



UNIVERSITY OF
LEICESTER

Exploring results from simulation studies interactively

A Gasparini¹ IR White² T Morris² MJ Crowther¹

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

²MRC 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?

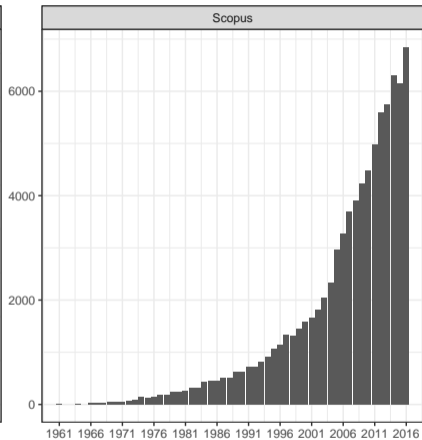
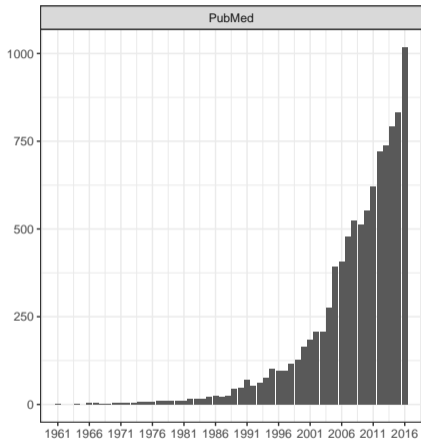
About me

- ▶ 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!



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 quickly¹:

	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}

¹Source: 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 studies²:

- ▶ 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.

²White IR, Morris T and Crowther MJ, unpublished

Enter SiReX

- ▶ Interactive tools can supplement the ADMEP framework very effectively
- ▶ SiReX [saɪə(ɪ)-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

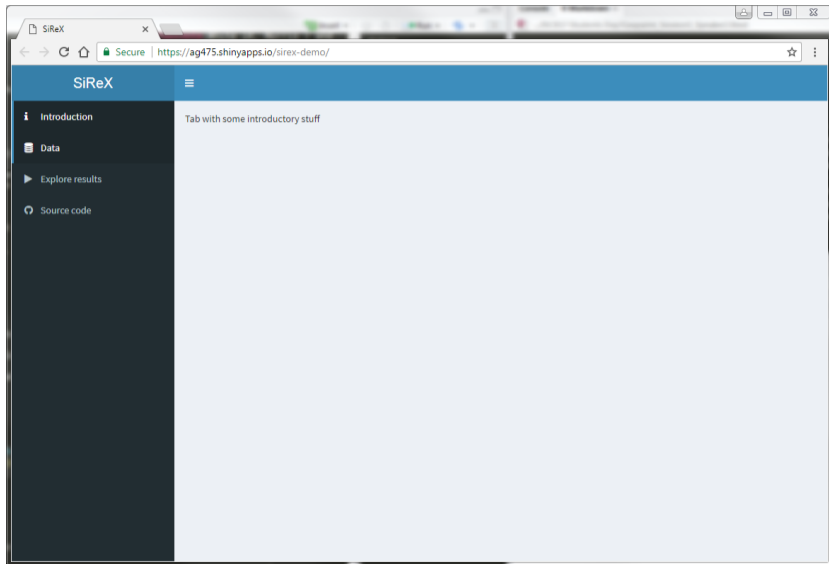
Key messages:

