Three Billboards Inside Basel, Switzerland

Gary Kennedy

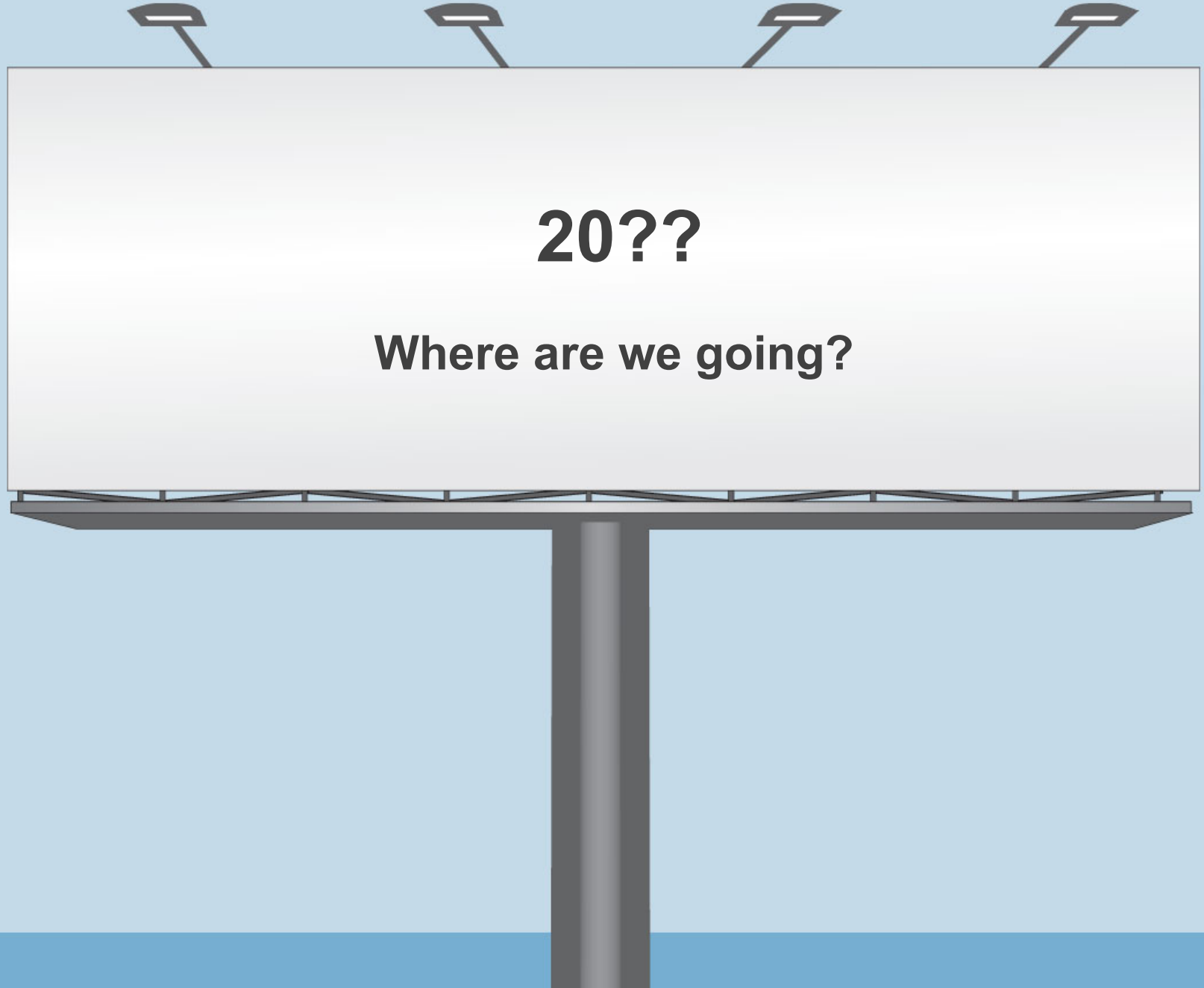
PhUSE One Day Event, Basel

28th June 2018

Shafi Consultancy Ltd.

