Turning science into medicines

Implementing Statistical and Data Science Innovative Approaches in Drug Development

26 Oct 2021

Gary Cline, Early Biometrics & Statistical Innovation, Data Science and Artificial Intelligence, R&D, AstraZeneca, Gaithersburg, US
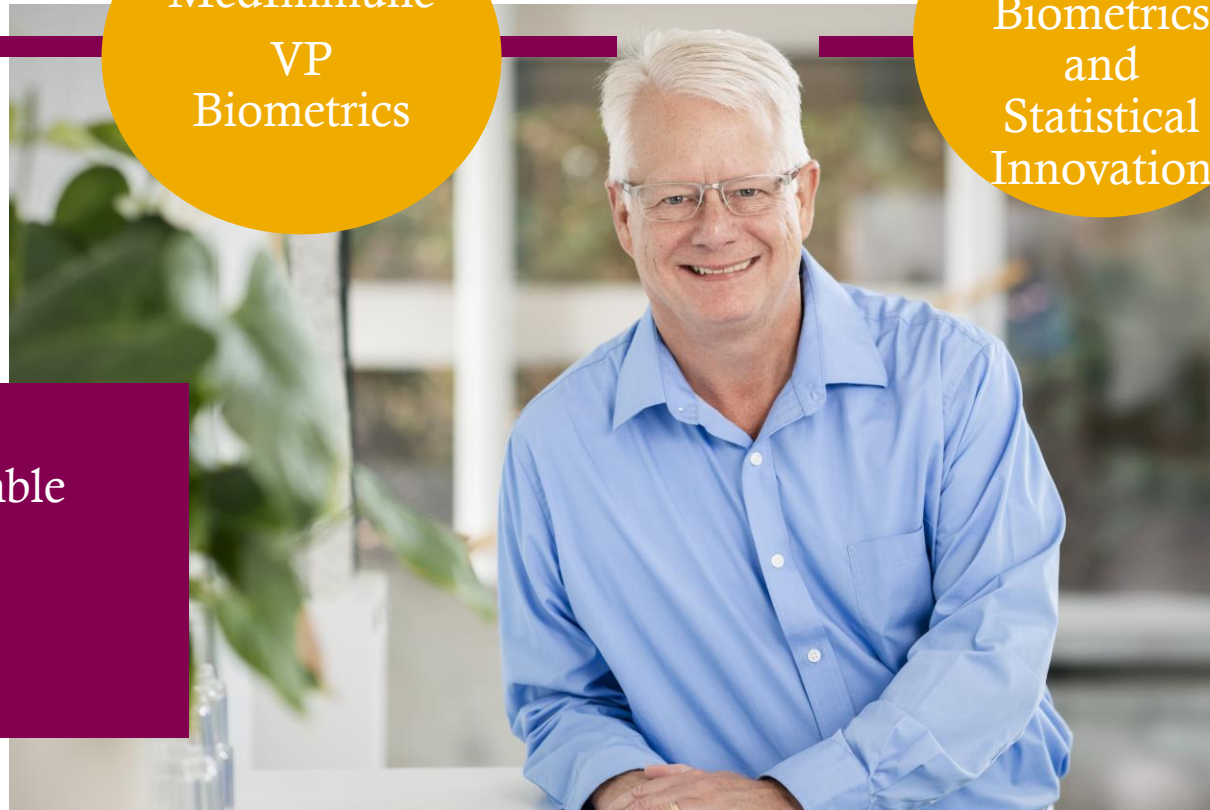
PhD, Statistics, University of Kentucky

MS, Math, Eastern Kentucky University

ClinTrials
Procter & Gamble
ICON
Celgene
Receptos

2018 MedImmune VP Biometrics

2020 AZ VP Early Biometrics and Statistical Innovation
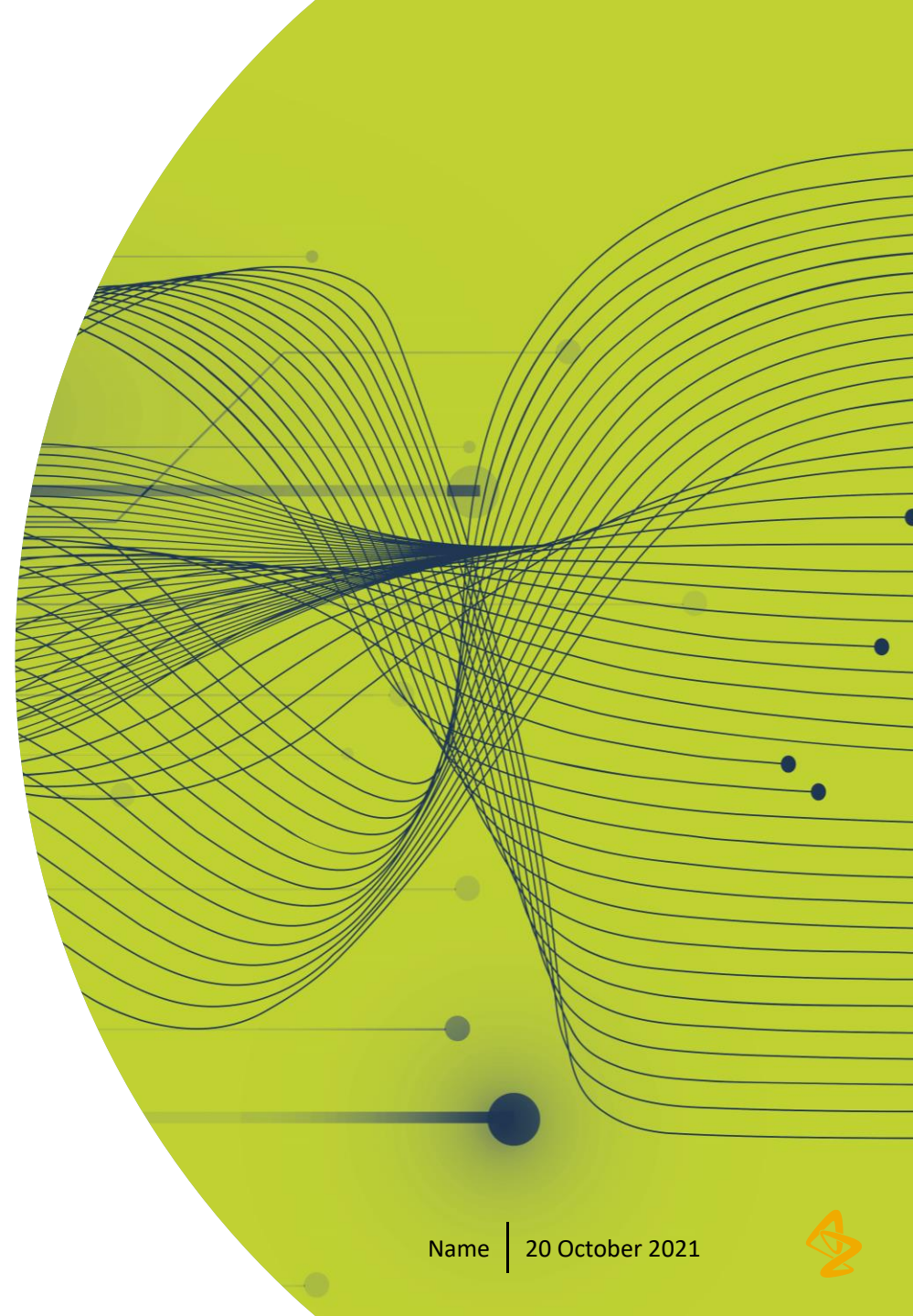
Agenda

1 Why we innovate

2 Data Science and AI & Biometrics

3 We follow the science – how AI is accelerating research

4 Expanding the scope with Biometrics

Why Innovate?

We push the boundaries of science to deliver life-changing medicines

Inspired by our **values** and **what science can do**, we are **focused** on **accelerating the delivery of life-changing medicines** that create enduring **value for patients and society.**

The 5R Framework

Right target
Uncover, select and validate new targets with a strong link to disease

Right tissue
Ensure that new drug candidates have good bioavailability and display the right effect in the intended tissue

Right safety
Establish safety as far as possible in humanised systems before initiating clinical trials

Right patient
Recognise that patients have unique, genetic, molecular and functional disease profiles, and target medicines to populations who will derive the greatest benefit

Right commercial
Develop a unique value proposition for new medicines based on the size and unmet needs of the target patient population

