The Wisdom of the Crowd

Leveraging advanced analytics at scale

Dr James Black
Associate Director, Personalised Healthcare Data Science, Roche/Genentech
james.black.jb2@roche.com

Topics For Today

Inner-sourcing advanced analytics (AA)

Advancements in data, analytics and technology

Bake-offs with external AA companies

Building an internal network

Hackathons

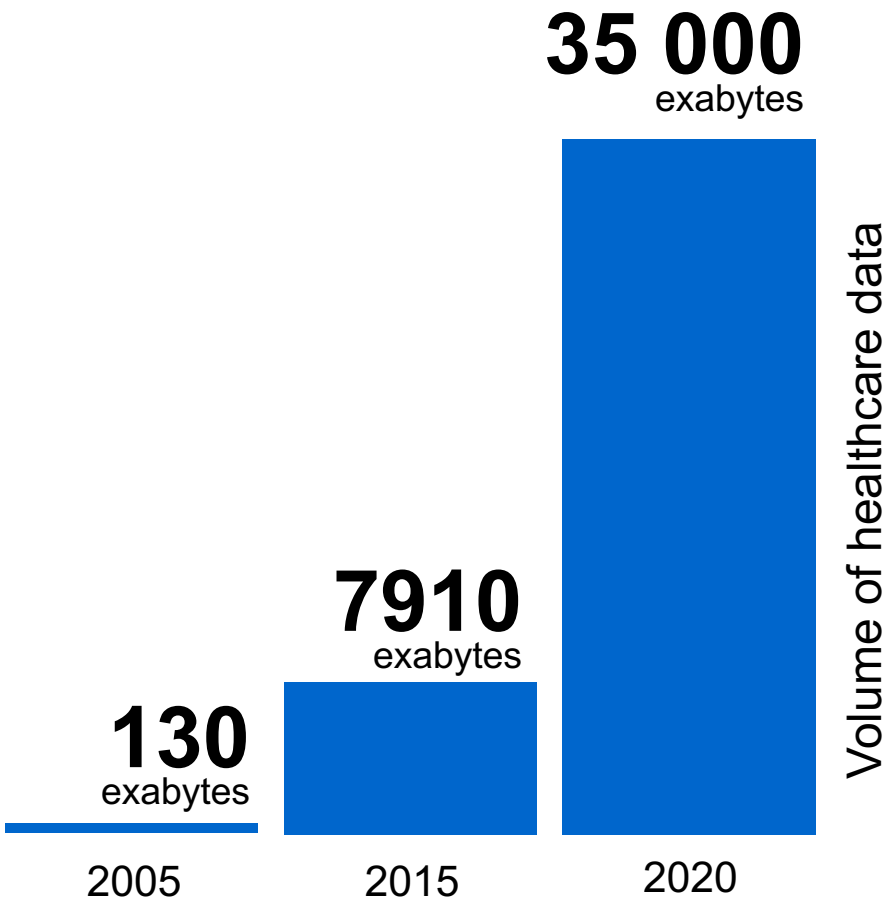
Final Reflections

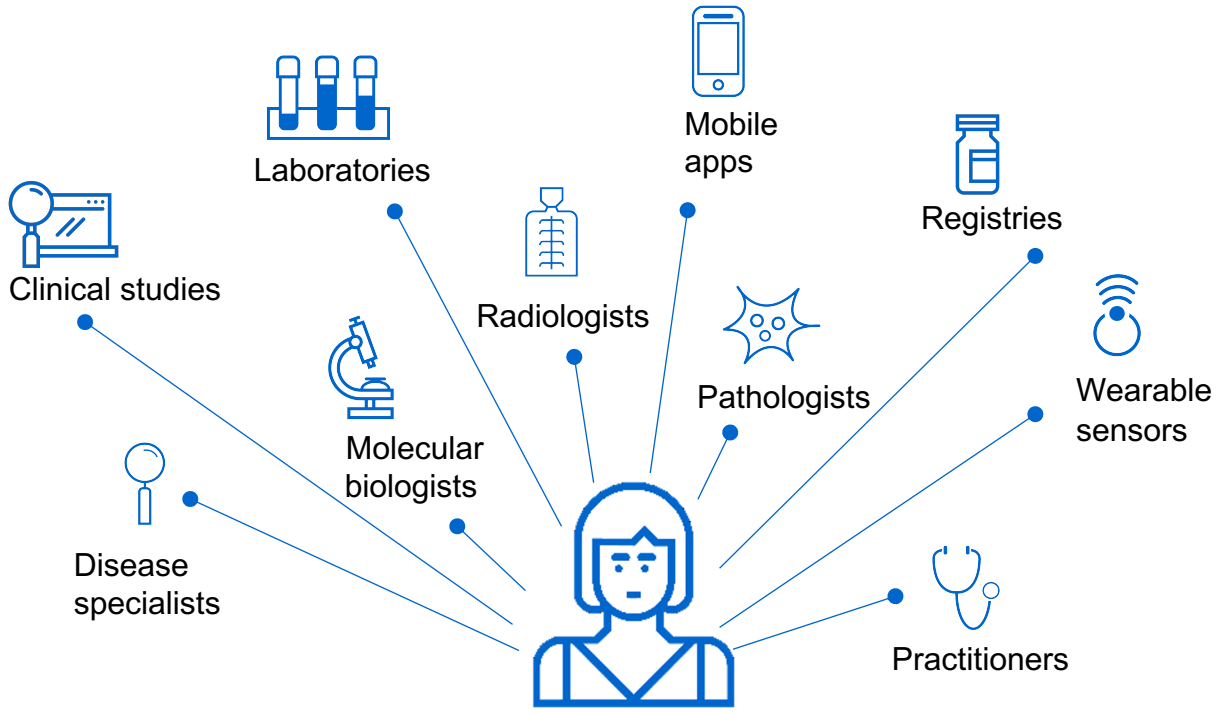
Traditional evidence generation ‘multi-modal’ oncology dataset

