

# How GSK is using R to Drive Agility in Clinical Reporting

Andy Nicholls, Head, Statistical Data Sciences

Tilo Blenk, Director, Statistical Data Sciences

Michael Rimler, Programming Leader, Clinical Programming

# General Overview



## Using R to Drive Agility in Clinical Reporting

---

- Introduction
    - The Biostatistics challenge
    - Why not R?
    - Why now?
  - Build capability
    - Training roadmap
- } Andy Nicholls, ~20 mins

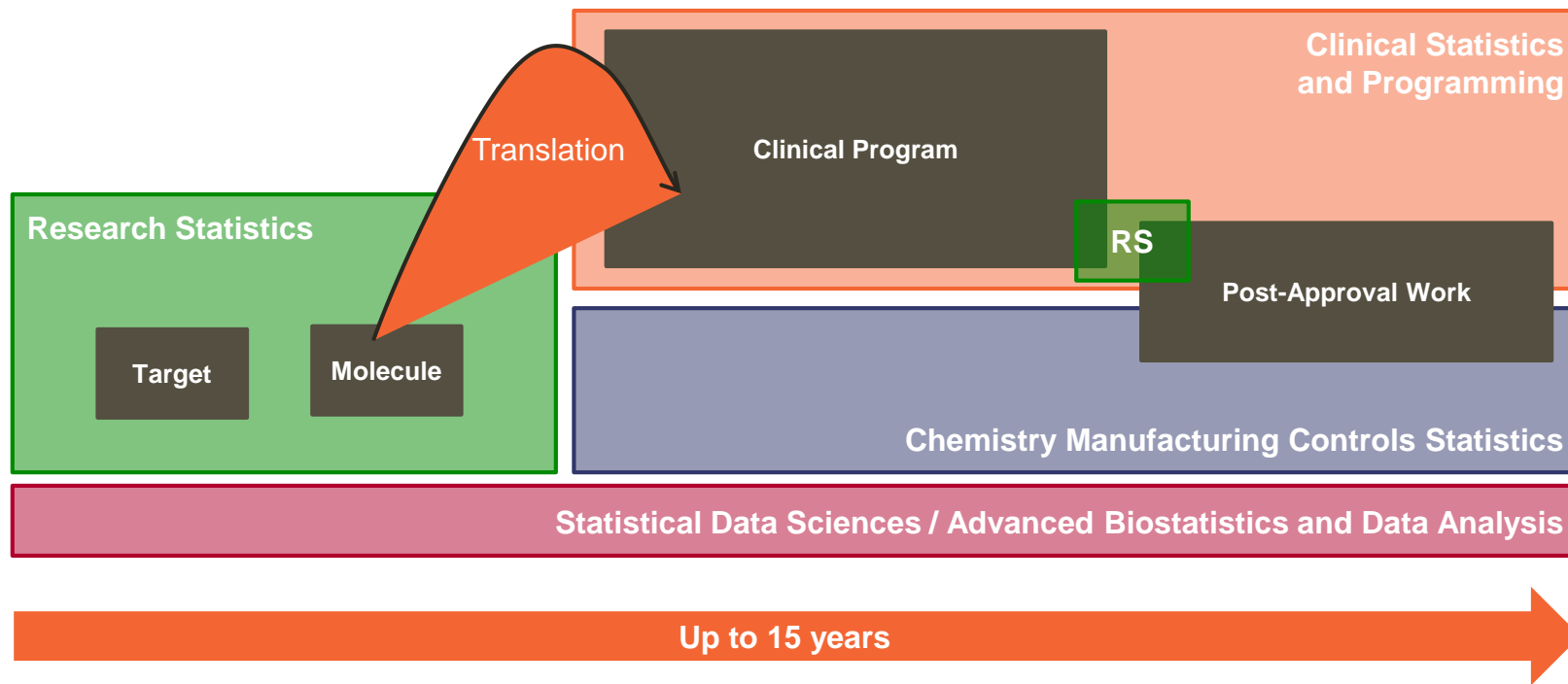
Then ...