- ▶ 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

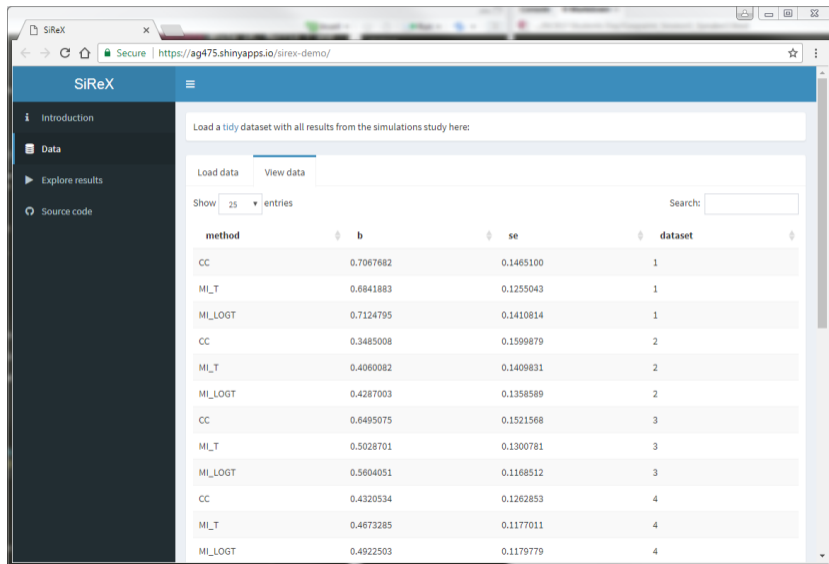


Demo: load data

The screenshot shows a web browser window with the URL `https://ag475.shinyapps.io/sirex-demo/`. The application header is blue with the text "SiReX" and a menu icon. A dark sidebar on the left contains navigation links: "Introduction", "Data" (highlighted), "Explore results", and "Source code". The main content area has a white background and contains the following elements:

- A text box: "Load a tidy dataset with all results from the simulations study here:"
- Two tabs: "Load data" (active) and "View data".
- A section titled "Upload your .csv file" with a "Browse..." button and a "No file selected" status.
- A "Load demo data" button.
- A paragraph of text describing the demo data: "Demo data: simulation study comparing different ways to handle missing covariates when fitting a Cox model (White and Royston 2009). One thousand datasets were simulated, each containing normally distributed covariates x and z and a time-to-event outcome. Both covariates had 20% of their values deleted independently of all other variables so the data became missing completely at random (Little and Rubin 2002). Each simulated dataset was analyzed in three ways. A Cox model was fit to the complete cases (CC). Then two methods of multiple imputation using chained equations (van Buuren, Boshuizen, and Knook 1999) were used. The `MI_LOGT` method multiply imputes the missing values of x and z with the outcome included as $\log(t)$ and d , where t is the survival time and d is the event indicator. The `MI_T` method is the same except that $\log(t)$ is replaced by t in the imputation model. The results are stored in long format, with variable 'dataset' identifying the simulated dataset number, string variable 'method' identifying the method used, variable 'b' holding the point estimate, and variable 'se' holding the SE."

Demo: landing page



SiReX

Introduction

Data

Explore results

Source code

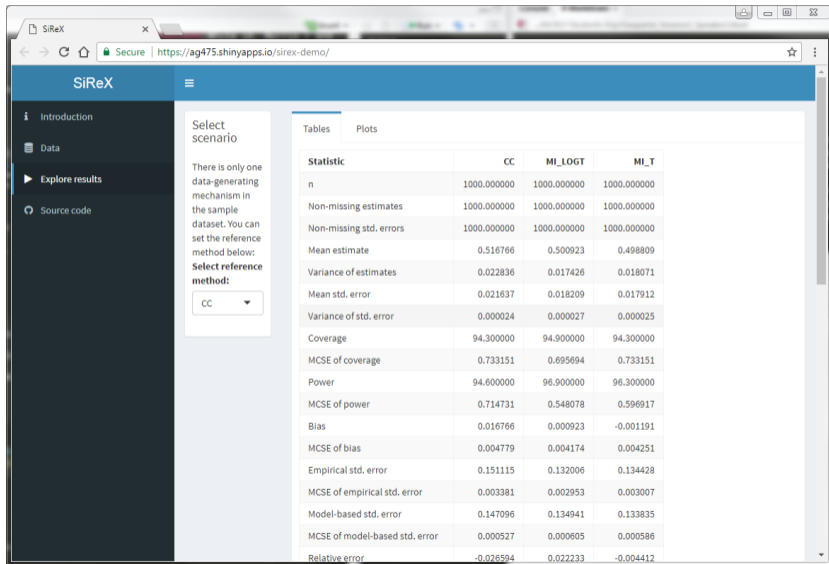
Load a tidy dataset with all results from the simulations study here:

Load data View data

Show 25 entries Search:

method	b	se	dataset
CC	0.7067682	0.1465100	1
ML_T	0.6841883	0.1255043	1
ML_LOGT	0.7124795	0.1410814	1
CC	0.3485008	0.1599879	2
ML_T	0.4060082	0.1409831	2
ML_LOGT	0.4287003	0.1358589	2
CC	0.6495075	0.1521568	3
ML_T	0.5028701	0.1300781	3
ML_LOGT	0.5604051	0.1168512	3
CC	0.4320534	0.1262853	4
ML_T	0.4673285	0.1177011	4
ML_LOGT	0.4922503	0.1179779	4

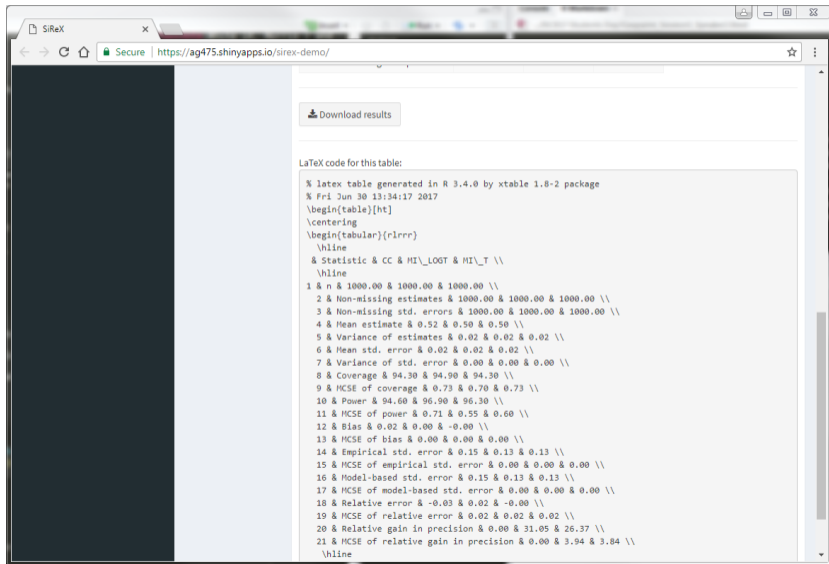
Demo: table of summary results



The screenshot shows the SiReX web application interface. On the left is a dark sidebar with navigation links: Introduction, Data, Explore results (active), and Source code. The main content area is titled "Select scenario" and contains a message: "There is only one data-generating mechanism in the sample dataset. You can set the reference method below: Select reference method:". Below this is a dropdown menu with "CC" selected. To the right, under the "Tables" tab, is a table with 4 columns: Statistic, CC, MI_LOGT, and MI_T. The table lists 20 different statistics and their corresponding values for the three methods.