Patient	Age	Hx	Smoker	ECOG	KRAS	Squamous	Days to censor	Had progression
Patient 1	79	HBP	5+ a day	4	TRUE	TRUE	180	TRUE
Patient 2	57	Type 2 Diabetes, HBP	No	3	FALSE	FALSE	203	FALSE

We now have Meaningful Data at Scale (MDAS)

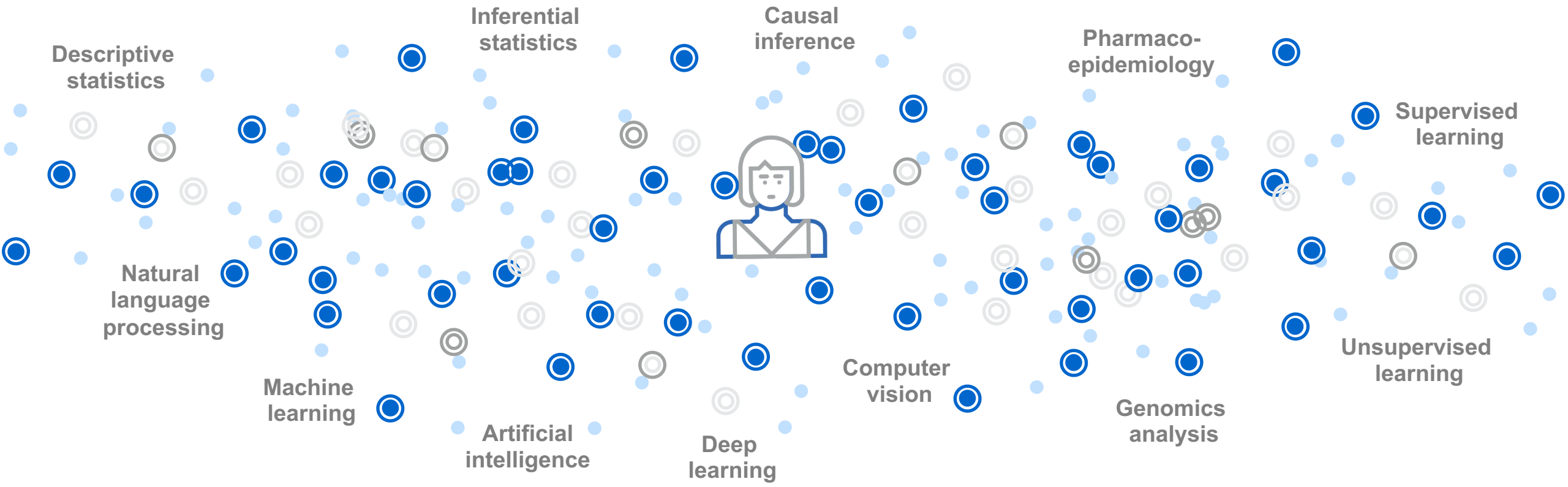
This means deeper (e.g. raw genomics and images) and more broad (e.g. more data modalities) patient datasets.





