
Biostatistics into the future....

Chris Holmes



Disruption

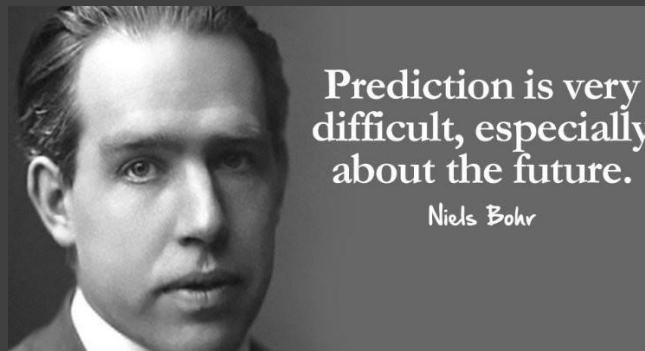
- We are living through a period of dramatic innovation in the biomedical and pharmaceutical sciences
 - akin to a 4th industrial revolution
 - pace and scale of developments in multiple areas that impact our field, including
 - AI, genetics, wearables, routine capture of eHealth records, imaging, health-relevant data
 - Biobanks, observational studies, data lakes
 - Regulation
- The ability to measure is no longer the principle bottleneck

Disruption

- The challenge comes from how to capture, store, assimilate, and make sense of data from multiple sources
 - This will be the major focus going forward
 - And design better studies that maximise the opportunities to learn from rapidly available heterogenous data
- As leaders an overarching challenge is to create an environment and culture for analytic teams that can keep pace with the developments.....

Where we might be heading

- Forecasting is a dangerous game
 - like driving a car looking out the back window
 - you tend to crash
- However, with that caveat....



Some thoughts on where we're heading

- This prediction will be wrong (with probability 1)
- Biostats will be integrated as a core discipline within biomedical Data Science
- Data will be streaming (wearables), highly multi-modal, longitudinal
- Everyone will have their genome sequenced (at / pre birth)
- There will be a shift towards well-being, early intervention, preventative medicine and balancing multiple long-term conditions
- **Causal inference will be mainstream**
- **Simulation will increasingly be used as an inference tool**
- **Move towards real-world continuous (Bayesian) adaptive clinical trials**

Biostatistics – death or glory?

- Biostats will remain a core discipline within biomedical data science
- Increasing demand for statistical skills in “data fusion”, multi-modal data integration, simulation (of outcomes, patient trajectories)
- Cultural change towards statisticians working in small teams alongside data wranglers, research engineers, AI experts, MLops
- We’re way off “automated statisticians”
 - pandemic response showed AI “missing in action” highlights the importance of information outside the data
- Not implausible that highly repetitive analyses, such as arising from SAPs in RCTs, will become fully automated
 - c.f. AlphaCode <https://www.deepmind.com/blog/competitive-programming-with-alphacode>



Biostatistics – core values – our USP

- Statistics provides the formal language and rigour for the communication and interpretation of scientific findings from data
 - The fundamental properties of sampling variability don't go away just because we use AI
 - They become even more important
 - “*Those that cannot remember Statistics are condemned to repeat it*” -
- adapted from G. Santayana
- Statisticians are the guardians of reproducible research and stability of findings – accurate uncertainty quantification
 - These are essential skills
 - Of ever increasing importance

Biostatistics – core values – our USP

- Much of the art of applied statistics and the skills of a trained statistician involve factors that cannot be captured by algorithms or are poorly covered in current ML training and practice
- These include
 - the careful design of experiments,
 - a thorough understanding of the research questions in the context of an existing knowledge base,
 - accounting for ascertainment and selection bias in data,
 - adjusting for censoring and truncation mechanisms,
 - the tailoring of models to the research question, and a healthy suspicion around results that look too good to be true,
 - careful model checking.

Biostatistics – core values – our USP

- Statisticians have strong self-selection for a feel for data
 - difficult to coach – not strongly selected for in ML
- That said, we all benefit by moving to a common ground
 - ML   Statistics
 - More rigour More agility
- Create teams with a blend of skills, shared appreciation of cultures and domain strengths – this has the potential for a super exciting future!
 - E.g. Turing data labs

The Future we need to prepare for...

- Data will be streaming (wearables), highly multi-modal, longitudinal, irregular
- Well-being will be monitored, real-time updating from wearables and dashboards of personal health
- World data doubling time every two years – pharma relevant data probably more so
 - more data generated in the last two years than in the entire human history
 - E.g. we may see clinical trial recruits given a smart watch (or app??), or some tech that provides real-time data capture (and nudging?)

Future -- preparing for it

- Everyone will have their genome sequenced (at / pre birth)
 - Our Future Health – 5m prospective cohort now recruiting (target 10% representative of UK adult population by 2025)
 - \$600 genomes – trend forecast from 2002 = 1c genome in 2042
 - Pipelines and analysts will be needed to integrate this into workflows
- There will be a shift towards well-being, early intervention, preventative medicine and balancing multiple long-term conditions
 - Ageing society will be the number one health challenge (in the West)
 - Biomedical companies will respond to this
 - This will impact on analytics and study designs (comorbidities)
 - Resilience to possible disruption: e.g. impact of a general blood screening tool for cancer?

Future of analytics

- Causal inference will be mainstream
 - We should be teaching this now in the context of real-world evidence and RCTs
 - E.g. methods such as “Target trial emulation” will be common place
- **Simulation as an inference tool** – and core tool/skill for analyst teams
 - Insilico trials complementing real patient data
 - Digital twins coupled with wearables
 - AB testing and refinement of trial recruitment strategies
 - Data fusion for predicting outcomes
- Real-world continuous (Bayesian) large scale adaptive clinical trials
 - Nudging for better outcomes, re-designing for better learning, utilising reinforcement learning from AI

Other possible disruptors

- More agile regulation
 - Governments critically aware of the need for change
- Increasing patient / public empowerment and patient owned recruitment
 - <https://waroncancer.com/clinical-trial-finder/>
 - <https://www.meet.nye.health/>
 - What's the current worldwide number of people using pharma products? Compare this with the number of people currently being monitored and actively providing information to pharma
 - That ratio needs to change

Summary

“Continual learning from everyone”

Delivered through talented agile teams with mixed skills (including biostatistics!)



Thank you!