



# How I'm selling R at GSK

Andy Nicholls  
(GSK)

# Outline

- My background / Industry background
- What I've done to promote the use of R
- Has it worked?
- The future...

# My Background

- Graduated from university with degrees in Maths and in Medical Statistics
  - Used R for some basic statistics and linear modelling for undergraduate degree
  - Further R & S training in postgraduate degree
  - Simulation project using R for postgraduate dissertation
- Joined GSK as a Statistician in 2007
- Told by anyone I spoke to in the industry that R is generally NOT used and that I should focus on SAS!!!

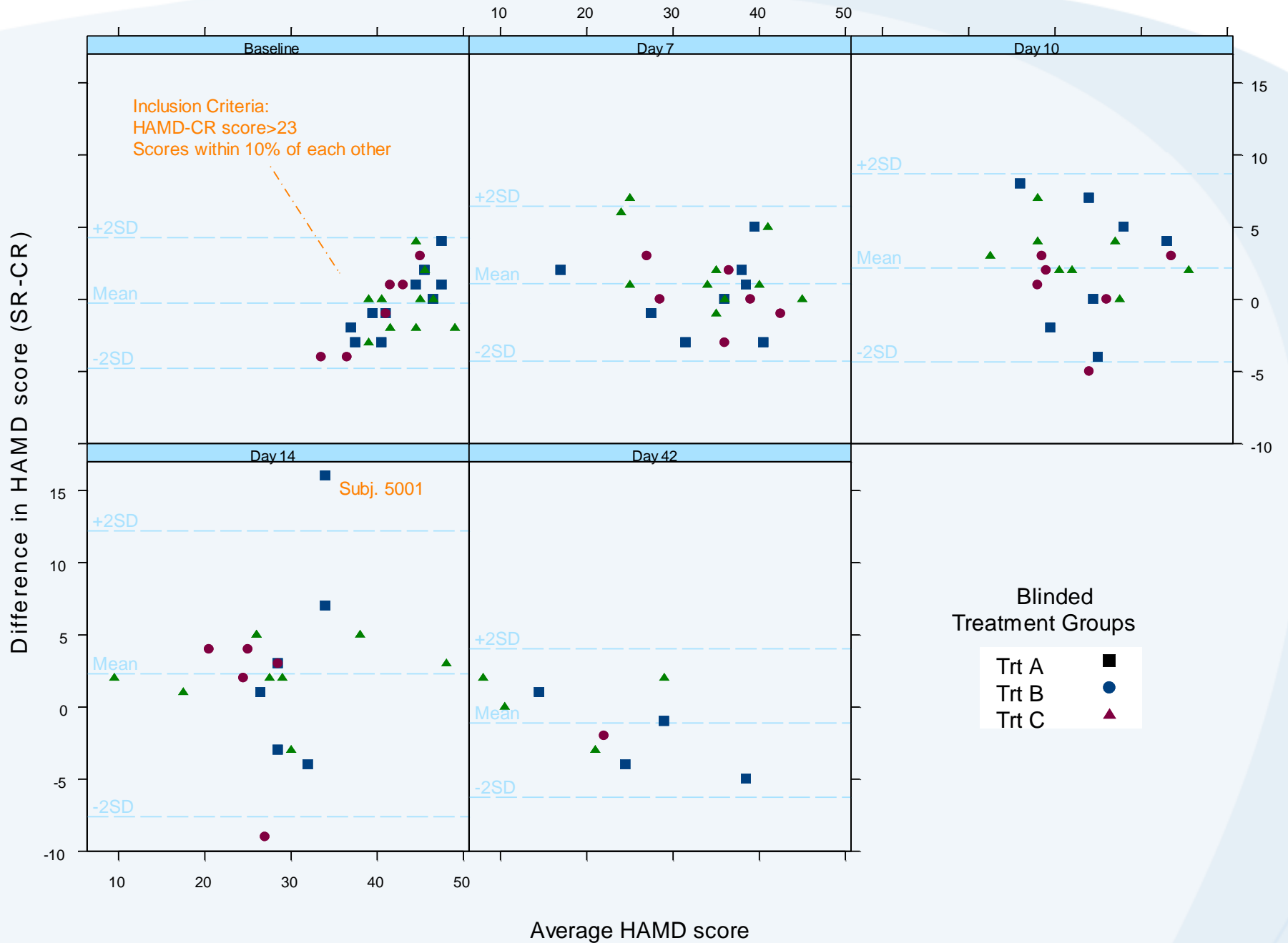
# Why I have to fight to use R in my industry

- Pharmaceutical Industry traditionally uses SAS
  - SAS (generally) thought to be the only software that regulators will accept (not true)
  - Statistical Analysis Plans written with SAS in mind, sometimes even containing example SAS code
  - Industry is not very progressive statistically
    - Same methods and therefore the same SAS code
  - The expertise is generally in SAS
- However, like me most new graduates entering the industry now have experience of R (or S)