These data are necessitating and facilitating advanced analytic approaches

A diverse set of disciplines, often contained in both department and mindset silos, is needed to leverage this data



Example 1: predicting outcomes from clinical trial data (Project Turing)

Understanding which patient characteristics drive patient outcomes will be vital in providing truly personalised healthcare

The goal

Evaluate and compare approaches to predicting patient outcomes in DLBCL

Prediction of OS, PFS and treatment response using baseline clinical and genomic data

Compare the performance of externally-developed, proprietary machine learning models with Roche's open source-based approaches

The data

Cleaned de-identified clinical trial data on DLBCL from a Phase 3 trial that included > 1400 patients

DLBCL, diffuse large B-cell lymphoma; OS, overall survival; PFS, progression-free survival

Motivating example: Bake off of proprietary vs internal teams predicting outcomes from clinical trial data (Project Turing)

The results

Outcome	Performance metric	Single biological feature	Open-source ML models with internal biological knowledge	Proprietary ML models	
				Model 1	Model 2
OS	Weighted C-index	0.56	**0.66**	0.62	0.60
PFS	Weighted C-index	0.64	**0.67**	0.45	0.64
ORR	AUC	0.68	**0.70**	0.68	0.66

Key learnings

Combining machine learning and biological understanding within-Roche generated the best predictions

Proprietary models can be consistently outperformed by internal cross-functional teams

AUC, area under the curve; ML, machine learning; ORR, overall relative risk; OS, overall survival; PFS, progression-free survival

Selecting the right analytical approach for a specific question is key

A deep understanding of the question, the data and biological context should inform the approach

Data modalities and the nature of the research question **should determine the type of approach selected**

A thorough understanding of the biology being modelled is critical for any approach

Roche has an industry-leading internal network of more than 900 data scientists across the globe with specific expertise in data and advanced analytics to solve key challenges in healthcare

So what if we focussed first on our internal expertise?

Roche PHC Strategy: AA Capability Building

OVERALL PURPOSE

To strengthen Roche's Advanced Analytics capabilities

METHOD TO ACHIEVE

Roche Advanced Analytics Network

Internal & Roche-wide focus

To strengthen Roche's advanced analytics (AA) capabilities across functions and sites and to become an industry leader in AA, bringing impact to patients.

We want to achieve this by:

1. *Connecting and empowering the Roche AA community*
2. *Fostering knowledge sharing and developing our AA expertise across the Roche internal organization*
3. *Impacting our research, business and patients by empowering our AA and DS teams to create further insights from data*

Academic Collaborations

External & ad-hoc focus

To rapidly evaluate and integrate emerging AA methodologies & approaches into our work through collaborations with leading academic institutions.

We want to achieve this by:

1. *Advancing research and creating impact by leveraging academic sites at the forefront of AA*
2. *Strengthening internal AA capabilities by consulting with academic experts on key projects/initiatives*

Roche has built a global network of data scientists and advanced analytics practitioners

The Roche Advanced Analytics Network (RAAN) is an enterprise-wide global network across functions and sites, established in **2017**

The expertise

Experts with diverse experience from pharma, diagnostics, IT, early R&D and non-medical applications

The people

> 900 members, spread over **> 40** sites.
<1% of Roche's employees

The goal

Strengthen Roche's capabilities to become an industry-leader in advanced analytics to solve important questions in healthcare

Community	Knowledge sharing	Data challenges	Advisory group	Interns	Tech

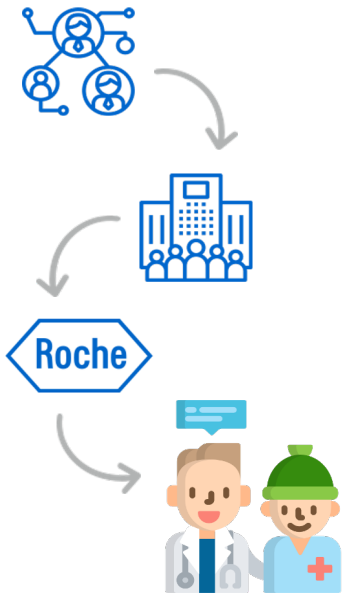
R&D, research and development

Building a network of Data Scientists at Roche proving to be impactful

Example: The Roche Advanced Analytics Network (RAAN)

The RAAN was established in 2017 and we now have over 850 members from Pharma, Dia, gRED, pRED and Group Finance/IT across >40 Roche sites.