We are focused on three specific Data and AI strategic priorities

Optimising our development programmes

By delivering savings, going faster, improving probability of success

Accelerating our research

AI-driven targets & molecules, optimising candidate drug decisions
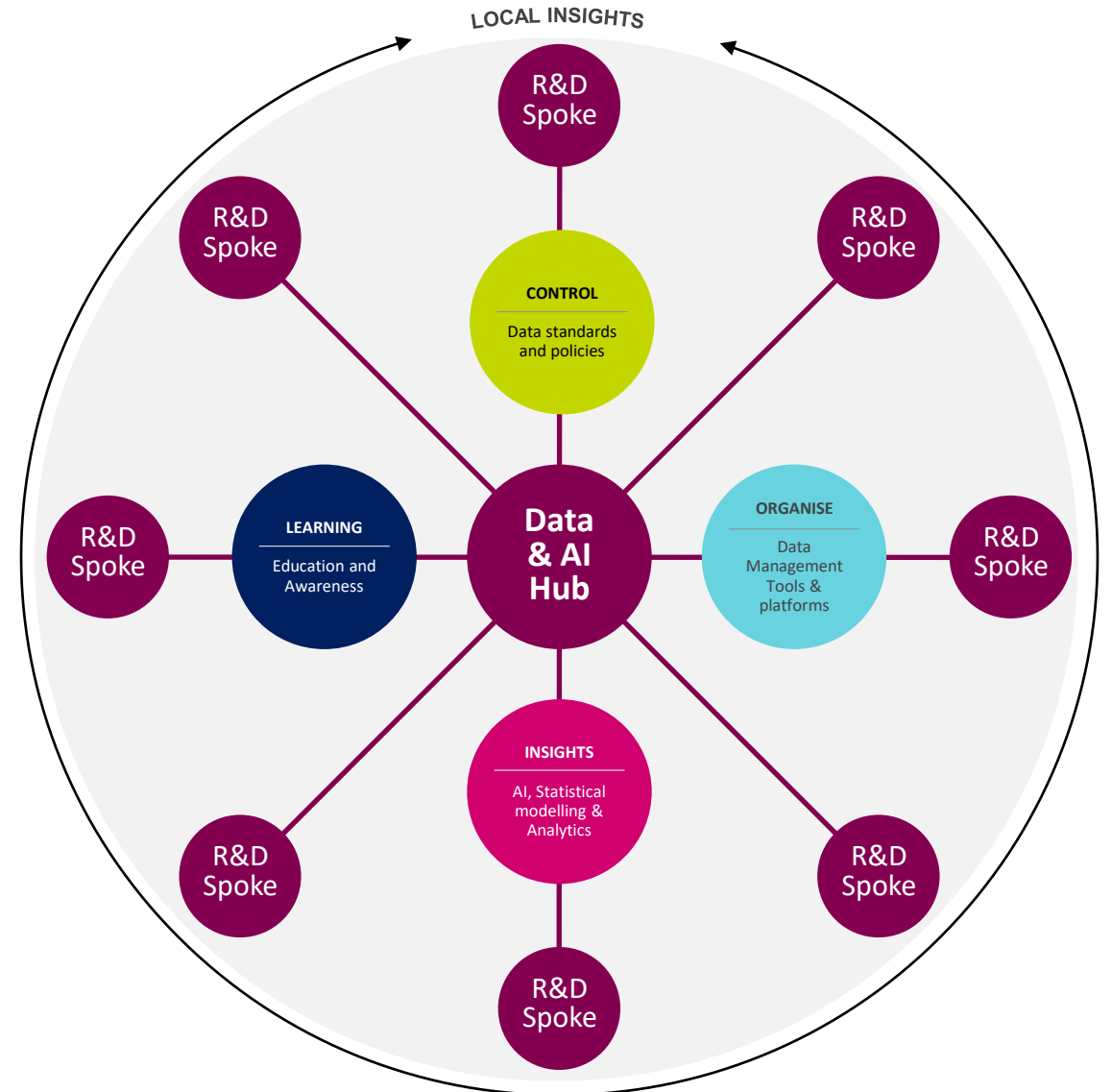
Transforming our science

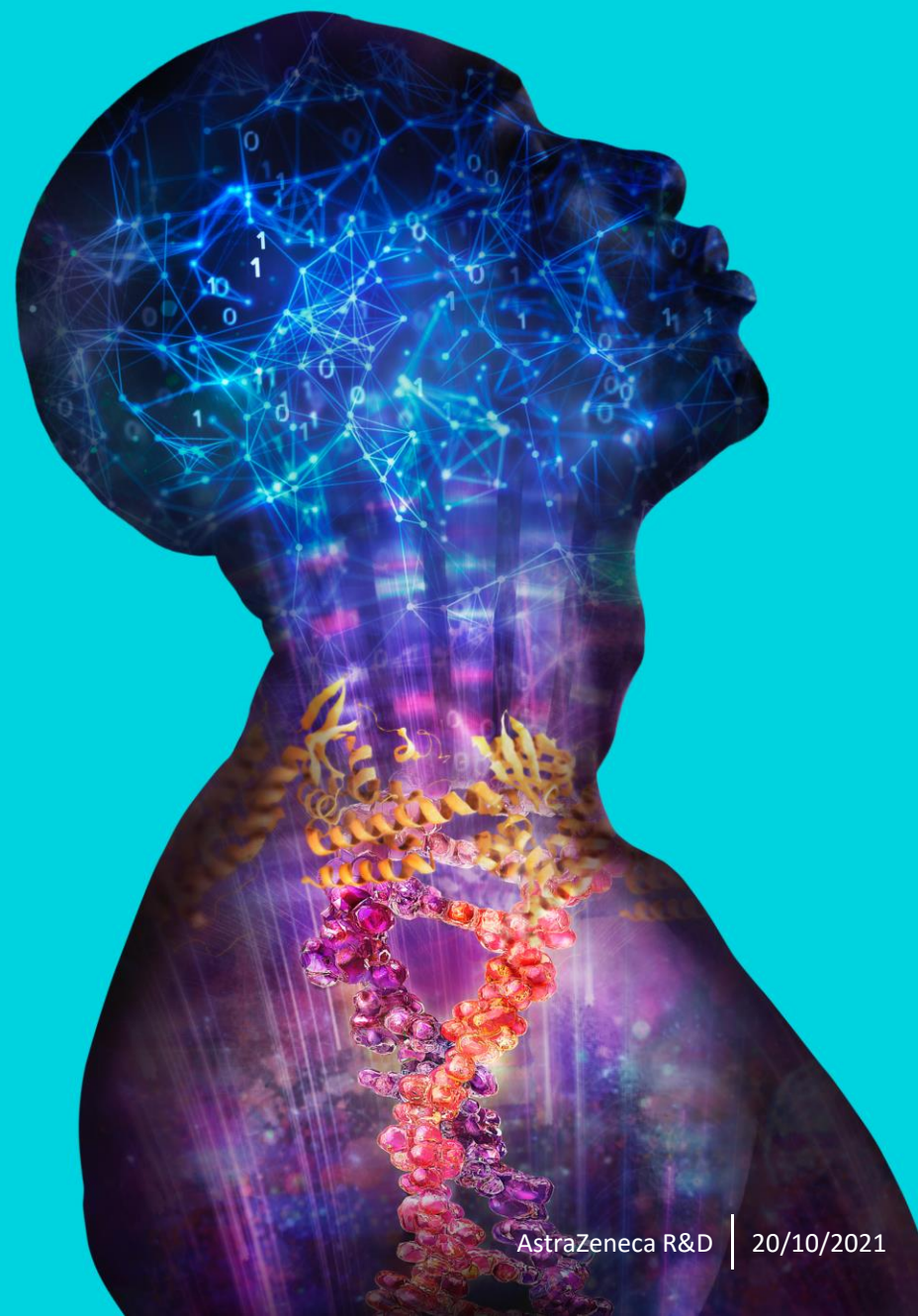*New methods for diagnosis and outcome prediction*

Our R&D Hub-and-Spoke model

We are organised in an R&D Hub-and-Spoke model combining central Data & AI enabling capabilities with TA and function-specific insights.