Agenda

- For the first 10 mins you will still be thinking about how did Switzerland win/lose and Germany win/lose last night (delete as appropriate)….
- The next 10 mins you will have gotten over the excitement or disappointment of winning/losing and your thoughts will be
  - "I hope England lose tonight"
- Last 10 mins there are three scenarios:
  - Why did he bring up the disappointing results from last night in his first slide!
  - I'm so glad he reminded me in his first slide we won, now I can imagine myself scoring the goals
  - Why is he talking about football? I hate football and wish the World Cup was over
- And after all of those thoughts you will awake from your daydream, leaving 5 minutes for questions



Where Have we Come From – 1995?

- No Programming groups worldwide
- All programming responsibility of statisticians
- Statisticians expected to sit SAS exams and get qualified
- We mainly validated our own work

Where Have we Come From?

- Let's look at a study I worked on soon after I arrived and also a more recent one

- Cardiovascular Drug A (over 300 patients)
  - Reported around year 2000

- Cardiovascular Drug B (682 patients)
  - Reported 2013

Drug A - 2000

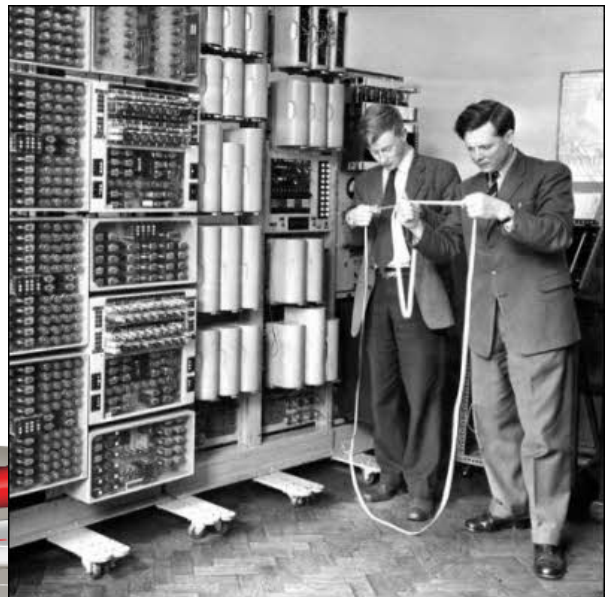
- We created around 6 main analysis datasets V001-V006

V001.SD2	02/04/1997 11:15	SD2 File	161 KB
V002.SD2	27/02/1997 11:48	SD2 File	49 KB
V003.SV2	24/02/1997 14:58	SV2 File	25 KB
V004.SD2	27/02/1997 11:49	SD2 File	105 KB
V005.SD2	27/02/1997 11:49	SD2 File	568 KB
V006.SD2	27/02/1997 11:50	SD2 File	93 KB