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

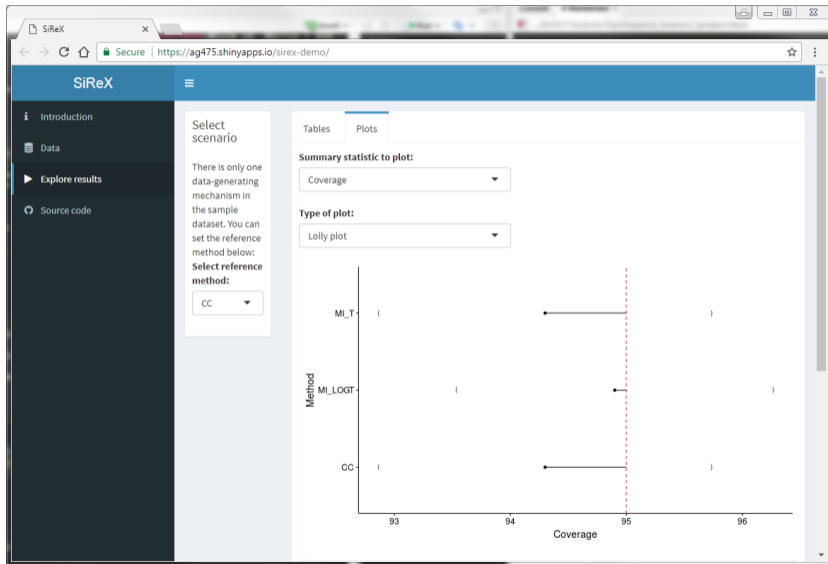
Demo: table of summary results



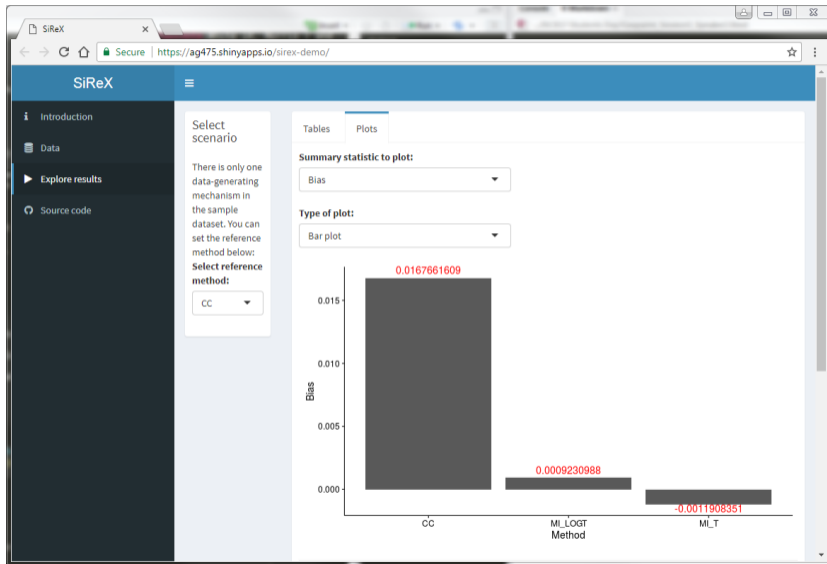
The screenshot shows a web browser window with the address bar displaying `https://ag475.shinyapps.io/sirex-demo/`. A button labeled "Download results" is visible. Below it, the text "LaTeX code for this table:" is followed by a code block containing LaTeX code for a table. The code uses the `xtable` package and generates a table with 21 rows and 4 columns. The table headers are: "Statistic", "CC", "MI", "LOGT", and "MI_T". The rows contain various statistical metrics such as "n", "Non-missing estimates", "Non-missing std. errors", "Mean estimate", "Variance of estimates", "Mean std. error", "Variance of std. error", "Coverage", "MCSE of coverage", "Power", "MCSE of power", "Bias", "MCSE of bias", "Empirical std. error", "MCSE of empirical std. error", "Model-based std. error", "MCSE of model-based std. error", "Relative error", "MCSE of relative error", and "Relative gain in precision".

```
% latex table generated in R 3.4.0 by xtable 1.8-2 package
% Fri Jun 30 13:34:17 2017
\begin{table}[ht]
\centering
\begin{tabular}{rllrr}
\hline
& Statistic & CC & MI\_{LOGT} & MI\_T \\
\hline
1 & n & 1000.00 & 1000.00 & 1000.00 \\
2 & Non-missing estimates & 1000.00 & 1000.00 & 1000.00 \\
3 & Non-missing std. errors & 1000.00 & 1000.00 & 1000.00 \\
4 & Mean estimate & 0.52 & 0.50 & 0.50 \\
5 & Variance of estimates & 0.02 & 0.02 & 0.02 \\
6 & Mean std. error & 0.02 & 0.02 & 0.02 \\
7 & Variance of std. error & 0.00 & 0.00 & 0.00 \\
8 & Coverage & 94.30 & 94.90 & 94.30 \\
9 & MCSE of coverage & 0.73 & 0.70 & 0.73 \\
10 & Power & 94.60 & 96.90 & 96.30 \\
11 & MCSE of power & 0.71 & 0.55 & 0.60 \\
12 & Bias & 0.02 & 0.00 & -0.00 \\
13 & MCSE of bias & 0.00 & 0.00 & 0.00 \\
14 & Empirical std. error & 0.15 & 0.13 & 0.13 \\
15 & MCSE of empirical std. error & 0.00 & 0.00 & 0.00 \\
16 & Model-based std. error & 0.15 & 0.13 & 0.13 \\
17 & MCSE of model-based std. error & 0.00 & 0.00 & 0.00 \\
18 & Relative error & -0.03 & 0.02 & -0.00 \\
19 & MCSE of relative error & 0.02 & 0.02 & 0.02 \\
20 & Relative gain in precision & 0.00 & 31.05 & 26.37 \\
21 & MCSE of relative gain in precision & 0.00 & 3.94 & 3.84 \\
\hline
\end{tabular}
\end{table}
```