- One stop shop for COIL™ services and capabilities

- Transfer solutions and methods between spokes

- Central experts to consult across spoke teams

- Critical mass of talent for key skills – hub role makes this uniquely possible

- Demand for hub services growing over time as spoke teams evolve and grow

- DSAIL[†] in place to align priorities across R&D

LOCAL INSIGHTS

R&D Spoke

CONTROL
Data standards and policies

R&D Spoke

R&D Spoke

R&D Spoke

LEARNING
Education and Awareness

Data & AI Hub

ORGANISE
Data Management Tools & platforms

R&D Spoke

R&D Spoke

INSIGHTS
AI, Statistical modelling & Analytics

R&D Spoke

R&D Spoke

R&D Spoke

[†]DSAIL = R&D **D**ata **S**cience & **AI L**eaders group

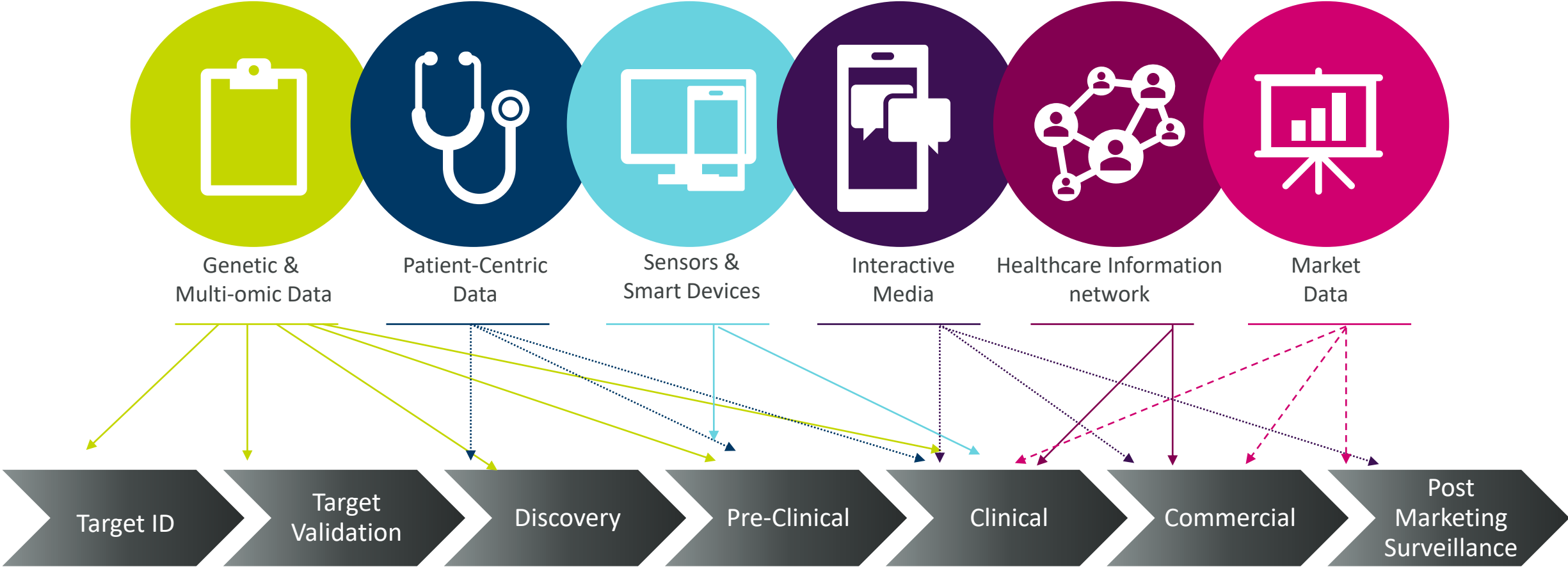# Data Science and AI & Biometrics

Why now is the right time for AI in healthcare

"AI will not replace drug hunters, but drug hunters who don't use AI will be replaced by those who do."
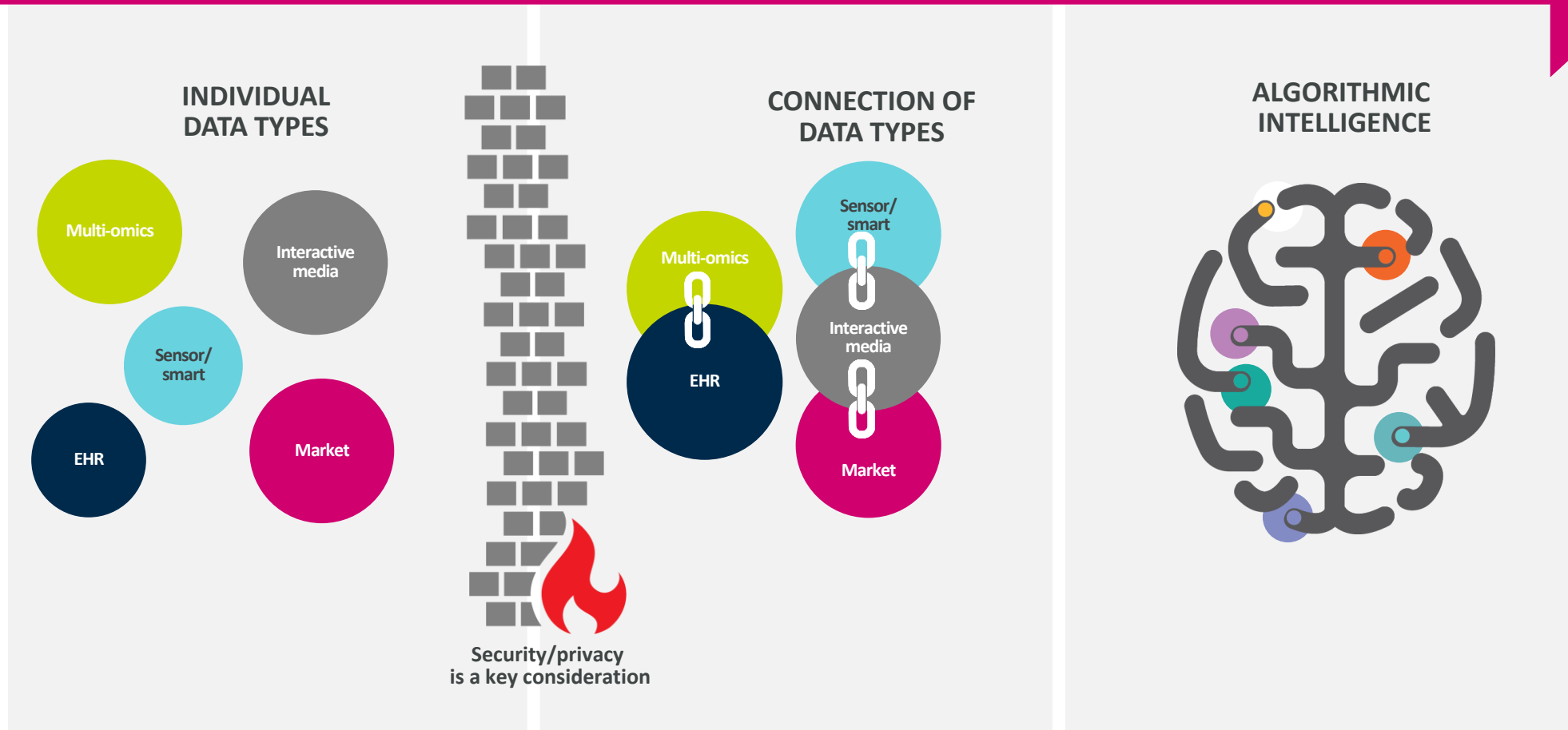
Access to more data than ever before

Genetic & Multi-omic Data

Patient-Centric Data

Sensors & Smart Devices

Interactive Media

Healthcare Information network

Market Data

Target ID

Target Validation

Discovery

Pre-Clinical

Clinical

Commercial

Post Marketing Surveillance

The way we analyse data is changing.

Connected data allows us to unleash the power of AI

✓ **Faster and more accurate**
   ✓ **Deeper and more sophisticated scientific insights...**
                                ...into patients, medicines & disease

INDIVIDUAL DATA TYPES

Multi-omics

Interactive media

Sensor/ smart

EHR

Market

Security/privacy is a key consideration

CONNECTION OF DATA TYPES

Multi-omics

Sensor/ smart

EHR

Interactive media

Market

ALGORITHMIC INTELLIGENCE

Our E2E approach also delivers **faster** and more **accurate** results *via* a **deeper** and more sophisticated scientific insights in **patients**, medicines & disease
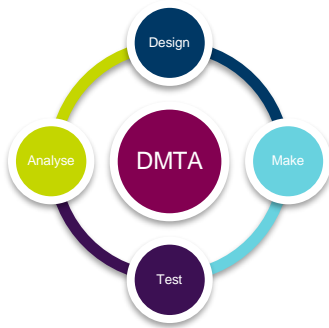

1. Disease classification and prognosis

(*Multiomics*)
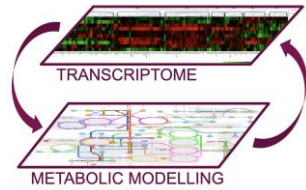
2. Disease understanding

(*AI & knowledge Graphs*)
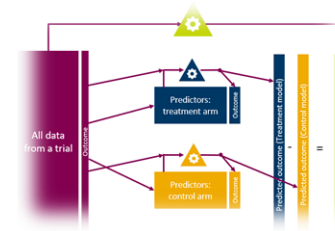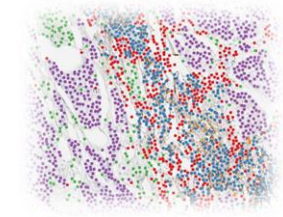
3. Drug Design & Synthesis
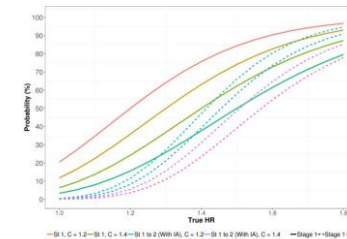
(*AI for drug design*)

Speed

Accuracy

4. Personalized Medicines

(*Virtual Twins*)

5. Digital Pathology

(*Imaging*)

6. Predict treatment response

(*Enhanced clinical trial design*)

7. Best Drug Delivery

(*product Development*)

8. Digital Health

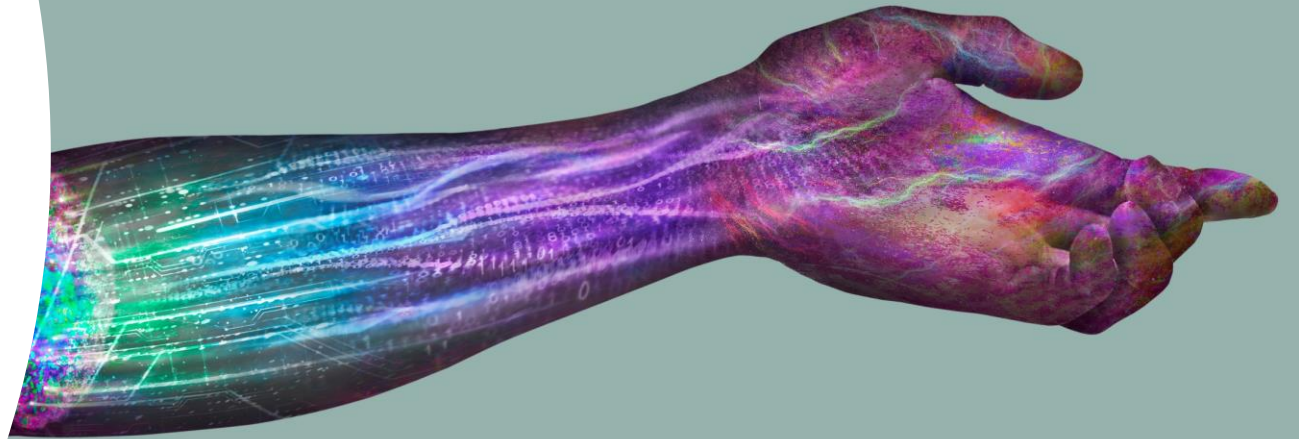| 1	2	3	5	4	6	7	8	2
Target Discovery		Drug Discovery	Pre-clinical Science	Clinical Trials, Product Development Regulatory & Safety				*Launch/post-launch*

We follow the science –
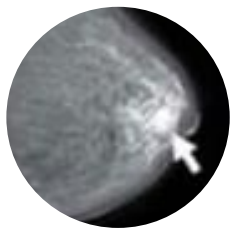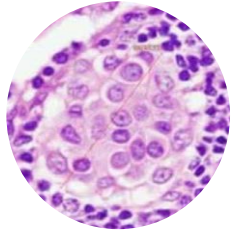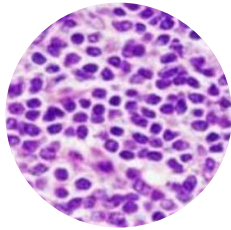how AI is accelerating
research

AI is already transforming the diagnosis of cancer from images…

Benign   **Cancer**

Photographs

e.g. AI can diagnose skin cancer from photographs with a competence comparable to a trained dermatologist