- Typical ADS program on next slide

- Biggest program:
  - 15 procedures
  - Approx. 200 lines of code

- Outputs of around 70 tables/figures



Shafi Consultancy Ltd.

```
OPTIONS FMTSEARCH=(Clean.FTXXXXXX);


DATA V00201;
SET TDD.V001;
 BY PTNO ATRCD;
DROP TRTDURA;
 IF ATRCD IN ('D1','A1','B1','C1','BL1');
 IF ATRCD IN ('D1','A1','B1','C1','BL1') THEN DO;
IF ATRCD='A1' THEN ATR='A';
 IF ATRCD='B1' THEN ATR='B';
IF ATRCD='C1' THEN ATR='C';
 IF ATRCD='D1' THEN ATR='D';
  IF ATRCD='BL1' THEN ATR='BL';
  ATR1=ATRCD;
ATRDURA1=TRTDURA;
  END;
 ATTRIB ATRDURA1 LABEL='Treatment Duration First Dose Level (days)';
RUN;


DATA V00204;
 SET CLEAN.DISPENS(WHERE=(discol2=2));
BY PTNO VISNO;
KEEP PTNO VISNO;
IF FIRST.PTNO;
RUN;
```

```
PROC SQL;
 CREATE TABLE V00203 AS
  SELECT V00201.*, V00204.VISNO AS DOSECH
   FROM V00201 LEFT JOIN V00204 ON
      V00201.PTNO = V00204.PTNO
   ORDER BY V00201.PTNO, V00201.ATRCD;

 create table V00204 as
  select v00203.*, date01.actdays as dosechd label='Dose Change (days)',
       date01.nomwk as dosechw label='Dose Change (NOMWK)'
   from v00203 left join tdd.date01 on
      v00203.ptno=date01.ptno and
      v00203.dosech=date01.visno
   order by v00203.ptno, v00203.atrcd;

DATA V00202;
  SET TDD.V001;
   BY PTNO ATRCD;
   KEEP PTNO ATRCD ATR2 ATRDURA2 ATRSPDT;
    ATTRIB ATRDURA2 LABEL='Treatment Duration Second Dose Level (days)';
  IF ATRCD IN ('D11','A11','B11','C11','BL2');
  IF ATRCD IN ('D11','A11','B11','C11','BL2') THEN DO;
   ATR2=ATRCD;
   ATRDURA2=TRTDURA;
  END;
   IF ATRCD='A11' THEN ATRCD='A1';
   IF ATRCD='B11' THEN ATRCD='B1';
   IF ATRCD='C11' THEN ATRCD='C1';
   IF ATRCD='D11' THEN ATRCD='D1';
   IF ATRCD='BL2' THEN ATRCD='BL1';
  RUN;

DATA TDD.V002(COMPRESS=YES);
 MERGE V00204 (IN=A) V00202(KEEP=PTNO ATRCD ATR2 ATRDURA2 ATRSPDT);
 BY PTNO ATRCD;
  IF A;
  TRTDURA=SUM(ATRDURA1, ATRDURA2);
  LABEL ATR    ='Overall Treatment Regimen';
  LABEL ATR1   ='First Dose of Regimen';
  LABEL ATR2   ='Second Dose of Regimen';
  LABEL TRTDURA='Total Treatment Duration (days)';
  LABEL DOSECH ='Visit at Which Dose Changed';
  IF ATR2 NE '' THEN ATRCD=ATR2;
  IF DOSECH=. AND ATR2=' ' THEN DOSECH=8;
 RUN;
```

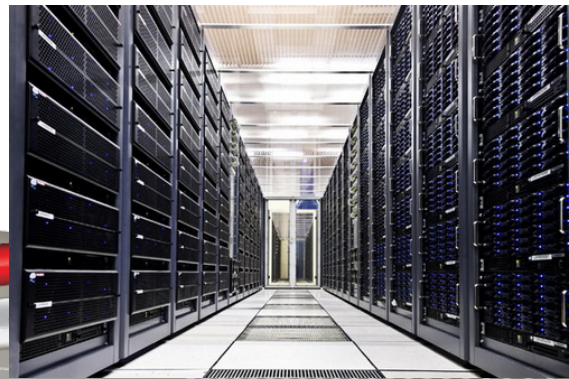
Drug B - 2013

- There are around 50 ADS's and one is 95MB
  - That's more than 160 times the size
  - Program used to create it is 4300 lines of code

- One of the main reporting programs was 2000 lines of code

- The outputs for only the **efficacy**:
  - Just under 2000 pages of tables and figures
  - The supporting statistical outputs were also 2000 pages

Shafi Consultancy Ltd.



What are some fears for programming in 2018?

- Will we have enough work?
- Will we have a job?

- If you look at how the size/quantity of the material required for trials/submissions has expanded since 23 years ago
  - This is one part to answering the question above
  - Will this expansion hit a ceiling? Maybe not

- This will justify the industry as it stands now
  - The programming role has seen massive growth in 23 years
  - Now much bigger groups in most companies than Statistics and Data Management

Will we have enough work?