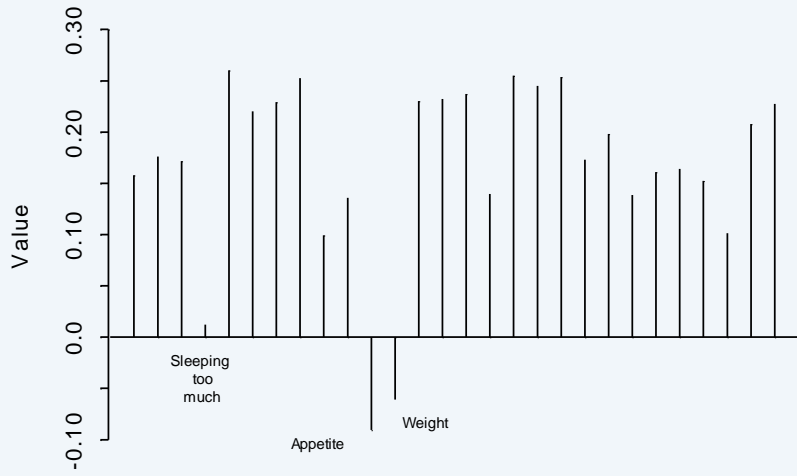
# How can I get people using R in my team/company?

1. Look for advantages over alternative software (i.e. SAS)
  - R graphics (arguably) look better than SAS
  - Lattice graphics in R

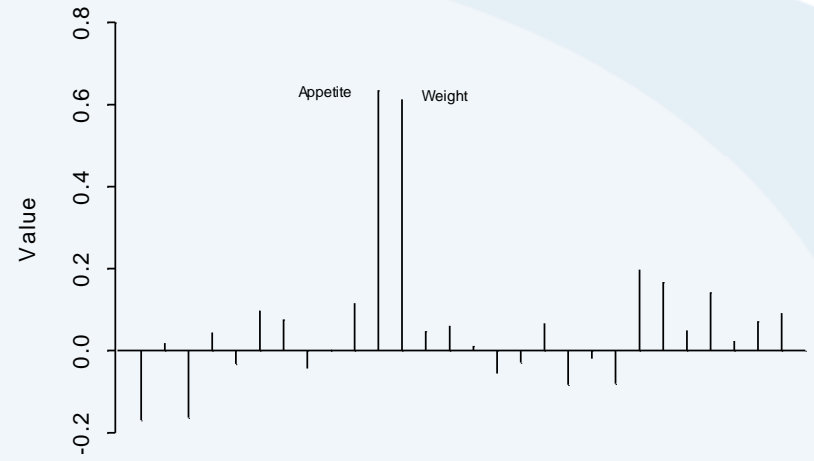
# Bland-Altman Plots to Compare HAMD-CR and HAMD-SR scores at each visit



### First Principal Component



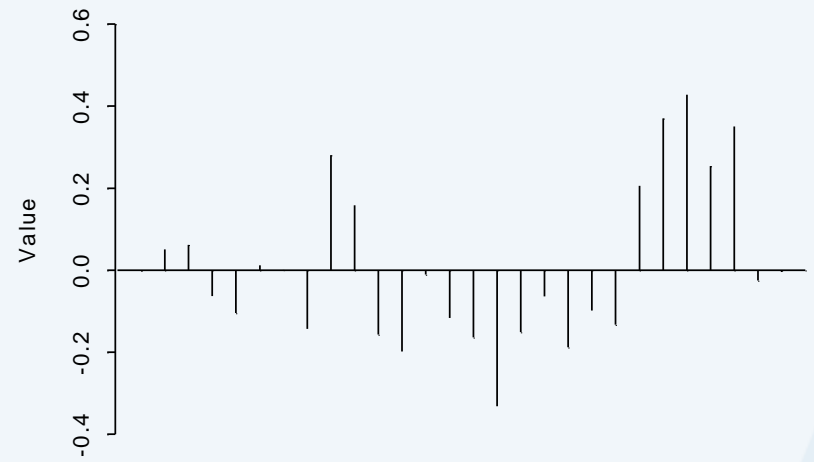
### Second Principal Component



### Third Principal Component



### Fourth Principal Component



# How can I get people using R in my team/company?

1. Look for advantages over alternative software (i.e. SAS)
  - R graphics (arguably) look better than SAS
  - Lattice graphics in R
2. Seek to use R for any non-standard work
  - Study teams are easily impressed by new and different analyses and/or graphics so I used R for any non-standard work
    - Simulation
    - Highly customised graphics