- Building a fit for Purpose R Environment (Tilo Blenk, ~15 mins)
  - R4QC – An R Adoption Story (Michael Rimler, ~ 20 mins)
-

# The Biostatistics Challenge



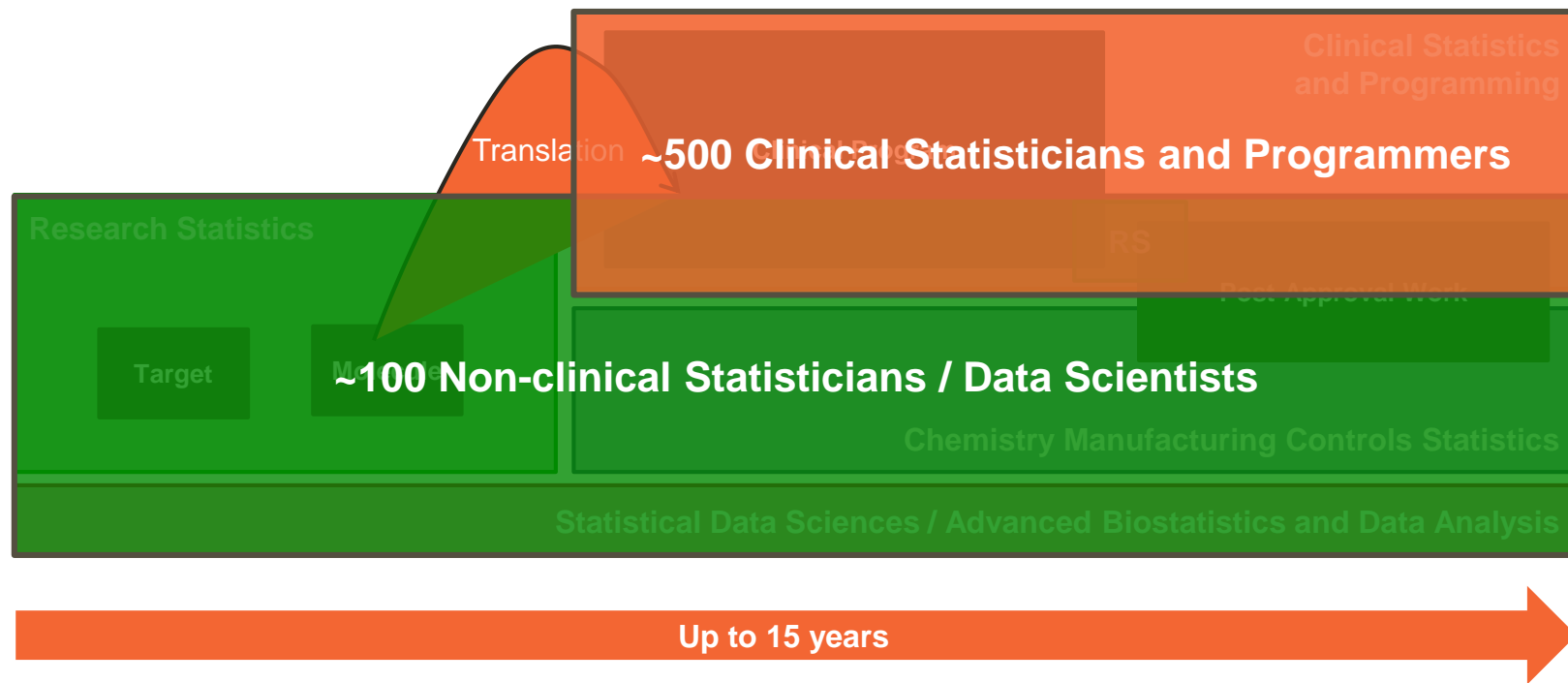
A department of ~600 'Data Scientists'



# The Biostatistics Challenge



A department of ~600 'Data Scientists'



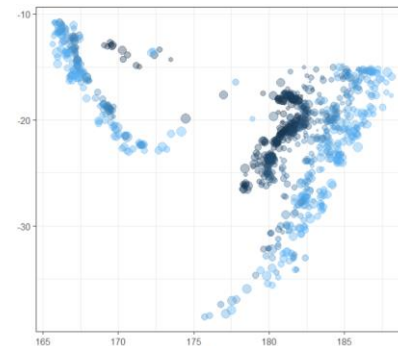
# The Biostatistics Challenge



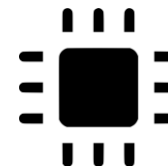
## Clinical Statistics and Programming

- 500 Clinical ‘Data Scientists’
  - ~200 Statisticians
  - ~300 Programmers
- Estimated proficiency – Dec 2018
  - SAS ~ 99%
  - R ~ 10%
  - Python ~ 1%
  - Julia == 0%

Graphics



Trial simulations



# Why not R?



- Existing expertise
- SAS-based tools
- Controlled workflows / platforms built for SAS
- Conservative nature of clinical development
- Common misconceptions around regulations
  - ‘The regulators only accept analyses performed in SAS’

## SAS Transport File Format

- We are able to archive data sets in SAS XPORT transport format, also called version 5 SAS transport format. This is an open, published file format developed by SAS Institute. For additional information, see the SAS Institute's [Standards for Electronic Submissions and SAS XPORT transport format](#) [web site](#).

```
haven::write_xpt(my_data,  
                 "my_file.xpt",  
                 version=5)
```

<https://www.fda.gov/drugs/electronic-regulatory-submission-and-review/electronic-regulatory-submissions-and-review-helpful-links>



[ABOUT US](#) ▾

[BLOG](#)

[REGULATIONS](#)

[VALIDATION](#) ▾

[RESOURCES](#) ▾

[CONTACT US](#)

## WHITE PAPER

PLEASE SEE OUR WHITE PAPER  
TITLED "A STATISTICAL APPROACH  
TO ASSESSING R PACKAGE  
ACCURACY WITHIN A VALIDATED  
INFRASTRUCTURE"

...another time, see  
[www.pharmar.org!](http://www.pharmar.org!)

