Visualisation of results from large simulation studies: introducing the multi-performance plot and its Shiny app

Wang Pok Lo[1], Christopher Weir[2]

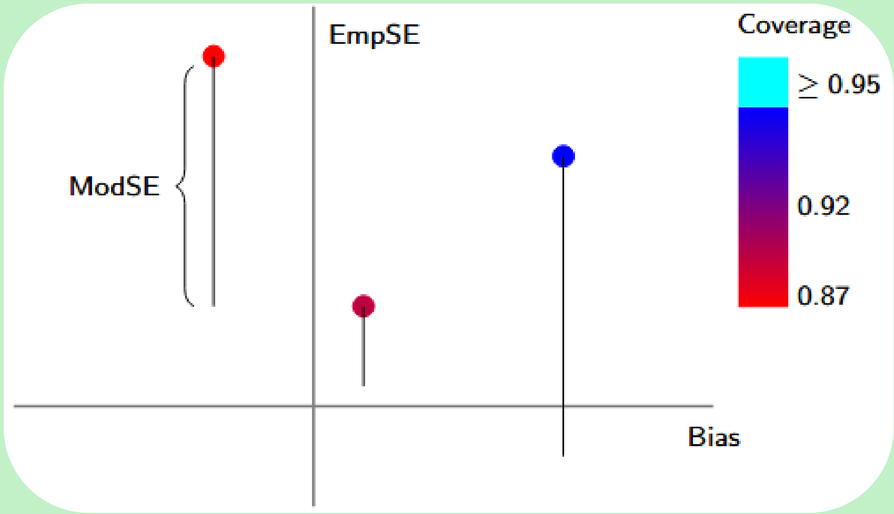
Contact email address: pok.lo@gtc.ox.ac.uk

[1]School of Medicine and Biomedical Sciences, Green Templeton College, University of Oxford, UK
[2]Edinburgh Clinical Trials Unit, Usher Institute, School of Population Health Sciences, University of Edinburgh, UK

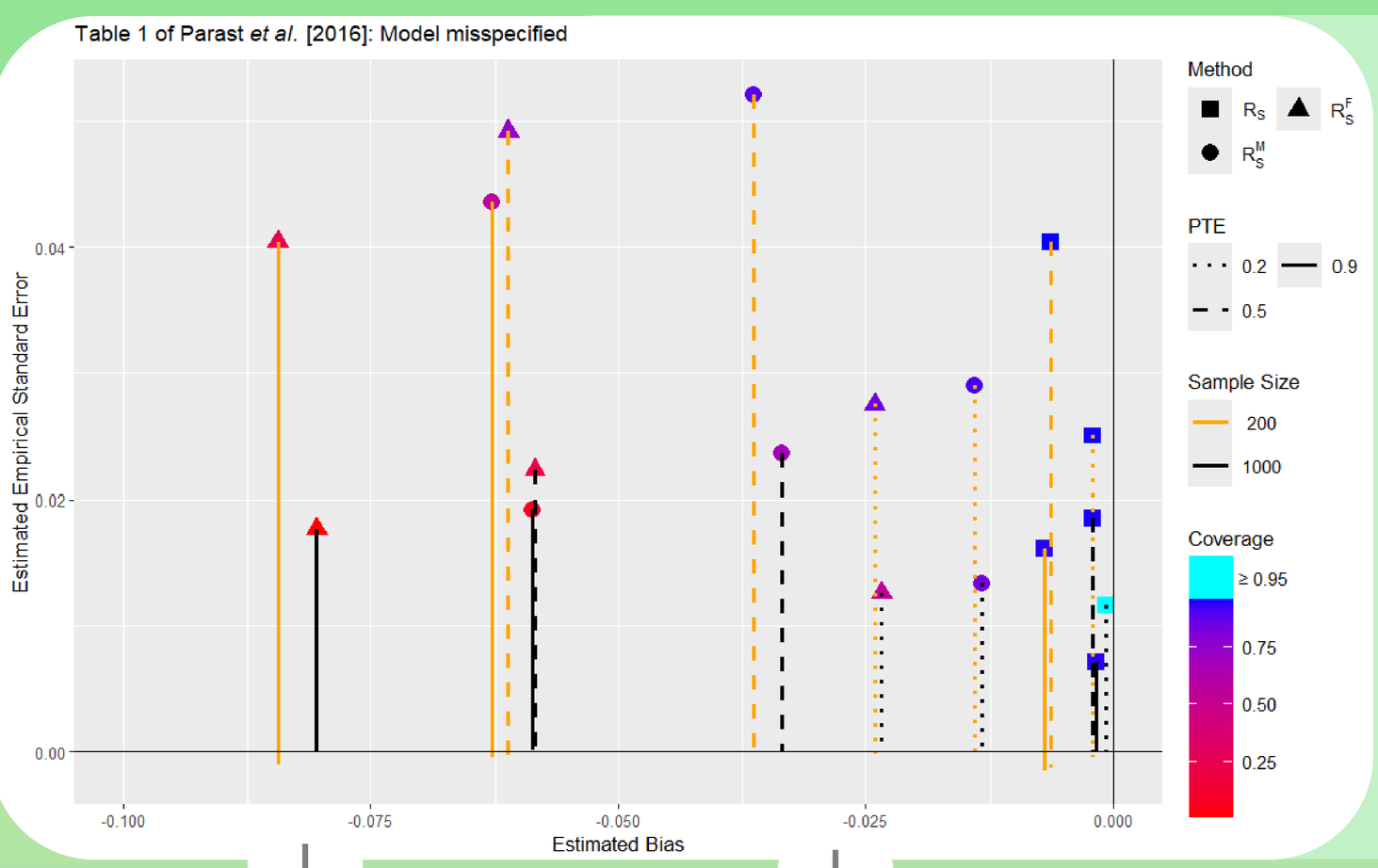
Simulation Studies: Performance Measures

