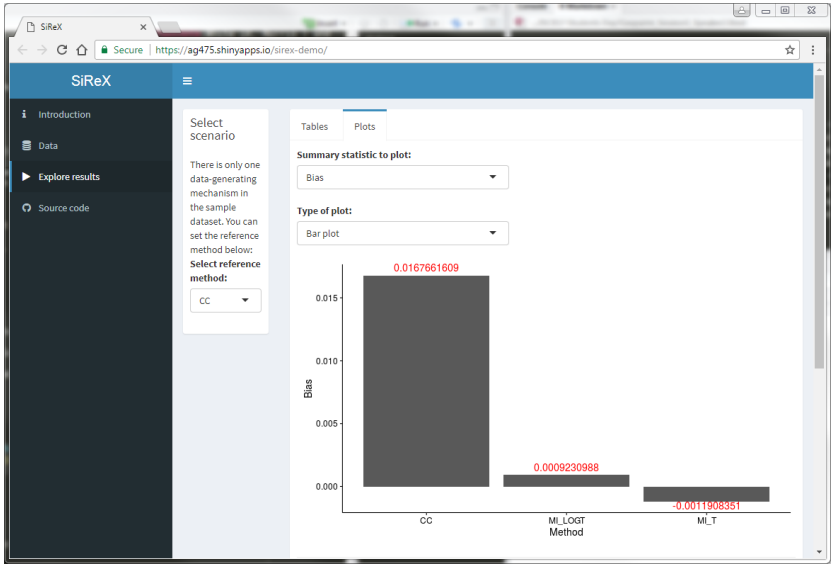
Demo: table of summary results



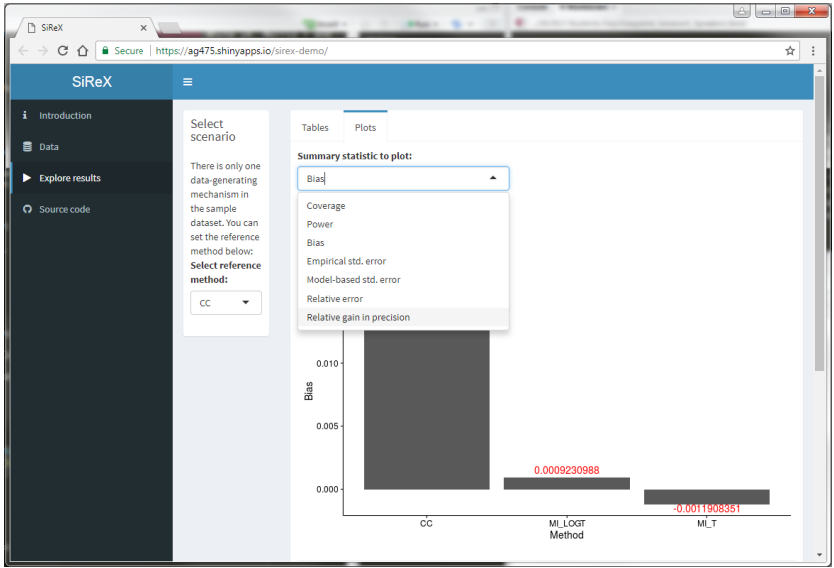
Demo: plot of summary results



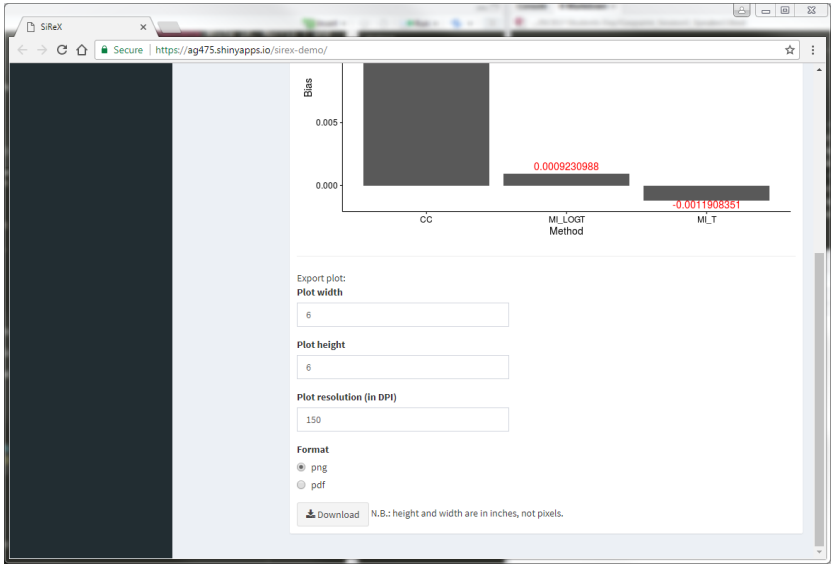
Demo: plot of summary results



Demo: plot of summary results



Demo: plot of summary results



Key messages and questions

Key messages:

- simulation studies are being increasingly used
- dissemination of results is key
- interactive tools can effectively supplement reporting of simulation studies

Questions:

- What is your experience with presenting results from simulation studies?