Histopathological images

e.g. AI can identify lymph node metastases in tissue sections with a performance that is comparable to an expert pathologist
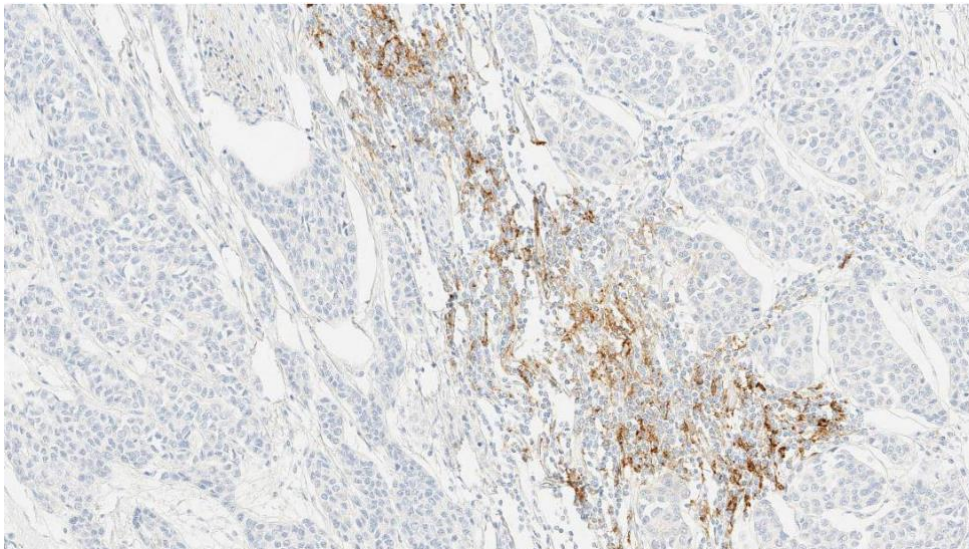
Radiological images

e.g. AI can distinguish cancerous from non-cancerous lesions in mammographic scans of the breast

Slide credit: Andrew Reynolds

…and radically changing pathology

Human assessment of PD-L1 staining

Complexity	+++
Training	years
Time	20 min
Error rate	10-20%

AI-based assessment of PD-L1 staining
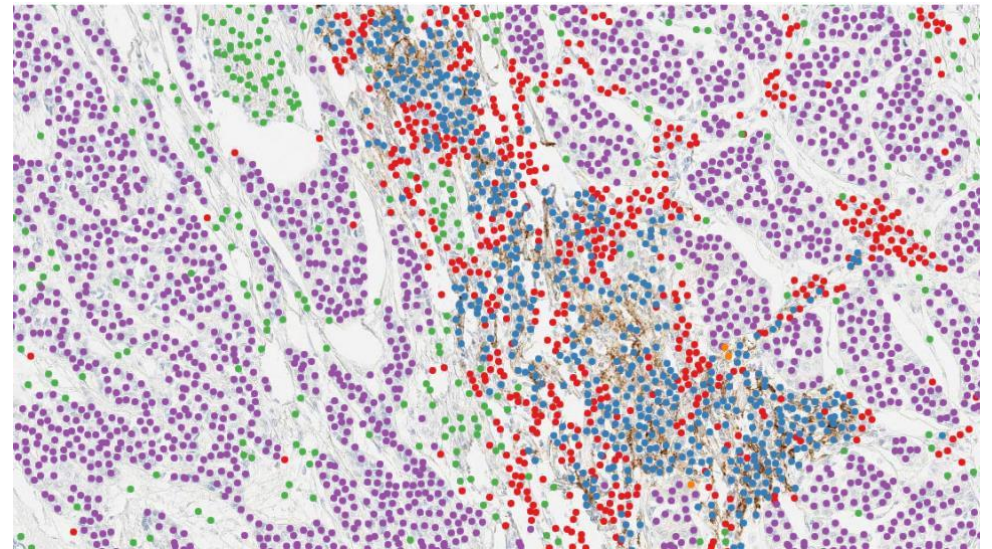
- 🟧 Tumour cell positive
- 🟦 Immune cell positive
- 🟩 Fibroblast
- 🟪 Tumour cell negative
- 🟥 Immune cell negative

Complexity	+++
Training	days
Time	seconds
Error rate	0.65%

How AI can enhance diagnosis and disease understanding

Retinal imaging can predict multiple cardiovascular risk factors

- AZ Healthy Route aims to screen millions of people for cardiovascular risk in developing countries

- Retinal images taken by smartphones are being considered as a screening tool

Input image | Age | HbA1c | Diastolic BP | BMI

Lower estimate ← Effect magnitude → Higher estimate
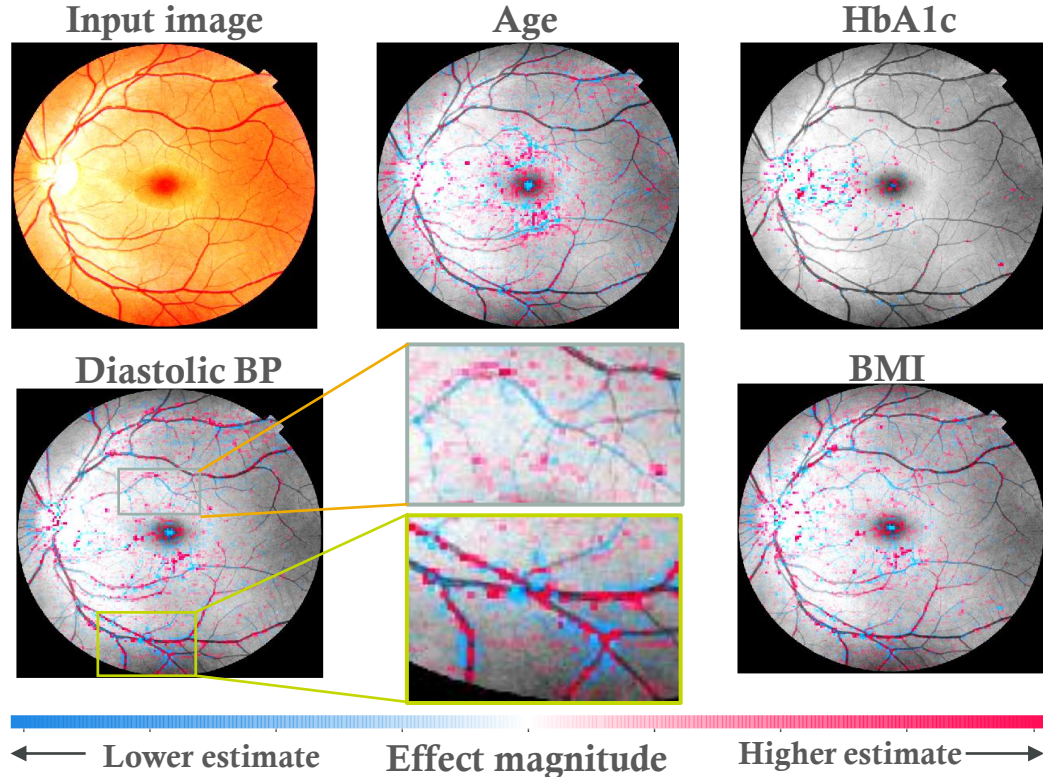
Machine Learning analysis of health records could replace liver biopsies for diagnosis of NASH

- This is an invasive procedure and poses a crippling challenge to clinical care and trial recruitment, with screen failure rates being prohibitively high

NASH Model Feature Importance & Directionality

Feature	Importance rank	Direction of association with NASH	Inflection point*
Factor A	1	⬆	0.7
Factor B	2	⬆	40.8
Factor C	3	⬆	3.3
Factor D	4	⬆	43.5
Factor E	5	⬆	12.7
Factor F	6	⬆	7.3
Factor G	7	⬇	165.9
Factor H	8	⬆	39.6
Factor I	9	⬆	103.7
Factor J	10	⬇	241.7

Features anonymised pending future publication / IP assessment

Our ambition:
Creating the Google map of cancer
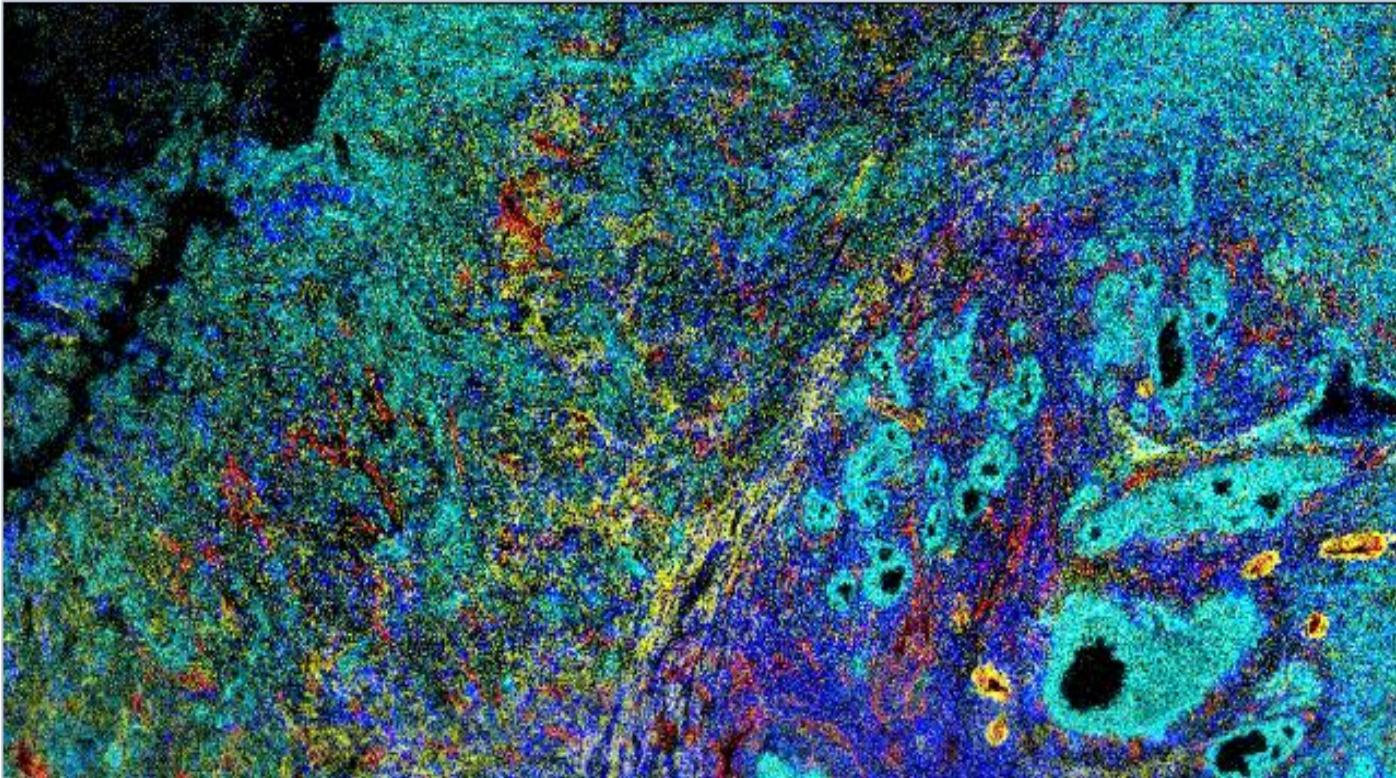
A Complete Cartography of Cancer
Through Multiscale Molecular Imaging