# Simulated outcomes of an Interim Analysis

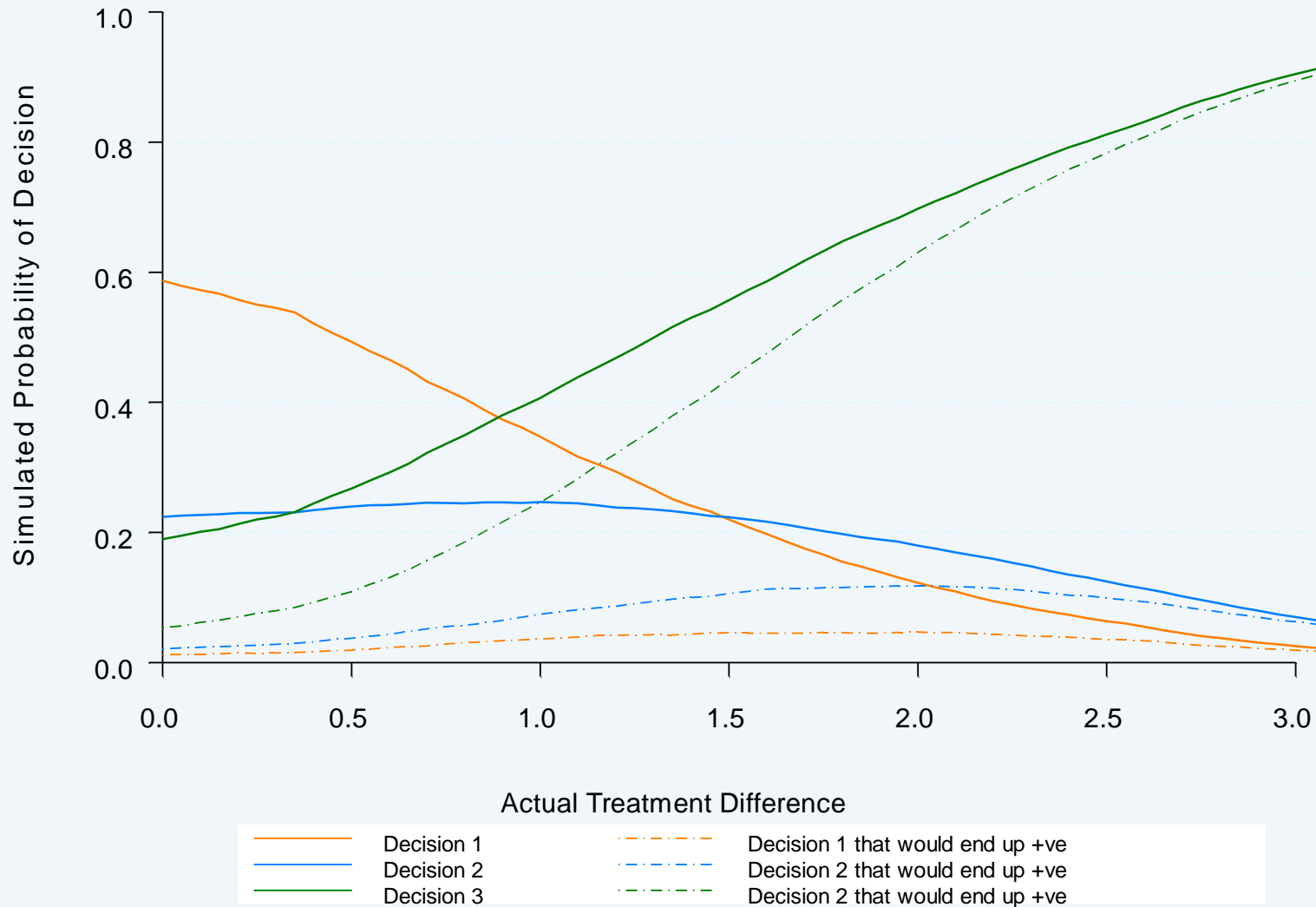
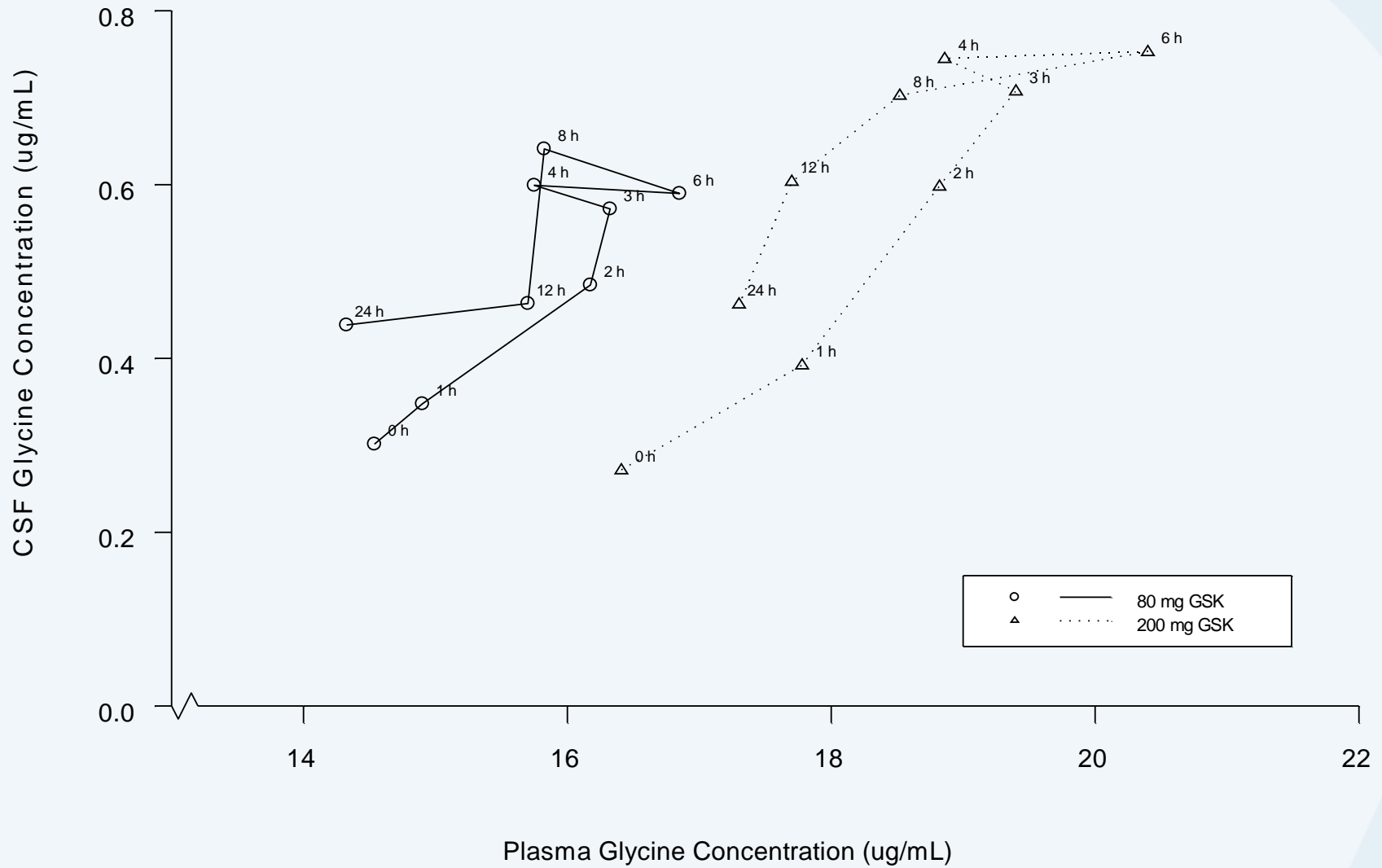


