Health Economic Modeling in R: A Hands-on Introduction

Why use R?

QUEEN'S UNIVERSITY BELFAST

UCL

University of BRISTOL

BresMed™

Introduction

- Challenges with health economics modelling include:
  - “Big Data”
  - Different data types need to be combined
  - Making decisions with a lot of uncertainty
  - Many stake-holders (patients, hospitals and governments)

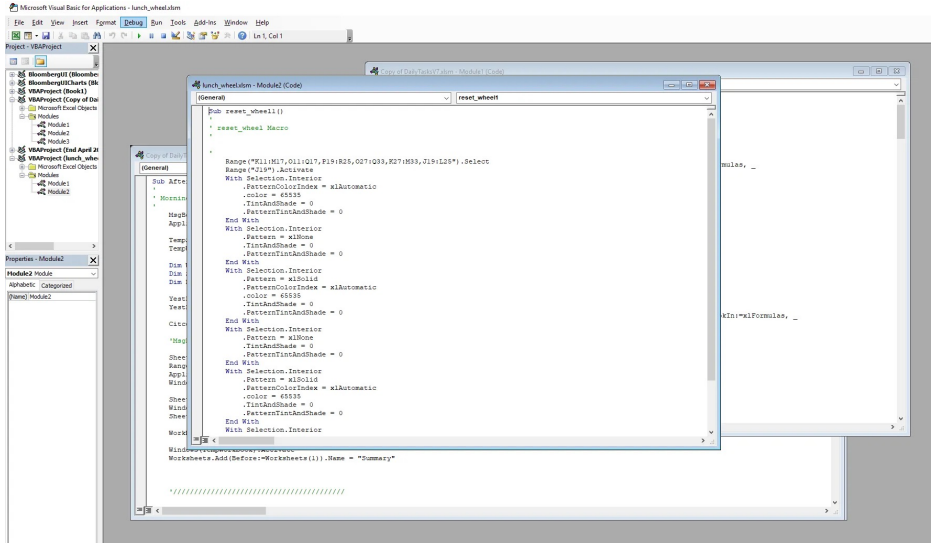
Microsoft Excel

- Excel is a spreadsheet developed by Microsoft
- Features calculation, graphing tools, pivot tables
- Macro programming language called Visual Basic for Applications (VBA)
  - Visual Basic Editor (VBE)
- Industry standard for spreadsheets
- Excel forms part of Microsoft Office

Why use Excel?

- Learning curve
  - Easy and efficient for constructing simple models
  - Point-and-click GUI
  - Often used in introductory courses and textbooks
- Excel models are implicitly transparent
- Extensible
  - With the use of macros
  - Numerous third-party packages to enhance functionality

Extensible via VBA



- Allows spreadsheet manipulation that is awkward or impossible with standard spreadsheet techniques
- Programmers may write code directly using the Visual Basic Editor (VBE)
- A common and easy way to generate VBA code is by using the Macro Recorder

Why I shouldn’t use Excel?

- Constructing more complex models can be hectic
- Models can become opaque through “cell chasing” where references and names lead to a tangled web of variables
- Repeated (probabilistic) simulation and reproducibility is not straightforward - more complex simulations require an increasingly sophisticated command of VBA
- Steps of an analysis are not explicitly recorded
- Plotting function is limited – graphics not the nicest!
- Large data set reading, handling and saving in different formats can be difficult

In Summary

- Excel is great for what it is designed to do
- It’s intuitive, simple to use and transparent for smaller models
- Are we beyond simple models for health economics?
- Excel works well *complementary to* other tools


Excel
R

So what is R?

- R is a free, open source software program
- R is arguably the go-to software for data science and statistics
- R provides a wide variety of statistical and graphical techniques
  - Linear and non-linear models, clustering, classification, statistical tests
- High quality/publication-quality plots can be produced easily, including mathematical symbols and formulae where needed

Why use R?

- Free and open source
- Graphics - R comes with great abilities in data visualization including static, interactive and all levels of complicated
- Arguably better than SAS and Stata in terms of availability of advanced statistical methods and algorithms, through availability of user-created packages
- Packages for written reports and analysis code in one document
- Simple syntax
- Interacts with other software, including Excel, C, Python, SQL, stan, WinBUGs and others

Why use R?

- R uses command-line scripting, ideal for recycling that analysis on similar sets of data
- Upgrades to the software are much more regular than Excel
- R has a very large and active online community supplying documentation, tutorials and online query forums
  - It is now supplemented by more than 16,000 community developed open-source packages available for download from The Comprehensive R Archive Network (CRAN)
  - Authors often supplement the package submission with a publication in the Journal of Statistical Software, with more rigorous documentation and relevant theoretical material

Computational Speed

- Critical component of cost-effectiveness modelling software because it is one of the key determinants of
  - how long it will take to produce a given analysis
  - scope of analyses that are feasible within the time available for a project
- Complex models increase the computational burden and when performing multi-stage Monte-Carlo simulations, any inefficiency in model implementation is exacerbated
- Marked differences in the time it takes to run simulations has implications for the costs of undertaking research and for the ability to use decision analytic modelling as part of research and design processes and iterative evaluation processes

Why I shouldn’t use R?

- Programming is required!
- Is relatively slow compared to other lower level programming languages e.g. C ?
- People just may be more comfortable with analyses in something other than R

RStudio

- RStudio is a free and open-source integrated development environment (IDE) for R, a programming language for statistical computing and graphics
- RStudio is available in open source and commercial editions
- Its interface is organized so that the user can clearly view graphs, data tables, R code, and output all at the same time
- Also offers an Import-Wizard-like feature that allows users to import CSV, Excel, SAS (*.sas7bdat), SPSS (*.sav), and Stata (*.dta) files into R without having to write the code to do so.

In Summary

- R is very so useful in numerous fields and growing all the time
  - CRAN is a vast repository of tested and documented packages
  - Other source can also be used e.g. GitHub repos
- The learning curve for R is not steep
- Graphics/visualisation are a lot nicer than the likes of excel
- Its free!