CANCER RESEARCH UK GRAND CHALLENGE ROSETTA TEAM

$20 million funding to link imaging from subcellular to surgery

NPL National Physical Laboratory | Imperial College London | CANCER RESEARCH UK | BEATSON INSTITUTE | ICR The Institute of Cancer Research

Barts Cancer Institute Queen Mary University of London | AstraZeneca | UNIVERSITY OF CAMBRIDGE | THE FRANCIS CRICK INSTITUTE

AI knowledge graphs with billions of data points are driving our understanding of disease

Pharmacology

Chemistry

Biomarkers

Data sources

Genomics

Clinical trials

Multiomics

Images

Text - EHR, literature, patents

Data Types

Video

Audio

Expanding the scope with Biometrics

Position in Development Program

|	IB	IIA	IIB	III	
Phase 2					High P(Phase 3 Success)
Phase 2/3					Confirmatory evidence
Early Phase					Proof of concept
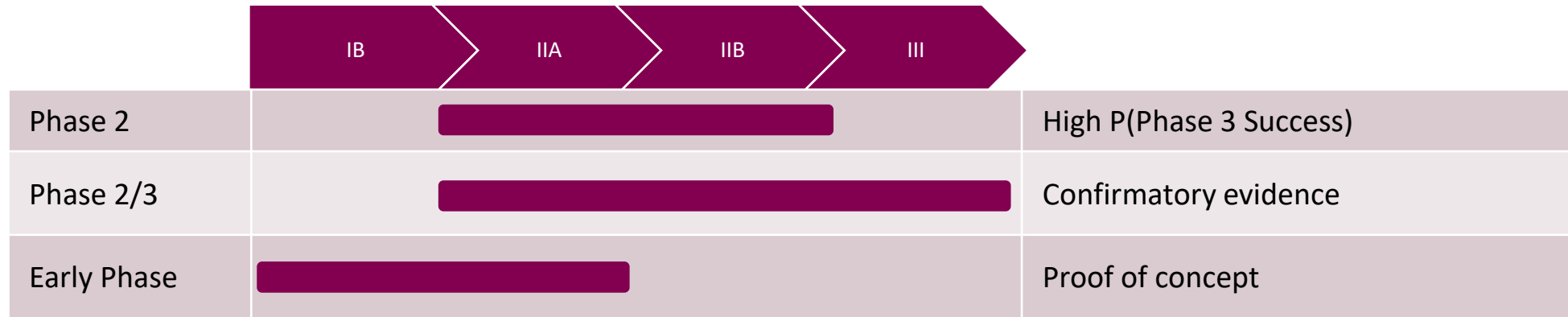
Drug Project Operating Model
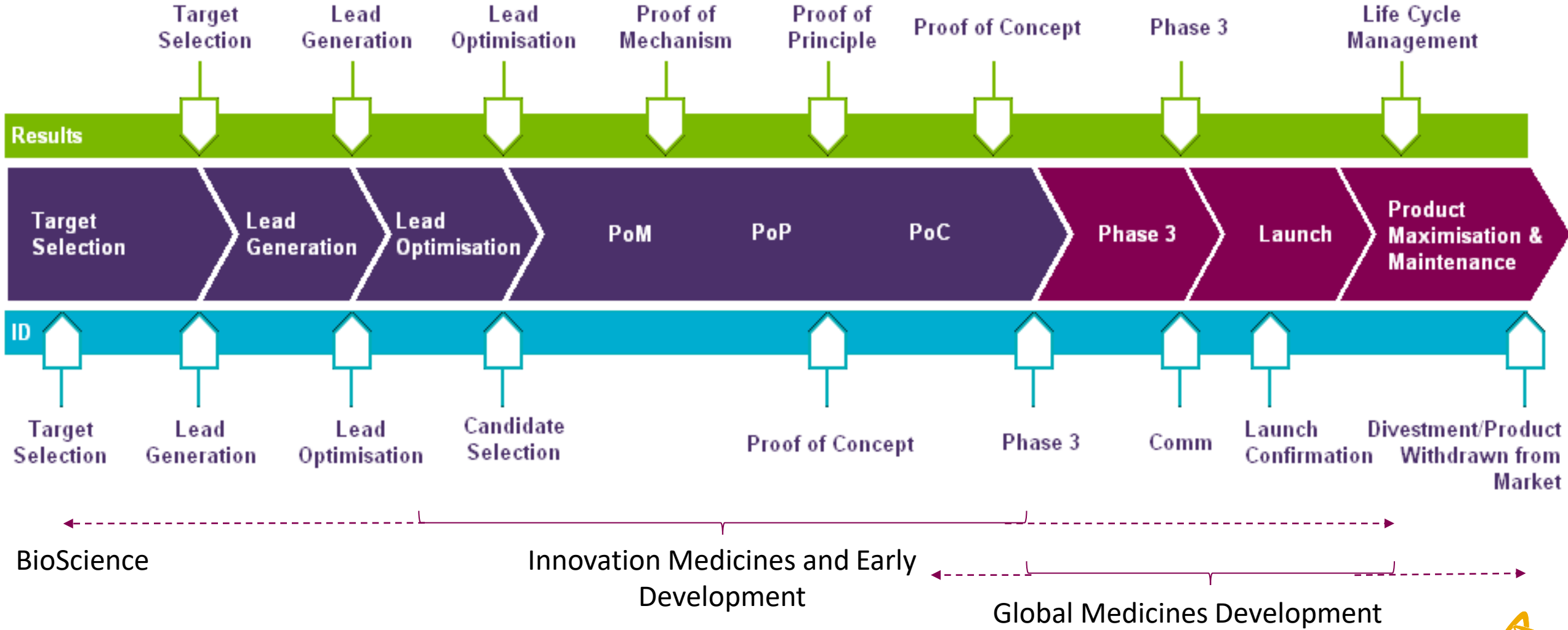
Biometrics Innovations

- **Adaptive Designs**
- **Platform/Basket/Umbrella Trials**
- **Historical Control**
- **Interim Analysis Approaches**
- **Endpoint development**
- **Large dataset approaches**
- **Trial Simulations**
- **Standardization in Reporting**
- **Tool Development**