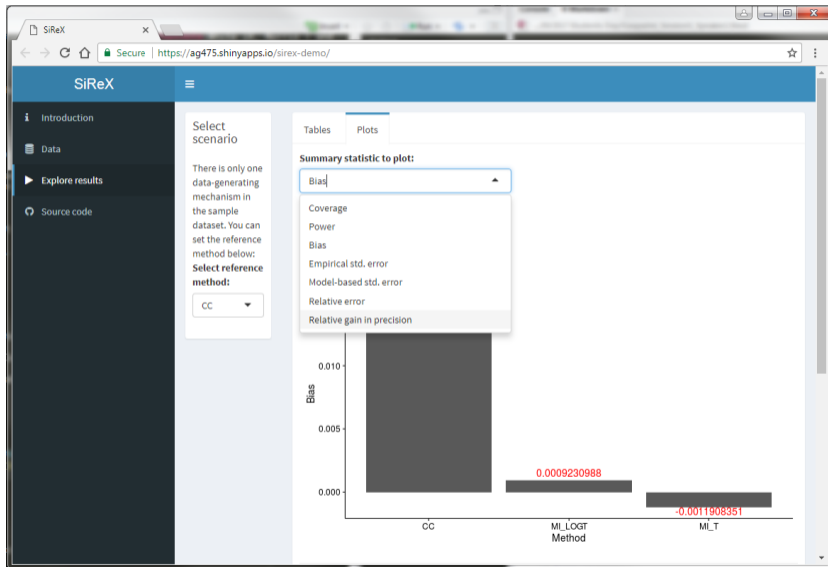
Demo: plot of summary results



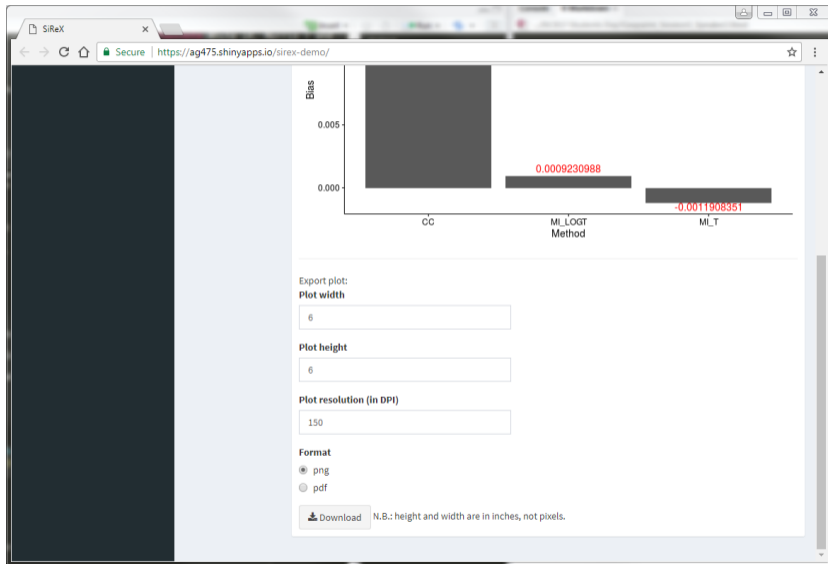
Demo: plot of summary results



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?