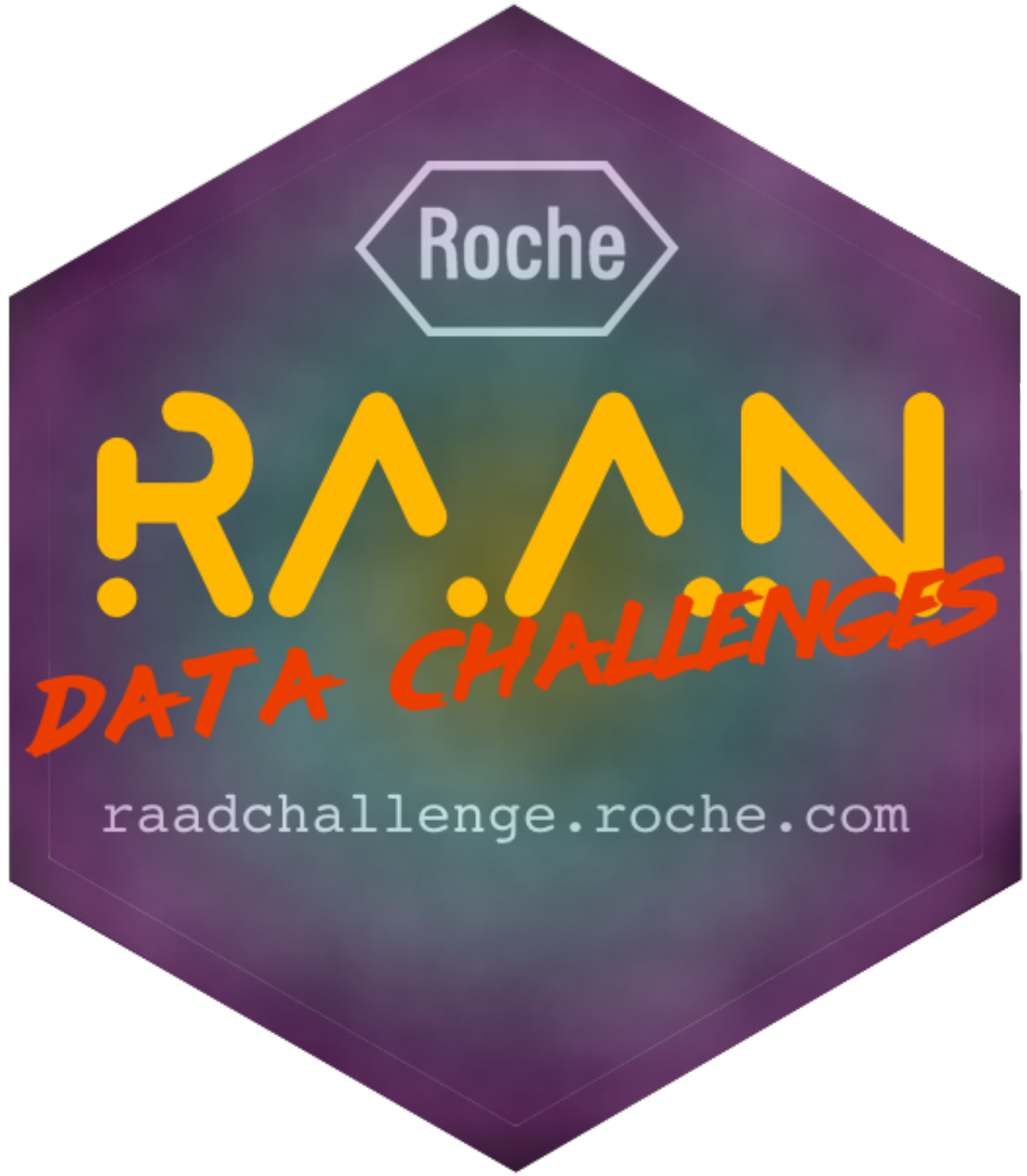
Our Goal is to strengthen Roche's advanced analytics (AA) capabilities across functions and sites and to become an industry leader in AA, bringing impact to patients.



We want to achieve this by:

1. **Connecting** and empowering an advanced analytics community
2. Fostering **knowledge sharing** and **developing our AA expertise** across the organization
3. **Impacting our research, business and patients** by creating insights from data

Community	• Developed RAAN website & newsletter • 30+ AA communities mapped • Engaging champions across sites/functions
Knowledge Sharing	• RAAN Day EU: 300 attendees, 60+ posters • RAAN Day North America: 200 attendees • Sponsoring trainings, requests for more
Data Challenges	• Support hackathons • Piloted common knowledge share platform & AA toolkit
Advisory Group	• Recruited steering committee and experts • Actively working on multiple projects • 2 mixers to link stakeholders and experts
Interns	• HR & UTAS to build external AA brand • Supporting 15 interns across functions • Implementing AA solutions to business
Tech	• Evaluating existing tech landscapes • Identifying existing tech solutions that could be rolled out to the RAAN community


Roche
RAAN
DATA CHALLENGES
raadchallenge.roche.com

2018 challenge: predicting mortality risk using electronic health records (RAAD)

Can training on a real-world, pan-tumour dataset benefit ultimate predictions?