Decision Making and Go/NoGo criteria

- Defining success

  - Stop ineffective compounds early

  - Accelerate development for effective compounds

- Consider scientific evidence for endpoints

  - external and internal data

- Understand risk

  - What actions will be taken given the decision

  - What is the probability of making each decision

Decision making drives the clinical development of compounds

What is done

- Prospective decision making criteria in place before the study begins

  - Promotes forward thinking

  - Provides context for future results

  - Speeds up decision making at the end of the study

- 3 outcome framework – Red Amber Green

  - Quantifies the risks attached to decision making

Go	Consider	Stop

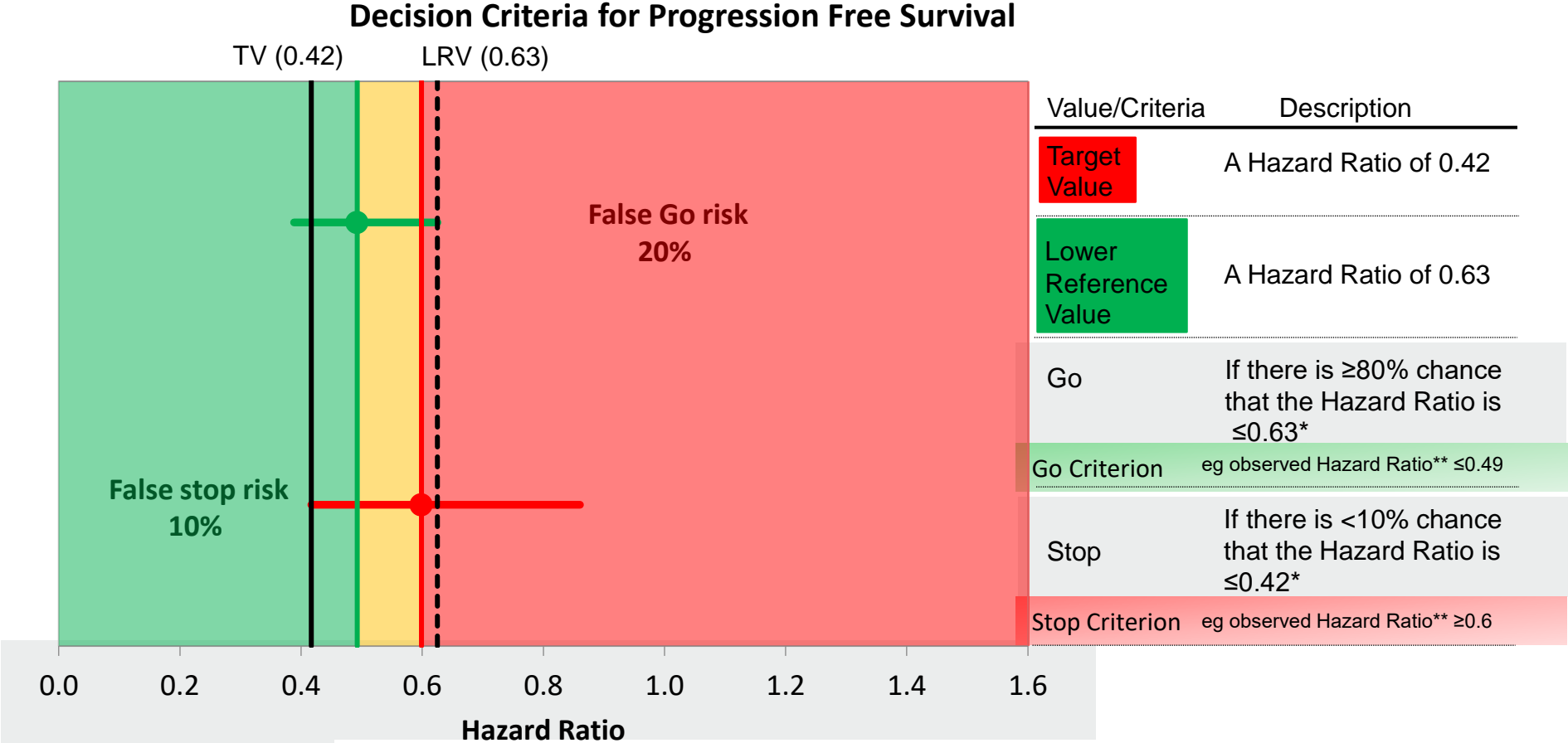Clear outcome   What to do?   Clear outcome
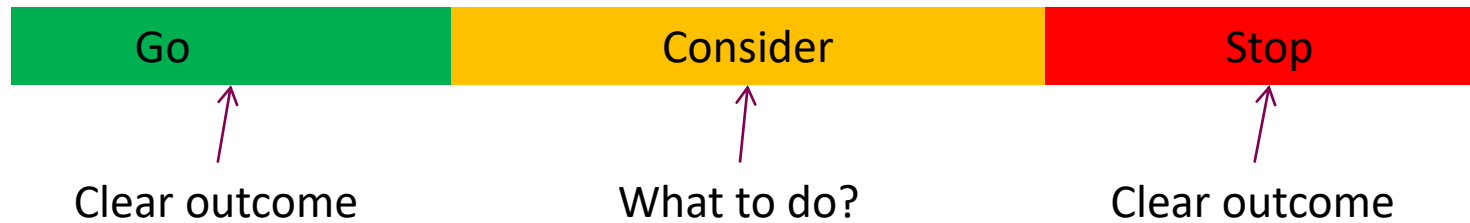
The framework, decision plot

Based on published method (Lalonde et al, 2007)

- **Target value**

  - TV - desired/meaningful performance, product profile

- **Lower reference value**

  - LRV - Minimally clinical acceptable performance

- **False Stop risk**

  - Risk of stopping the study when the truth is better than the Target value, default 10%

- **False Go risk**

  - Risk of continuing the study when the truth is less than the Lower reference value, default 20%

Example – Decision Criteria

Decision Criteria for Progression Free Survival

TV (0.42)   LRV (0.63)

False Go risk
20%

False stop risk
10%

Value/Criteria	Description
Target Value	A Hazard Ratio of 0.42
Lower Reference Value	A Hazard Ratio of 0.63
Go	If there is ≥80% chance that the Hazard Ratio is ≤0.63*
Go Criterion	eg observed Hazard Ratio** ≤0.49
Stop	If there is <10% chance that the Hazard Ratio is ≤0.42*
Stop Criterion	eg observed Hazard Ratio** ≥0.6

0.0   0.2   0.4   0.6   0.8   1.0   1.2   1.4   1.6

Hazard Ratio

** Assuming 72 patients, 50 events

* Stop and Go correspond to lower-limit of 1-sided 90% CI and upper-limit of 1-sided 80% CI

The actual criteria will be driven by the stated probabilities so that if the observed data do not follow the assumptions, the GNG values will change

# Re-purposing Ph2 decision criteria when Ph3 is expedited

- Decision criteria are predefined for phase 2 studies, which set out the different future development options for different outcomes from phase 2.

- A framework such as that described by Lalonde may be used, resulting in three potential outcomes 'Go', 'Amber', or 'Stop'.

- When a decision to proceed to phase 3 is expedited prior to the phase 2 read-out, the original decision criteria for the phase 2 studies may be redundant.

- Phase 2 studies can be re-purposed to contribute to futility decisions in the phase 2 program. New ph2 decision criteria may be defined such as to be the most informative for ongoing ph3 decision making.

Phase 2	Phase 3 futility analysis	Guidance to IDMC
Continue	Pass	Continue
Amber	Pass	Think, only stop if borderline pass in ph3
Continue	Fail	Think, consider secondary endpoints
Think	Fail	Stop

GO/NO-GO interim decision making incorporating short- and long-term endpoints

- Early Phase trials often planned and assessed using Lalonde GO/NO-GO framework (Lalonde et al 2007, Frewer et al 2016)

- Interim Analyses (IAs) are often performed at different stages of drug development

- To improve decision making at IA short-term information might be incorporated with Lalonde framework


GNG Criteria for Difference of Means
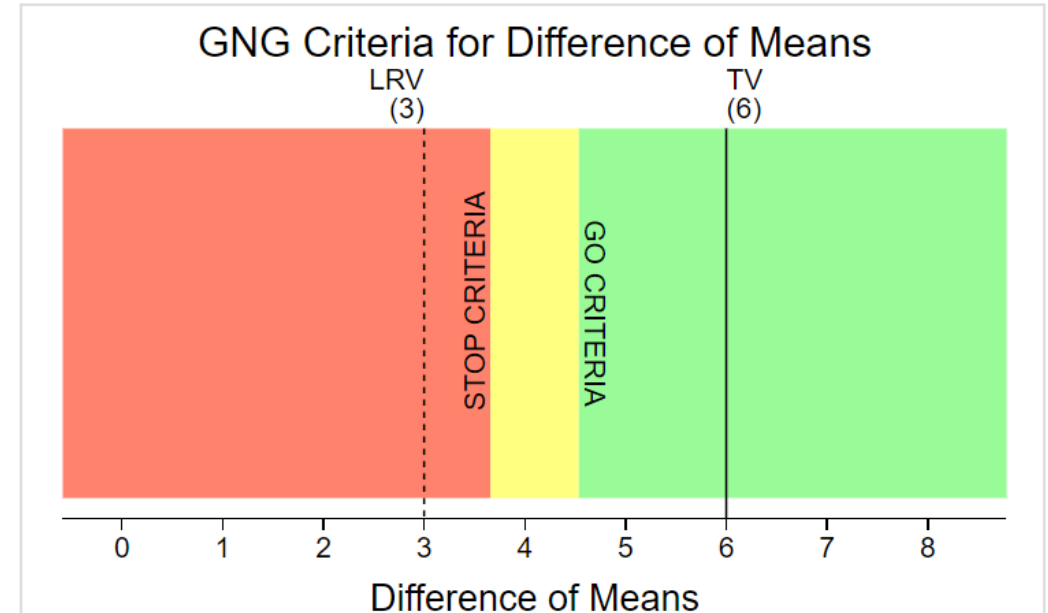
Probabilities	Go (4.54)	Consider	Stop (3.66)
Target Value (6)	78.9%	11.1%	10.0%
Lower Reference Value (3)	20.0%	15.9%	64.1%
User Interest Value (5)	60.0%	16.8%	23.2%

- LRV: smallest clinically meaningful treatment effect
- TV: desired treatment effect
- GO/NO-GO boundaries calculated based on upper/ lower confidence intervals of LRV/TV

GO/NO-GO interim decision making incorporating short- and long-term endpoints

- Estimation of the effect at interim using
  - Long-term primary outcome, L
  - Short-term outcome, S
  - Combination of both, S and L