Figure 13.7 Plot of Mean CSF Glycine Concentration by Mean Plasma Glycine Concentration



# Estimated Difference from Placebo for Main Efficacy Endpoints

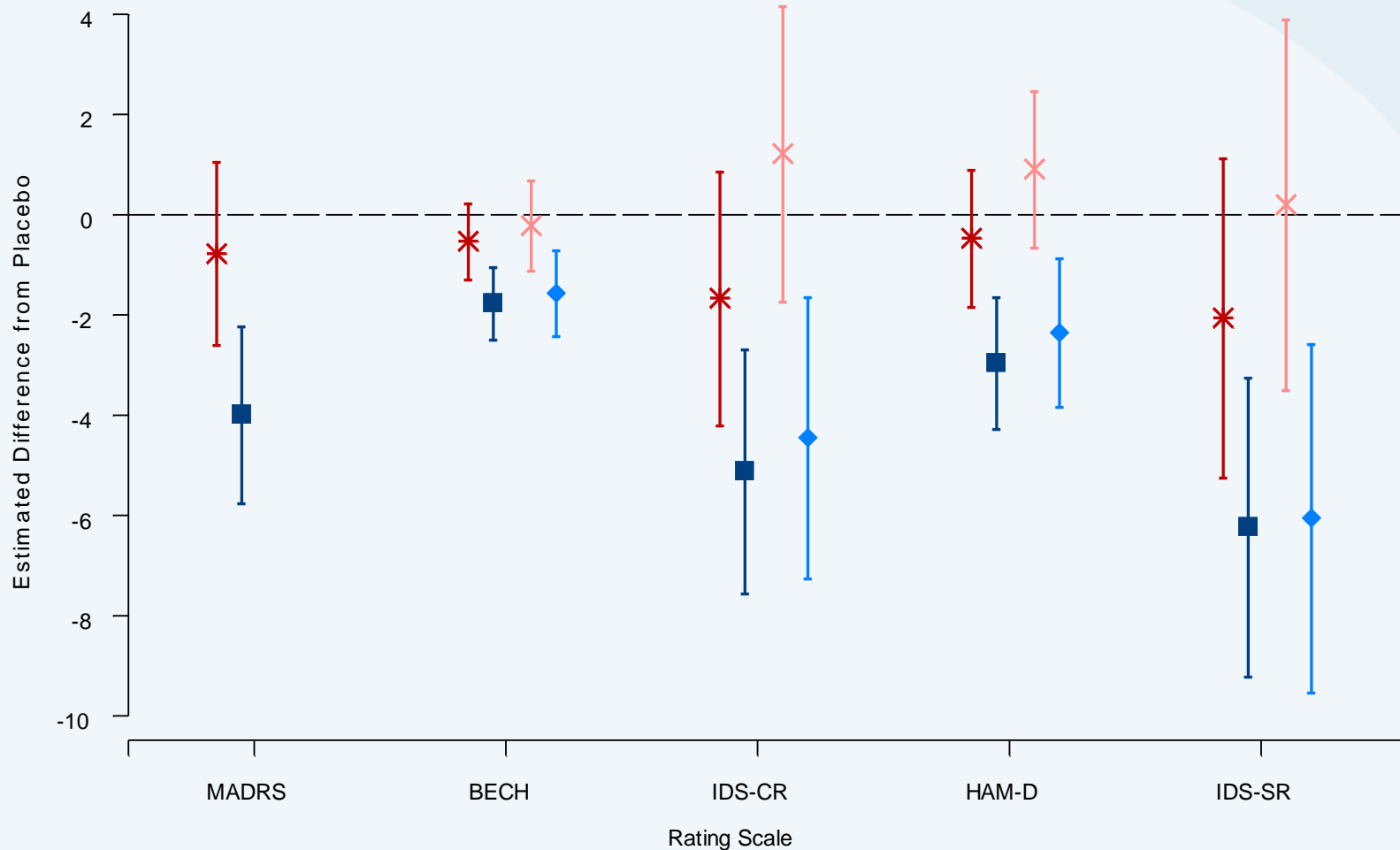
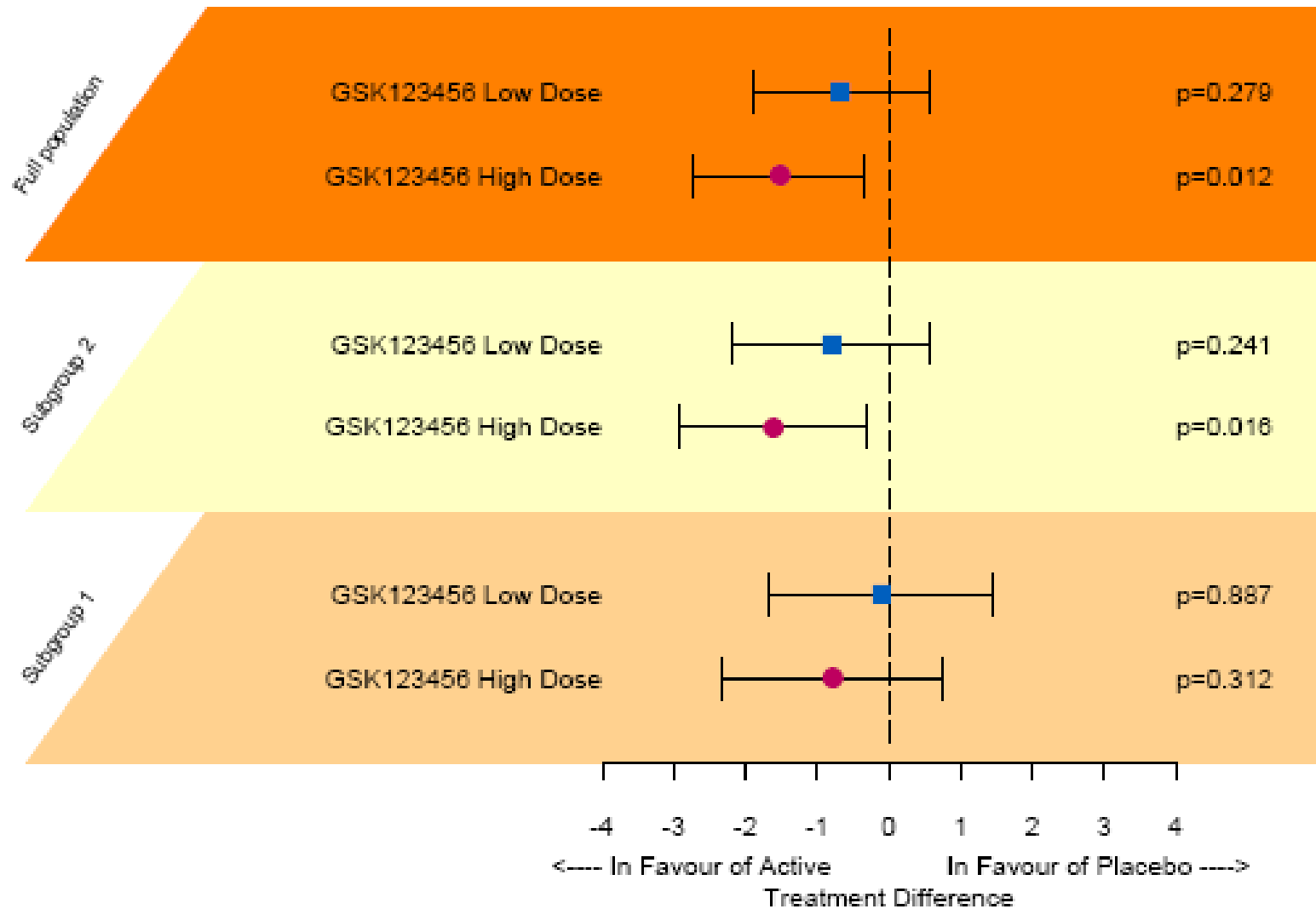
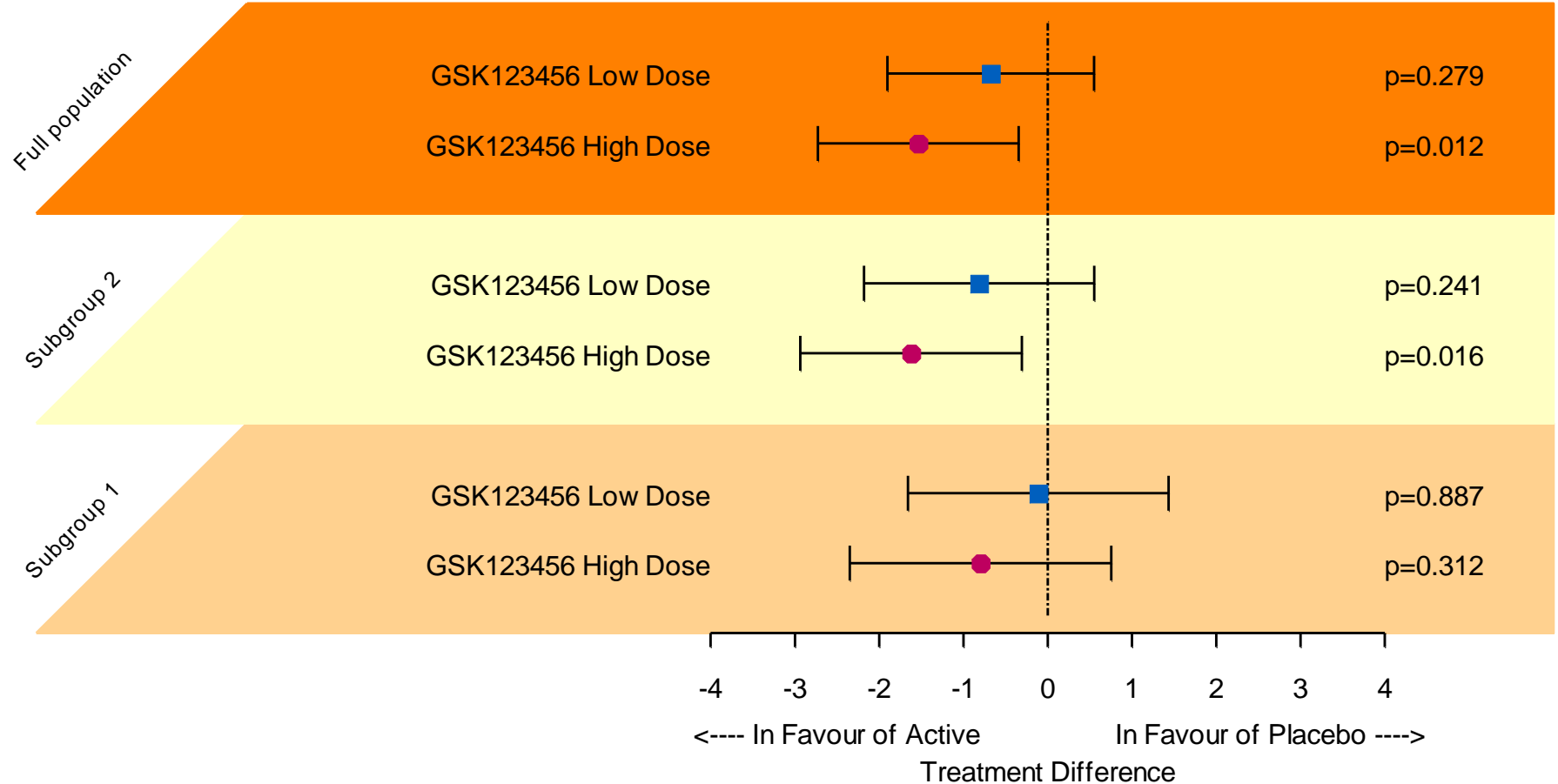