Roche Advanced Analytics Data Challenge

The goal

Predict the probability that a patient will be alive at 1 year after treatment initiation, using all the patient data available up to the start of treatment

The data

7000 patients' electronic health records were used across seven cancer types to build a model

Average patient had **~190** data points pre-treatment

3500 different patients used to test the models

500 Roche employees

132 teams

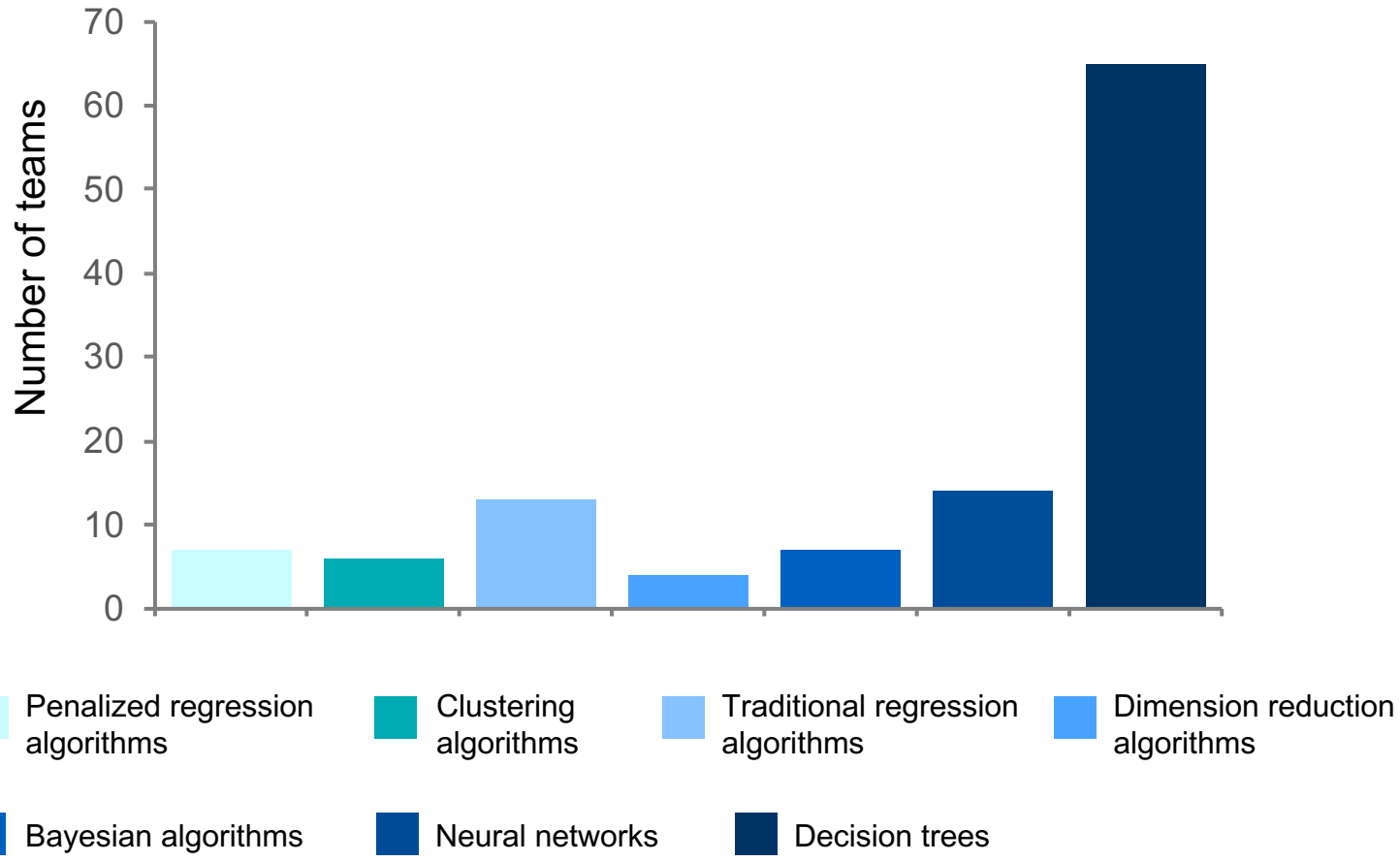
28 Roche sites

165 departments

RAAD, Roche Advanced Analytics Data challenge; RAAN, Roche Advanced Analytics Network

Example 4: predicting mortality risk from electronic health records (RAAD)

Key learning: Non-parametric, decision tree-based algorithms were an effective choice for this type of research question



Key learnings

The top 5 teams used XGBoost (in ensembles), but there was wide variety in the way in which models were applied

High-density risk scores offered an improvement over traditional measures

In predictive models for mortality risk using real-world data, consider changes over time up until baseline

Full use of EHR data for predictive models was more successful than using known predictors only