Health Outcomes

Global warming

Enhancement of graphics

Narratives and safety requests

Amount of data

Translational Medicine

Increase post submission ad-hoc work

Genetic data

Increase in country specific submission standards

Transparency

CT.Gov for EU and in more countries

CDISC

Methodology groups

Epidemiology

Support functions as they lose SAS Skills

Shafi Consultancy Ltd.

Real world data

Key challenges and tools in 2018…..

- So what issues will this vast amount of data from many different sources pose us
  - Big Data
  - Data Lakes: Massive variety of formats (structured/unstructured) and sources
  - Technical Advances:
    - Multiple processing
    - Server farms etc.

- What analytical tools can help us
  - Predictive Analytics: Data Mining, Predictive Modelling, Machine Learning…
    - How do we see patterns in these massive datasets?
  - Do we need the program/software/hardware to adapt itself
    - Artificial Intelligence – 1990 BSc

So what challenges are we facing in 2018 or soon to be asked to solve?

- Transparency
- Data Scientists
- Real World Data
- Risk Based Monitoring
- Digital Trials with no patients
- Standards for data, displays, etc.
  - MDR's (Meta Data Repositories)
  - Programs that write programs
  - Outputs generated by display templates rather than programs
  - Standards agreed within companies and across the industry
- ……

The Known Unknowns - Transparency Initiative

- Studies going back to ???? (depends on company – 1998)
- Can we access the datasets/programs of old studies in SAS V6?
- Do we need to converted to CDISC?
- Need to anonymise all data
  - Remove all comments, centre, date, renumber etc……
- Will companies ask to have access to other companies data to perform their own analyses
  - This is a minefield if we go down this road and an unknown quantity of extra work

How far in the future do we plan for?

- We tend to think of temporary solutions to problems rather than those 10-20 years down the line
- Take Risk Based Monitoring
  - Originally introduced to solve the problem of 100% SDV
  - So less SDV now……yes
  - Look at centres with high risk factors early to prevent issues before they happen
  - Introduce regular ongoing central monitoring during the trial
    - Some companies already did this some didn't
    - Now a more formal process and requested by FDA
  - Some nice applications to help with the front end of this process

- Overall have we reduced the amount of resources required?
- Overall have we reduced the cost?
- Overall have we improved the quality?

So what if we take ourselves to 20 years from now

- Maybe…..

- Many trials with no patients and the data comes from a variety of GP databases
- No SDV as the data itself is the SDV
- No cleaning as we know it now, as the data is what it is
- The future for analytics is…
  - How do you analyse data you can't clean?
  - How do combine data from completely different databases?
  - How do you measure the quality of the trial and the validity of the conclusions you reach?

- Surely this is an impossible task that no other industry has had to face?

So what if we take ourselves to 20 years from now

- Or is it…
- Doesn't it seem kind of antiquated to go back and forth to sites and ask

  "Did you really mean that?"

  "I need a complete date?"

  "The patient must complete this outcome even if they withdraw from the trial"
- Do Google, Facebook, LinkedIn, Twitter……
- Clean their data and change it?

  "PhUSE is the biggest worldwide electricians organisation"…… fake news

  "Are you sure Gary is rated as a 0 for looks?"

- Do we build in methods to ensure the data is as clean as possible
- And analyse in a way that our conclusions are robust to the quality of the data we collect
- Do we need any SDV or any cleaning
  - RBM was a valid solution to a problem that was raised
  - But did we go far enough and think far enough ahead

Conclusions

- So data cleaning and SDV is just one example
- Do we apply solutions to the challenges we encounter now, or…….
- Think further ahead to anticipate the challenges yet to come
- As an industry we are normally reactive in nature to the regulators
- What if we steered the regulators towards the processes needed for the future
  - In a risk averse industry (for good reasons!) this is not so easy