Figure 01.01  
Results of the Primary and Subgroup Analysis on Endpoint at Week X



# Results of the Primary and Subgroup Analysis on Endpoint at Week X



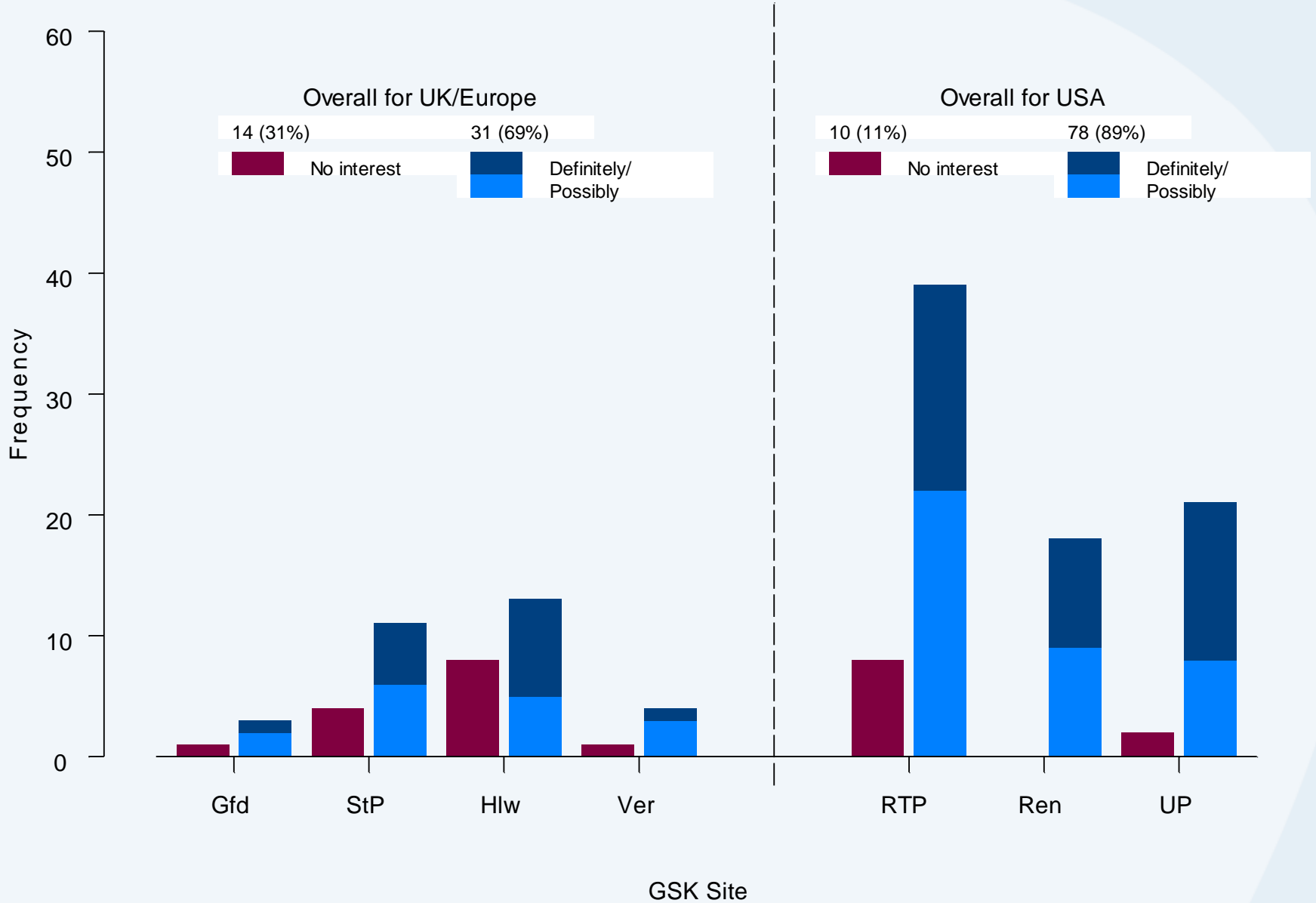
Acknowledgements for George Zannoupas for starting this work