SAS T

- W
- ve
- by
- fo

<https://>

# Why R now?



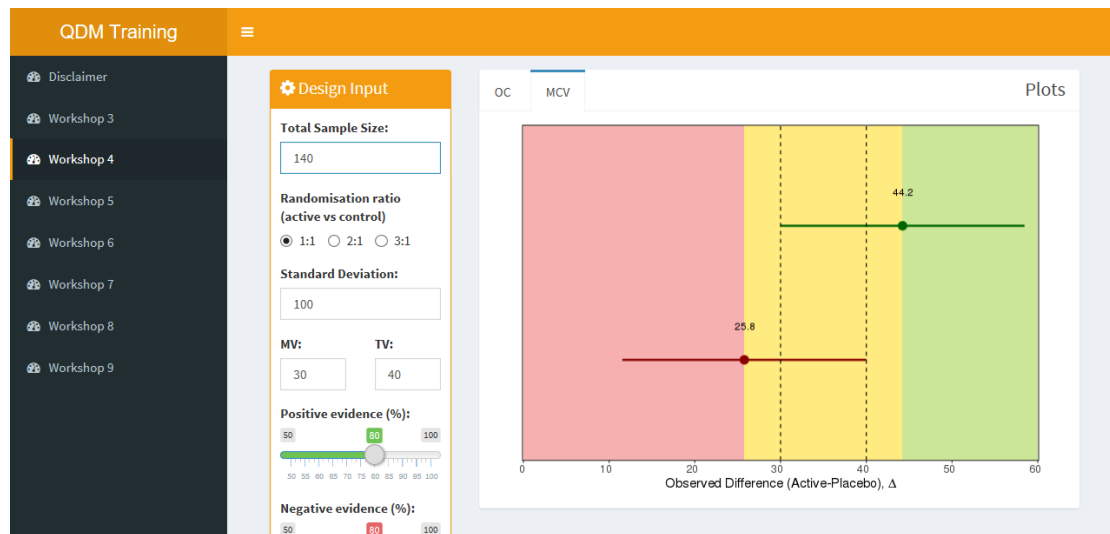
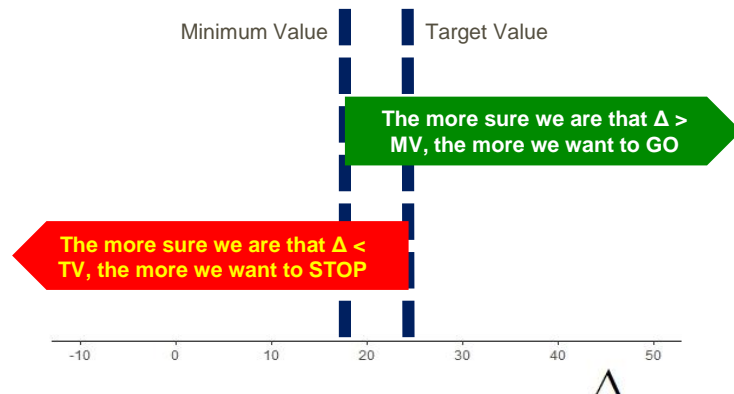
- 
- Influx of new talent skilled in R
  - The drive for innovation
    - Access to latest statistical methodologies
    - Emphasis on simulation
    - Industry trend towards interactive communication / visualisation
    - Productivity gains (automation, dynamic reporting)
  - Cost
  - Gateway to python / open source
    - Interoperability



## GSK Catalysts...

# Quantitative Decision Making

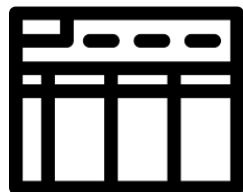
To describe in full the quantitative characteristics of the proposed **study design**, to transparently show the **probability of being able to make a decision at the end of the trial**