- Lalonde GO/NO-GO updated at IA
  - Apply Lalonde directly based on data at **interim**
  - Probability in NO-GO/CONSIDER/GO zone based on the observed effect at **IA**
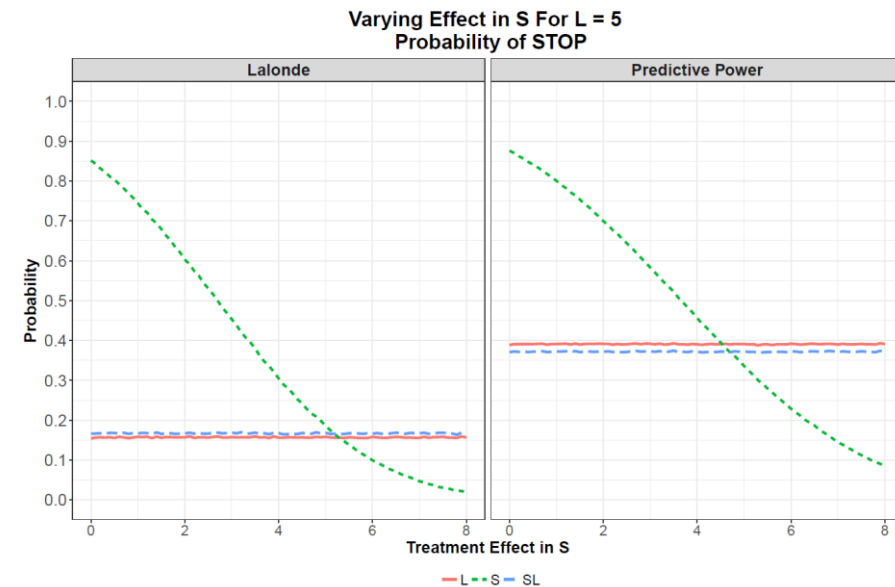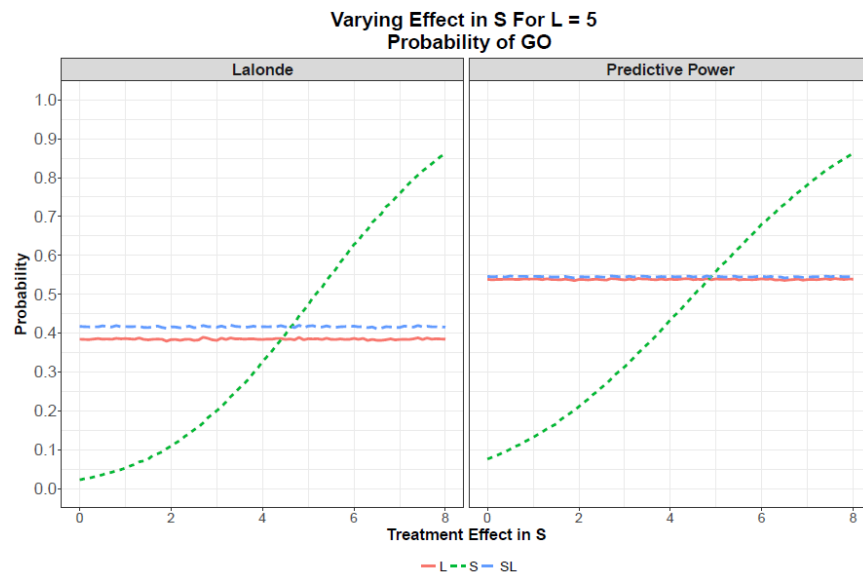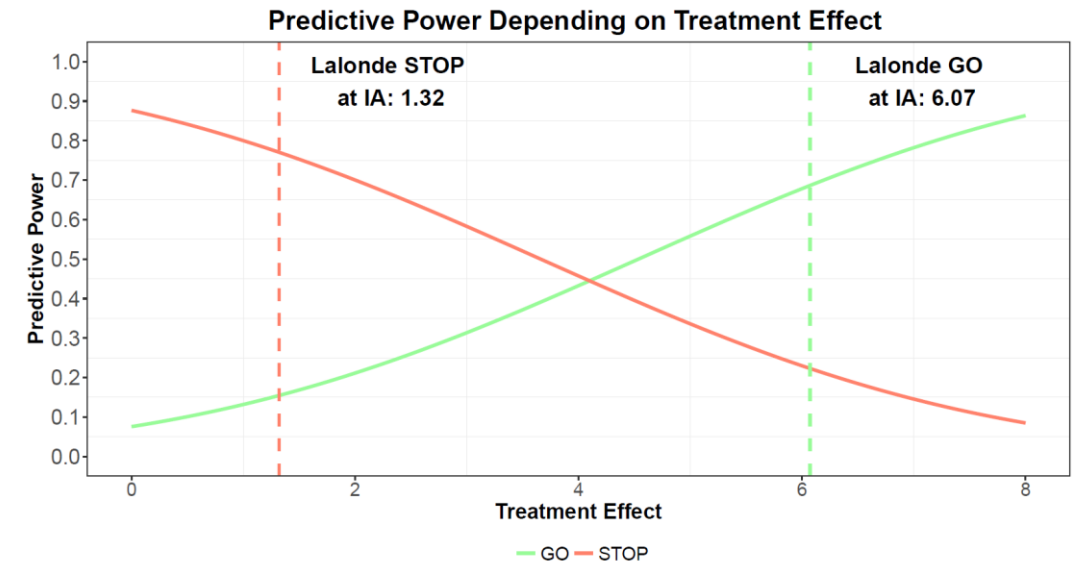
Simulation to assess impact of missing data due to COVID-19

- Clinical team requested an assessment of the impact of missing data due to COVID-19 for ongoing Ph2 study, to determine whether the sample size should be increased.

- Continuous endpoint, outcome assessed Q4W, primary analysis is MMRM

- Simulations were carried out for different amounts of missing data and different correlations between visits, to assess impact on the power and the chance of a GO decision at study end.
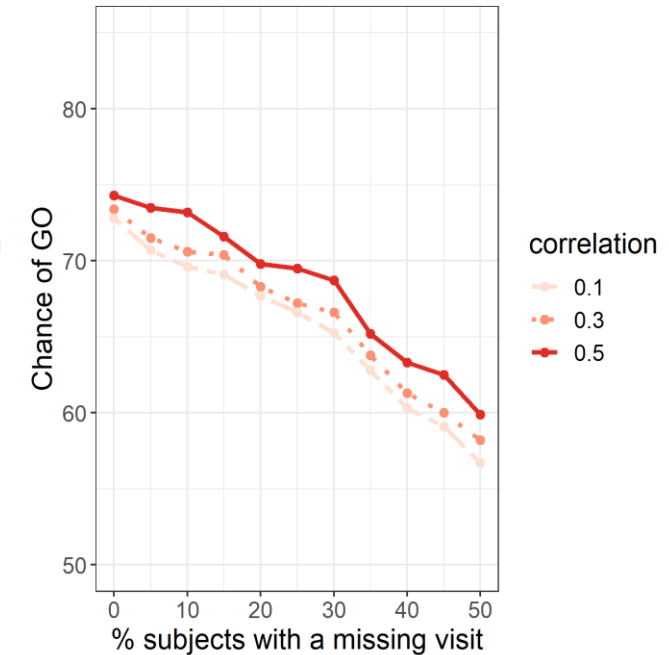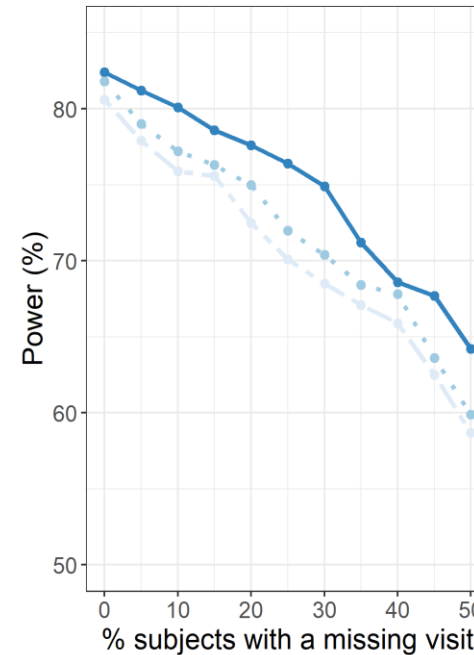
- Power and the chance of a GO hold up well when up to 15-20% of subjects have at least one visit missing, assuming the target effect was reached at the end of treatment visit.

- Team will monitor proportion of subjects with missing visits through the study.

Proof of Concept study in NASH 1(2)

Regulatory guidance in NASH

- **For full approval**
  - Outcome study for superiority in delaying disease progression

- **First accelerated approval**
  - As disease progression is slow
  - Based on histology as surrogate efficacy endpoints

Accelerated approval:

Resolution of steatohepatitis on overall histopathological reading *AND* no worsening of liver fibrosis

OR

At least one stage improvement in liver fibrosis *AND* no worsening of steatohepatitis
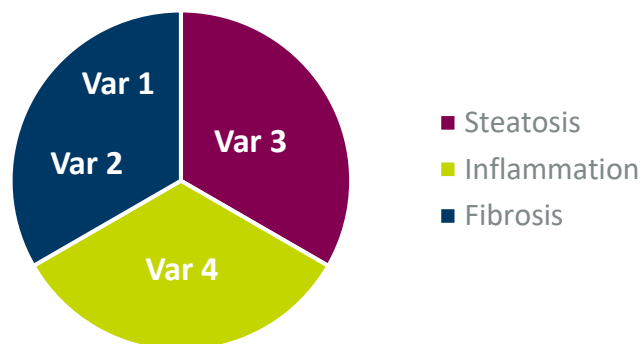
OR

Both resolution of steatohepatitis and improvement in fibrosis

Considerations for MPO

- **First study in NASH patients**
  - Generate Proof of Concept
  - Initial thought to use non-invasive markers

NASH domains:

- Steatosis
- Inflammation
- Fibrosis

Proposed decision framework
(To mimic decisions based on histology)

For each NASH domain:
Green: At least one marker GO
Amber: At least one marker amber and none GO
Red: Otherwise (i.e. STOP for all markers)