CRAN Package XGBoost: https//cran/r-project.org/web/packages/xgboost/index.html
RAAD, Roche Advanced Analytics Data challenge; RAAN, Roche Advanced Analytics Network

2019 challenge: Identifying likely responders to treatment using integrated clinical trial data (RAAD)

Predicting treatment response in oncology is a key step in individualizing care

Roche Advanced Analytics Data Challenge

The goal

Use advanced analytics to develop a prediction model to identify patients with non-small-cell lung cancer (NSCLC) at first line who are most likely to respond to Tecentriq® treatment versus standard of care

The data

Training dataset

10 curated clinical trials

~5000 patients

Test dataset

1 clinical trial with minimal indication overlap

~1000 patients

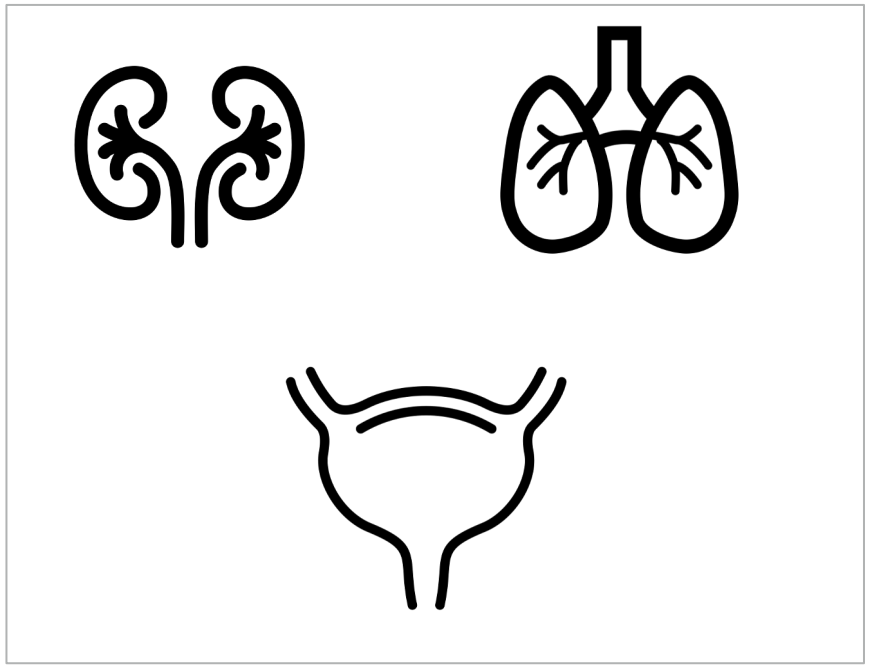
Clinical and omics data provided

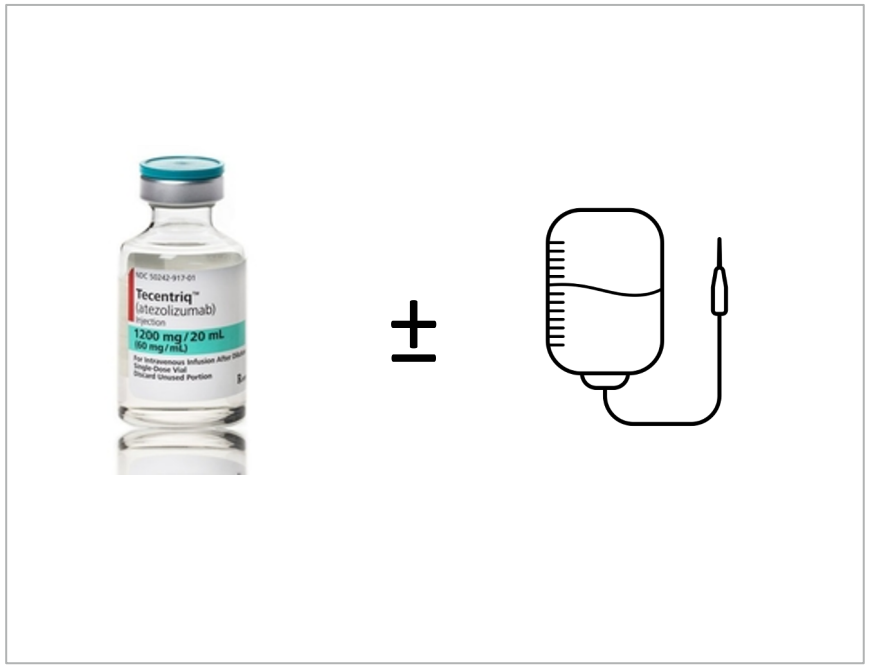
517 participants

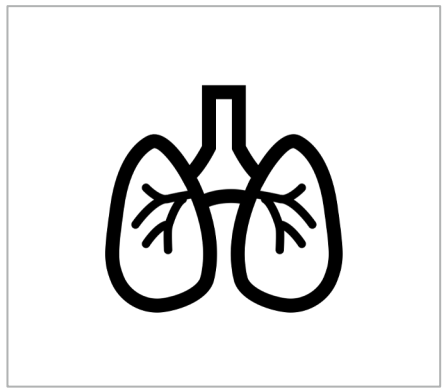
141 teams

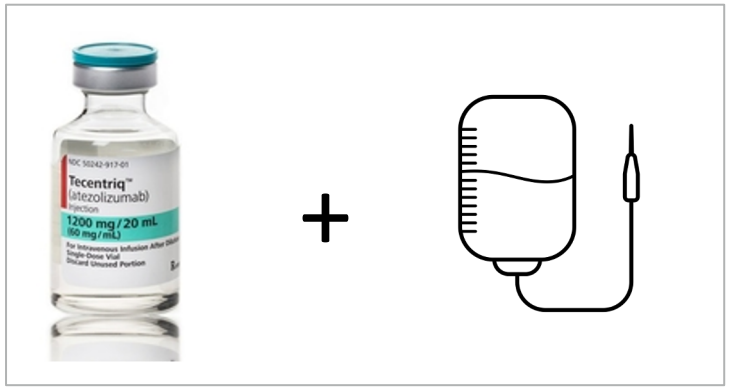
38 Roche sites

RAAD, Roche Advanced Analytics Data challenge; RAAN, Roche Advanced Analytics Network