- **Simulation studies** are often used to assess and compare the performances of different statistical methods. Measures of performance include:
  (Morris *et al.* [2019])
  - **Bias** *how far the method under/overestimates the true value*
  - **EmpSE** *how consistently the method estimates the true value*
    *[**Emp**irical **S**tandard **E**rror]*
  - **MSE** $Bias^2 + EmpSE^2$
    *[**M**ean **S**quared **E**rror]*
  - **ModSE** *how uncertain the method is in its estimates*
    *[Average **Mod**el **S**tandard **E**rror]*
  - **Coverage** *how often this range of uncertainty contains the true value*
- When **many methods** are assessed under **many combinations of parameters**, a graphical display of performance measures may be more conducive to **readily identify patterns** compared to tabular versions.

Multi-Performance Plot: Basic Design



The **line colours**, **line types** and **shapes of points** can be modified to reflect different parameters and statistical methods.

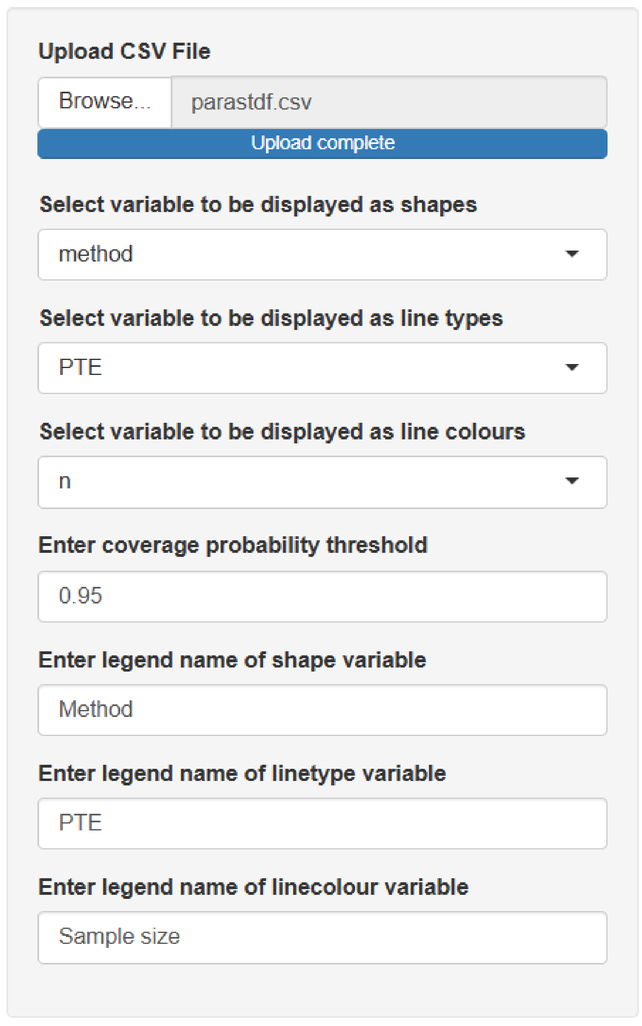


Multi-Performance Plot: Example

Parast *et al.* [2016] developed a **method R_S**, which extended previous methods **(R_M and R_F)** to estimate the statistical metric **PTE** ∈ [0,1]. PTE measures how well a substitute endpoint can replace a patient-relevant endpoint in a clinical trial. Performance was assessed using 3 true values of PTE under two sample sizes.

- All methods underestimate PTE, but R_S performs better **(blue squares closer to *y*-axis)**
- EmpSE is approximately equal to ModSE across all methods **(all vertical lines terminate around *x*-axis)**
- Larger sample size produces more precise estimates **(black lines are shorter than orange lines)**
- All methods tend to under-cover PTE, but R_S performs better **(cyan / dark blue points >> red points)**

Multi-Performance Plot



Integration with *rsimsum* package for simulation studies

```
library(rsimsum) # Package for running simulation studies (Morris et al. [2019])

initialdf <- tidy(simsum(data = relhaz,...))) # Tidy dataframe of performance measures

coredf <- (initialdf %>% filter(stat %in% c("bias", "empse", "modelse",
"cover")))[-3] # Selects only estimates of Bias; EmpSE; ModSE; Coverage

finalperfDF <- cbind(t(as.data.frame(split(coredf$est,
ceiling(seq_along(coredf$est)/4)),
row.names=c("bias","empse","modelse","cover"))), unique(coredf[,3:5]))
                # Generates dataframe that can be converted to
                  CSV compatible with RShiny app
```

RShiny App: Multi-Performance Plot

https://poklo.shinyapps.io/MultiPerformancePlot_1-0/

GitHub guidance

Current features:
(1) automated plot generation given `compatible CSV`
(2) cyan colouring changes depending on user-specified coverage threshold
(3) options to change representations of shapes, line types and line colours

References
Morris TP, White IR, Crowther MJ. Using simulation studies to evaluate statistical methods. Statistics in medicine. 2019 May 20;38(11):2074-102.
Parast L, McDermott MM, Tian L. Robust estimation of the proportion of treatment effect explained by surrogate marker information. Statistics in medicine. 2016 May 10;35(10):1637-53.