# Encouraging others

- Training courses
  - Encouraged attendance of Basic R Scripting Course\*
  - Set up an 'Advanced' Scripting course to push people forward to the next level
  - (Both run by Mango)
- Poster
- Presentation at annual company conference
- Provide support for others
  - Put my name out as a user and encourage others to ask me programming questions
- Presentation on R graphics at PSI Meeting on graphics

\* - Acknowledgements for Mark Jones and Rich Pugh for arranging the first of these

# Interest in the Basic S-Plus/R Scripting course by GSK Site



# Clinical Graphics using S-Plus and R

Andrew Nicholls  
Neurosciences Discovery Biometrics

## Introduction

Graphics is currently a hot topic at GSK, and indeed within the pharmaceutical industry as a whole. SAS remains the tool of choice for many and the release of SAS 9.2 has excited a lot of people with its new graphics capabilities. At GSK we also have the GWE, which has vastly improved both the quality of graphics we are producing, and the ease with which we can produce and export these graphics to make them available to colleagues. So why bother scripting and why use S-Plus, or indeed R? This poster highlights some of the areas in which S-Plus and R excel over rival packages.

## Customisation

Customisation remains one of the most appealing reasons for scripting in S-Plus and R. One reason for this is that both of them allow the user to build their graph from scratch, and then continue to build upon this template. The user can define everything from the thickness of an axis line to the orientation of text annotation. This line-by-line approach makes it very easy to target specific elements of a graphic and, if necessary, make a change to this element without impacting on other aspects of the graphic.

Figure 1 - Interest in Basic Scripting Course by Site



## Simulation

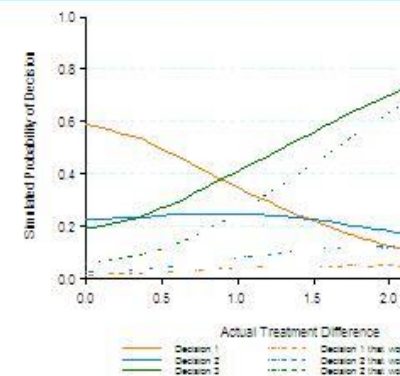
With adaptive designs becoming increasingly popular, it is important to have appropriate tools to handle the large number of simulations that will usually be required. This is an area in which S-Plus, and to an even greater extent R, really excel over rival software packages. The syntax of the S language, used by both packages, is relatively easy to follow and a range of scenarios can be simulated with minimal difficulty. Adding dimensions to a problem is also refreshingly straightforward.

Having produced thousands of simulations in order to address a particular question, simply inspecting the table of results is not going to be a particularly fast way learning about the possible scenarios. It is much more effective to produce a graphic. Of course, any piece of software worth its salt will have the capacity to produce an appropriate graphic, but S-Plus and R have the advantage of being particularly adept at both simulation and the production of quality graphics.

Figure 2 - Consequences of Interim Analysis for a Range of True Treatment Differences (v)



Figure 3 - Consequences of Interim Analysis for a Range of True Treatment Differences (v)



## Trellising

Historically, any poster advocating S-Plus and R have included a large section promoting their graphical capabilities. Indeed for many years S-Plus and R have had their version of the graphical lattice. As an idea, other companies have begun to develop similar features. However the trellis functionality of S-Plus, and it is worth devoting a little time to

Mastering the trellis functionality is not a small task. Most features in S-Plus and R however, there are several different ways of producing what you want. Looking for Figure 4 was produced in pro



# Has any of this worked?

- When I joined GSK in October 2007, there had not been any form of formal training in R
- In 2009 we ran 3 'Basic' training courses globally and 4 'Advanced' courses
  - Over 120 employees attended these courses
  - In the UK, more employees took the advanced course than any other course we offered in 2009
- I am now regularly asked for help from colleagues who are either using R themselves and have got stuck, or want me to produce something for them using R
  - So maybe people are actually using it too!

# The Future

- Further training planned for 2010-2011\*
- PSI also now running their first training course later in the year / early next year\*
- I'm going to keep plugging it in any way I can!

And most importantly...

- Students continue to leave university with a strong grounding in the basics of R

\* Mango providing training

# Recap

## Situation:

- Pharmaceutical Industry very aligned towards using SAS

## How I've been selling R:

- Looked for advantages over alternative software (i.e. SAS)
- Sought to use R for any non-standard work
- Organised training courses
- Submitted a poster and gave a presentation at annual conference
- Provided support for colleagues

## The future:

- More courses planned for 2010-11 (internally and externally)
- Graduates continue to leave university with skills in R

# References

1. Bland, Martin J. and Altman, Douglas G., 1986: Statistical Methods for Assessing Agreement Between Two Methods of Clinical Measurement. *The Lancet*, 327 (8476) pp. 307–310

**Questions?**