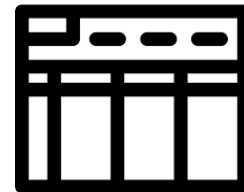
# Dynamic Reporting



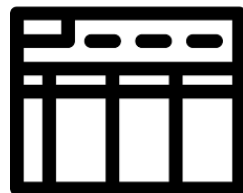
**Biostatistics**



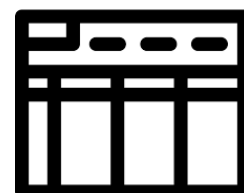
Submissions



Clinical  
Study Report



Presentations



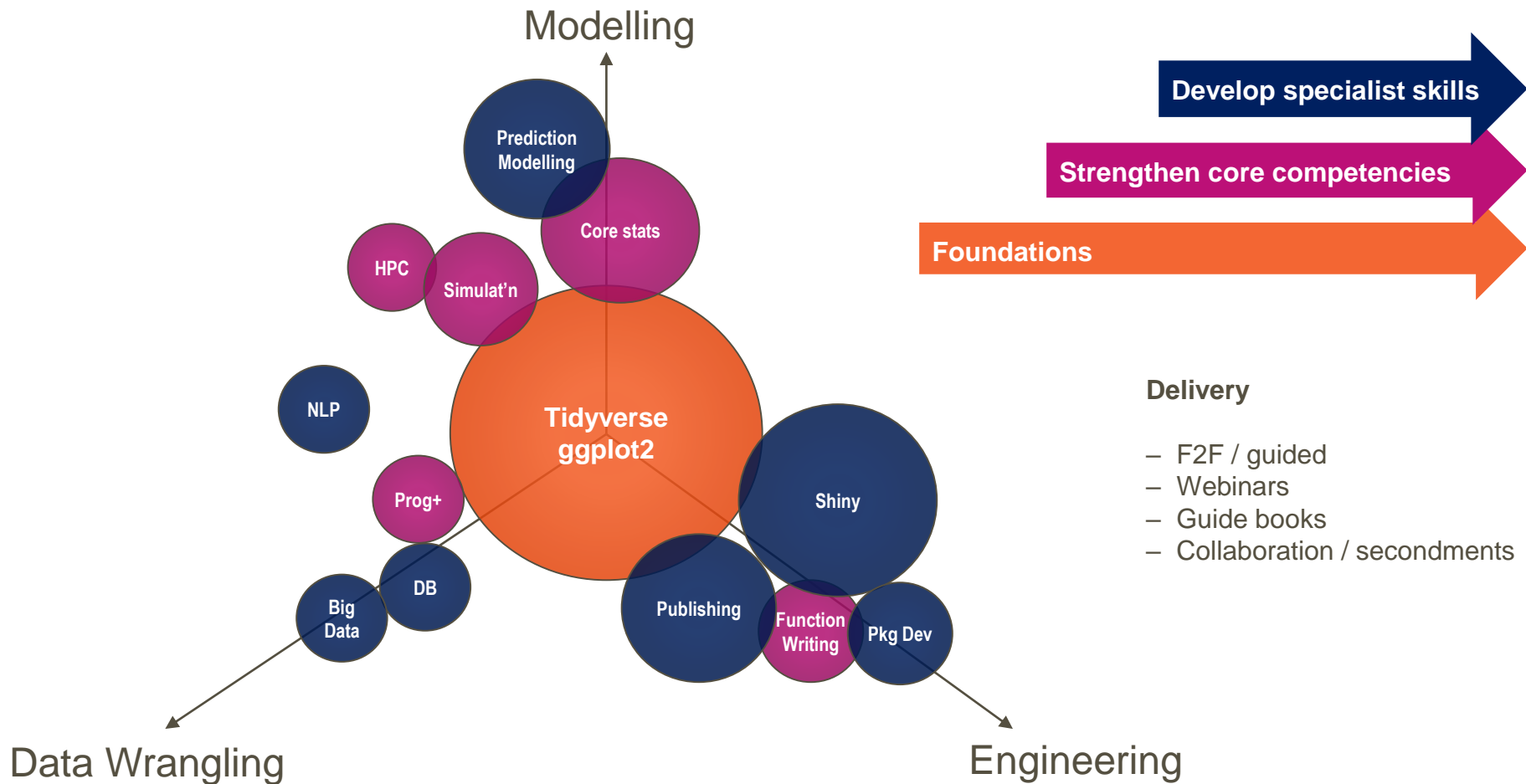
Publications



# Building Capability

- A simple ‘recipe’
  - Training / Support ← Andy
  - Environments ← Tilo
  - On the job learning ← Michael
  - Processes and standards } ← *Not today*
  - Marketing }
  
- Additionally for Pharma...
  - Convince everyone that R is OK to use!

# Training Roadmap



- A simple ‘recipe’
  - Training / Support ← Andy
  - **Environments** ← Tilo
  - On the job learning ← Michael
  - Processes and standards } ← *Not today*
  - Marketing }
- Additionally for Pharma...
  - Convince everyone that R is OK to use!