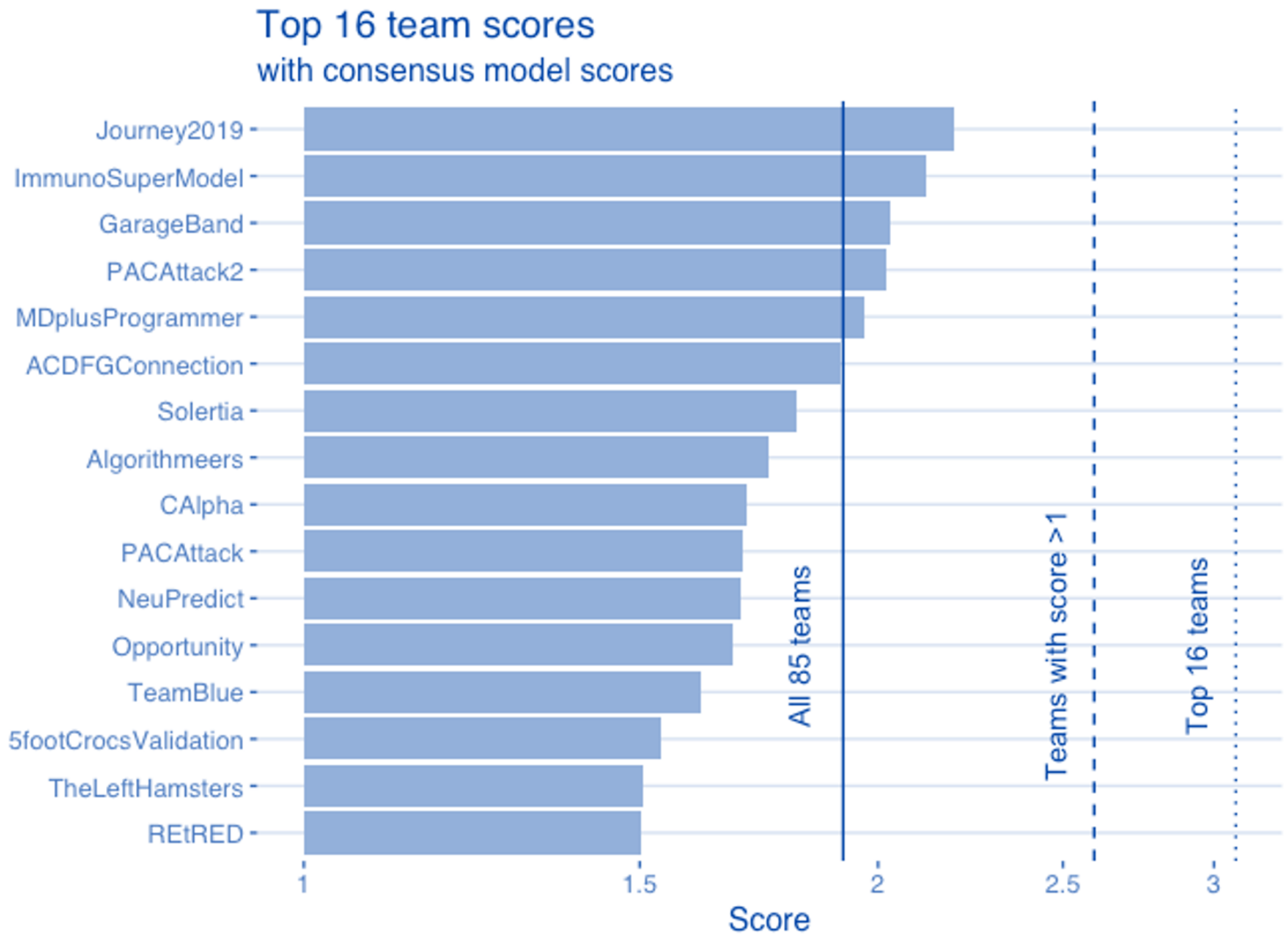



Tecentriq™
(atezolizumab)
Injection
1200 mg / 20 mL
(60 mg / mL)



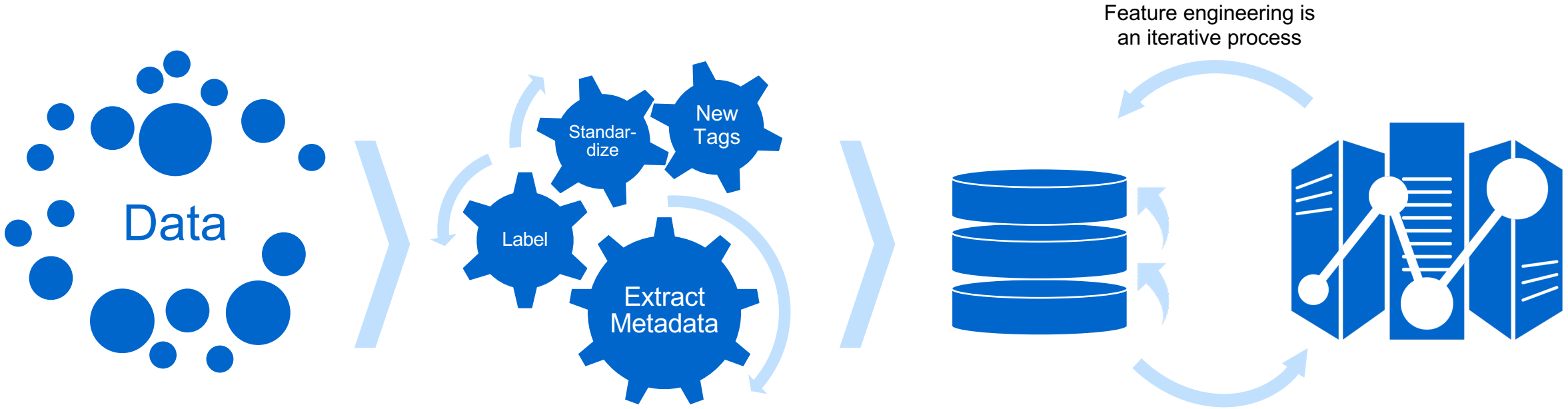

Tecentriq™
(atezolizumab)
Injection
1200 mg / 20 mL
(60 mg / mL)

A consensus model outperformed beyond expected biases



Making sense of complex healthcare data requires a considered approach

Significant time is invested in expert curation and preparation of the data before advanced analytics can be applied



Acquisition

High-quality data must be systematically collected[1,2]

Curation

Relevant data must be carefully categorized, linked and curated[1,2]

Feature engineering & Advanced analytics

Data must be transformed based on understanding of the question, biology and analytic approach to create relevant features. Analytics must be carefully designed to address specific research questions

1. Kruse CS *et al. JMIR Med Inform* 2016;4:e38; 2. Anagnostopoulos I *et al. J Supercomput* 2016;72:1494–1516

Final reflections

Collate the right data

We must continue to invest in meaningful data, data quality, data linkages and data at scale. Without the right data, even the best analytics cannot provide meaningful answers

Ask the right questions, using the right approach

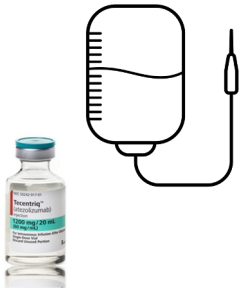
Our approach is centred around designing a fit-for-purpose analytic approach based on an understanding of the data and the healthcare question at hand

Bring together the right people

Quantitative science and advanced analytics has been an essential component of Pharma for decades, but that expertise can be silo'd onto specific data and questions. Enabling true cross-functional collaboration brings a breadth of knowledge on the data and science that can often outperform organisations billed as leader in AI and AA.

Internal and external collaborations with academia can further enable teams, and has the added benefit of enriching, rather than consuming, company resources.

Roche

Odds  / Odds 

Odds  / Odds