Why use R for Health Economics Modelling?

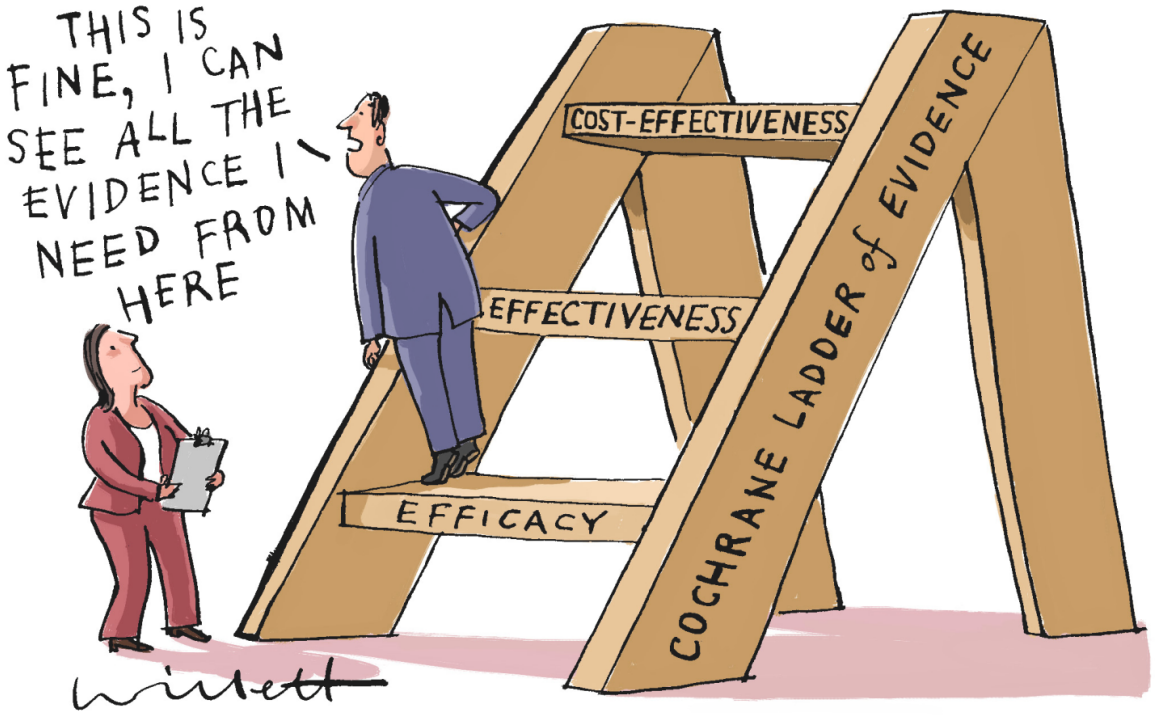
Introduction

- Health economics is concerned with issues related to efficiency, effectiveness, value and behaviour in the production and consumption of health and healthcare

- 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)

Introduction

- Model the real-world scenario
- Assumptions made
- Decision needs to be made
- “Evidenced based decision making”



Transparency

Eddy *et al.* state that

“transparency refers to the extent to which interested parties can review a model's structure, equations, parameter values, and assumptions.”. They identify two levels at which this is important; the first is to allow a general understanding of the model and the second a more technical understanding of the model.

What should we use to model build?

- What software is mostly used currently?
- What are the issues with the current software?
- What can and should a different software do better?
- What is the demand for new tools and by whom?
- Why do these need to be addressed? What will happen if they aren’t?
- How widespread is the issue?
- How long has it been a problem and what the impact?
- What evidence is there of this?

Microsoft Excel

- Excel is a spreadsheet developed by Microsoft for Windows, macOS, Android and iOS
- 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
- Available cross-platform
  - E.g. MS Windows and Mac OS X operating system
- Links between Excel and other components of the MS Office Suite e.g. Word and PowerPoint, support efficient production of reports and presentations for dissemination

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)
  - Includes a window for writing code
  - Debugging code
  - Code module organization environment
- User can implement numerical methods as well as automating tasks such as formatting or data organization
- Any desired intermediate results can be reported back to the spreadsheet
- A common and easy way to generate VBA code is by using the Macro Recorder

Transparency

- Parameters and model structure should be obvious
- Use of cell names is required for efficient updating of parameter values apart from the simplest models
- Models can become opaque through “cell chasing” where references and names lead to a tangled web of variables
- The ability to highlight these relationships, using the trace precedent and dependent facility, offers some reprieve
- Validation of macros requires additional technical skill
- The seed cannot be set so any model should be reproducible upon demand

Why I shouldn’t use Excel?

- Constructing more complex models can be hectic
- 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!
- Updated infrequently
- 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
- However, when models are more complex its strengths can become its weaknesses
  - E.g. navigating a model (cell chasing)
- Excel works well *complementary to* other tools
  - E.g. model validation, prototyping, etc


Excel
R

So what is R?

- R is a free, open source high-level software program
- R is arguably the go-to software for data science and statistics
- R provides a wide variety of statistical (linear and nonlinear modelling, classical statistical tests, time-series analysis, classification, clustering, ...) and graphical techniques
- Ease with which well-designed publication-quality plots can be produced, including mathematical symbols and formulae where needed

Why use R?

- Free and open source.
- Available for Windows, Macintosh, and Linux
- Graphics - R comes with great abilities in data visualization, should the visualization be static, interactive and even far more complicated. Publication quality graphics!
- Rivals (and in many cases, exceeds) SAS and Stata in terms of availability of advanced statistical methods and algorithms, through availability of user- created packages
- Packages for *literate statistical programming* - weaving 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, which is ideal for storing numerous series of complex data-analysis and recycling that analysis' on similar sets of data
- Upgrades to the software are much more regular
  - This is extremely advantageous for statistical programming languages and environments
- R's large and active online community supply a myriad of documentation, tutorials and online query forums
  - It is now supplemented by more than 8000 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, e.g. loops, to other lower level programming languages e.g. C
  - but can leverage this by linking with them
- 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 and runs on the desktop (Windows, macOS, and Linux)
- 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 powerful for 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 he likes of excel
- Its free!