GO for next study:
At least one domain **Green** and at most one domain **Red**
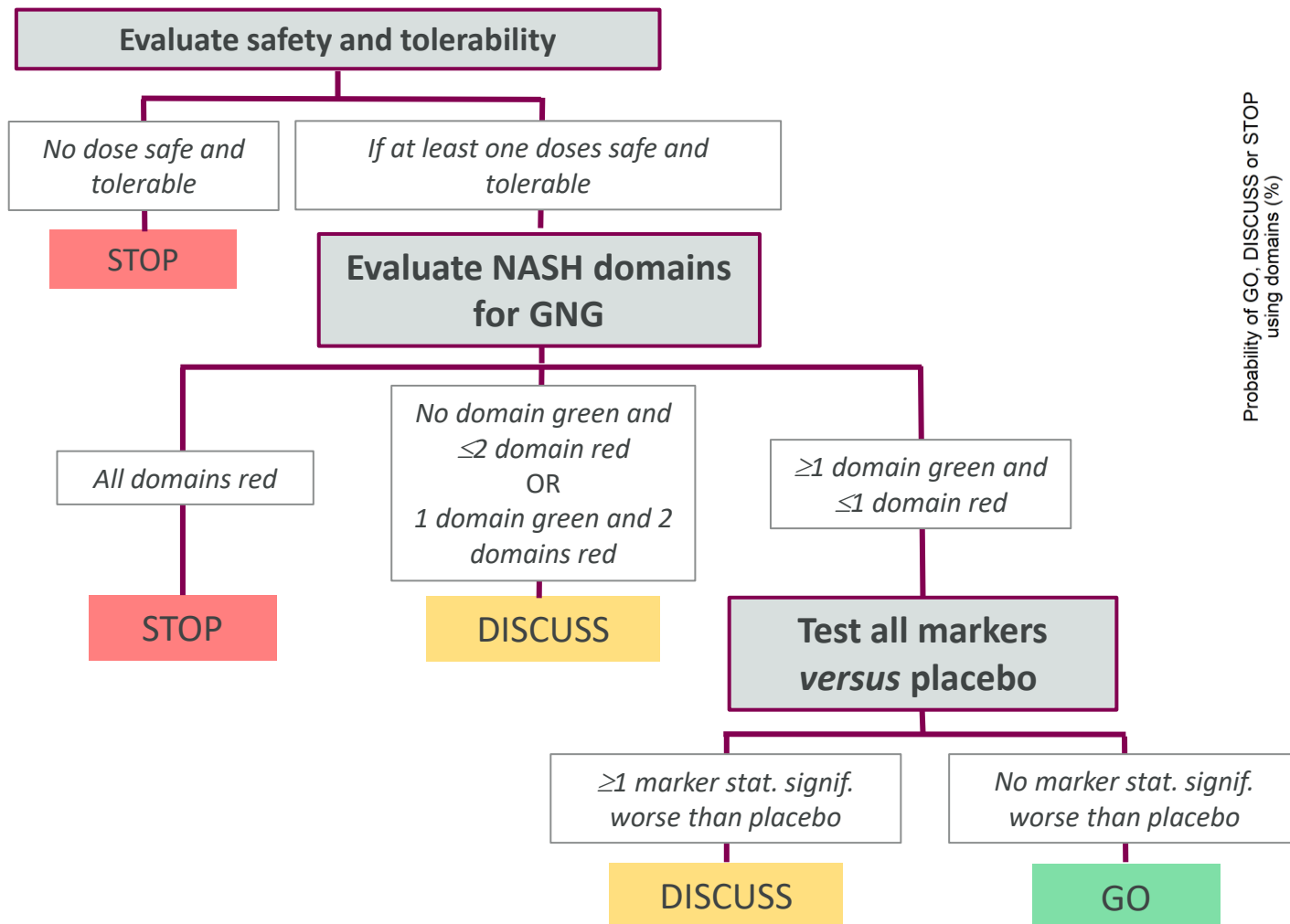
In addition for GO:
No marker statistically significantly worse than placebo

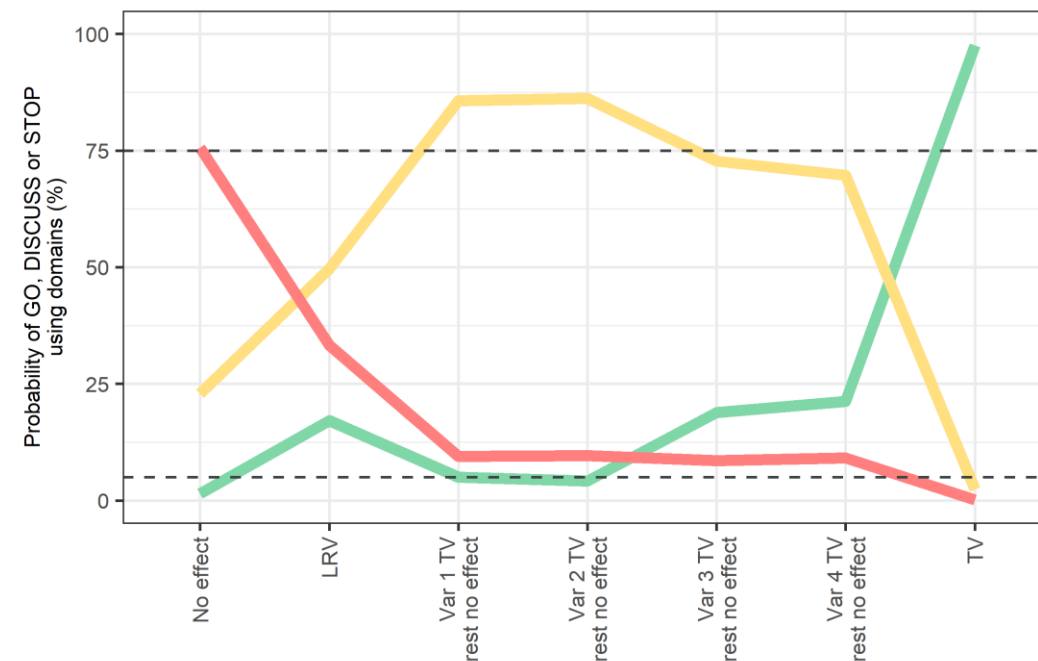NOTE: Green, amber and red according to standard decision framework

Proof of Concept study in NASH 2(2)

Proposed decision framework:

Evaluate safety and tolerability

No dose safe and tolerable

STOP

If at least one doses safe and tolerable

Evaluate NASH domains for GNG

All domains red

STOP

No domain green and ≤2 domain red OR 1 domain green and 2 domains red

DISCUSS

≥1 domain green and ≤1 domain red

Test all markers *versus* placebo

≥1 marker stat. signif. worse than placebo

DISCUSS

No marker stat. signif. worse than placebo

GO

Statistical properties:

No effect in any variable
- STOP: 76%
- DISCUSS: 23%
- GO: 1.5%

All TV effect
- STOP: 0.1%
- DISCUSS: 2.3%
- GO: 98%

Evaluation of proposed decision framework performed via simulation assuming multivariate normal distribution of log transformed variables

A Ph2a/b seamless study design to achieve 4 goals!

Part 1

Part 2

① Efficacy?

AZ Drug high dose, n=201

Placebo, n=201

|— 12 weeks —|

IA-1 based on Lalonde,
triggering Part 2

AZ Drug high dose, n=354

④ Biomarker?

Placebo, n=349

Comparator, n=305

② Competition?

AZ Drug med. dose, n=250

③ Dose?

AZ Drug low dose, n=250

|— 12 weeks —|

IA-2 based on Lalonde,
triggering Ph3 activities

BM-high patients only

BM-high & BM-low patients
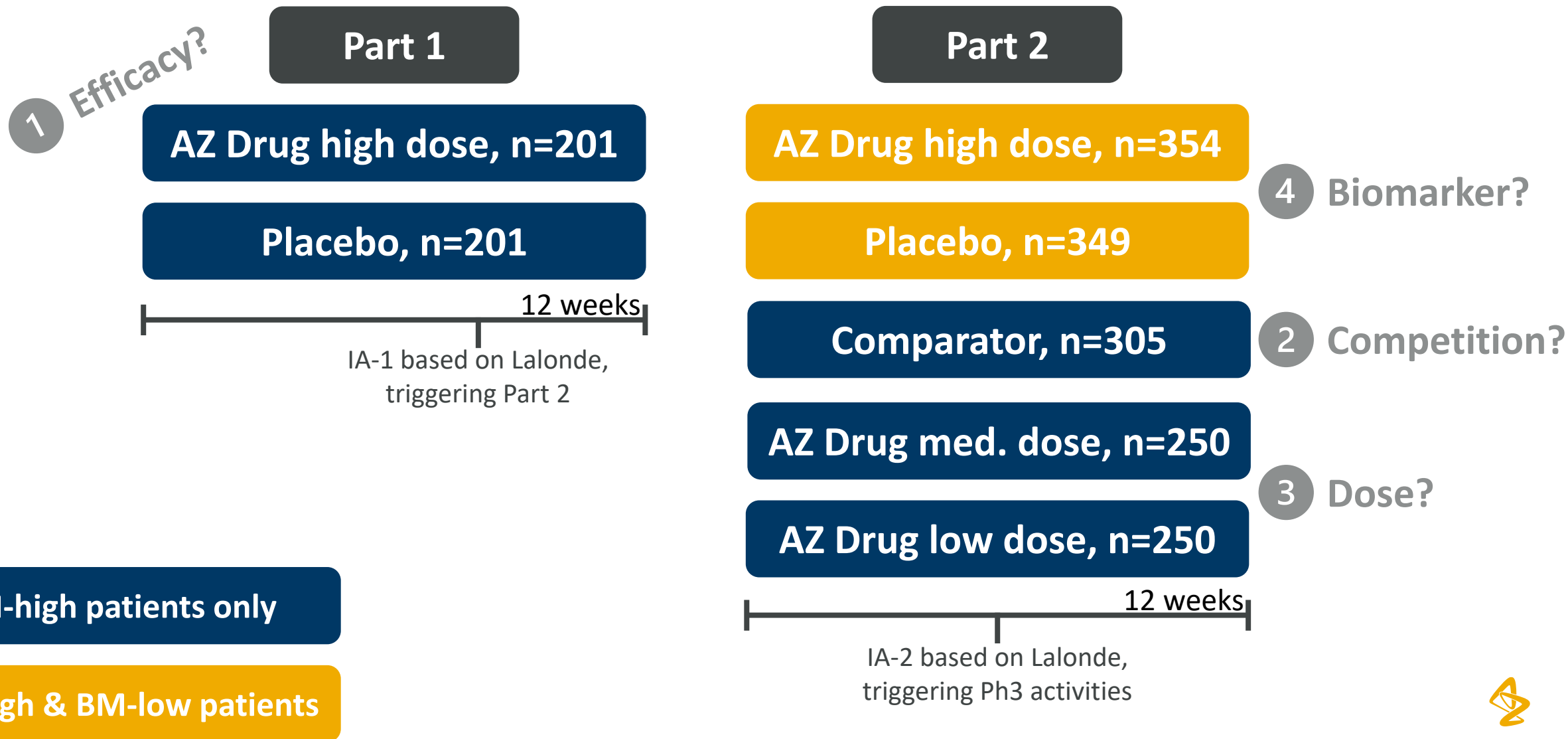
250 biomarker-low patients per arm for BM threshold fine-tuning (AZ Drug vs. Placebo)
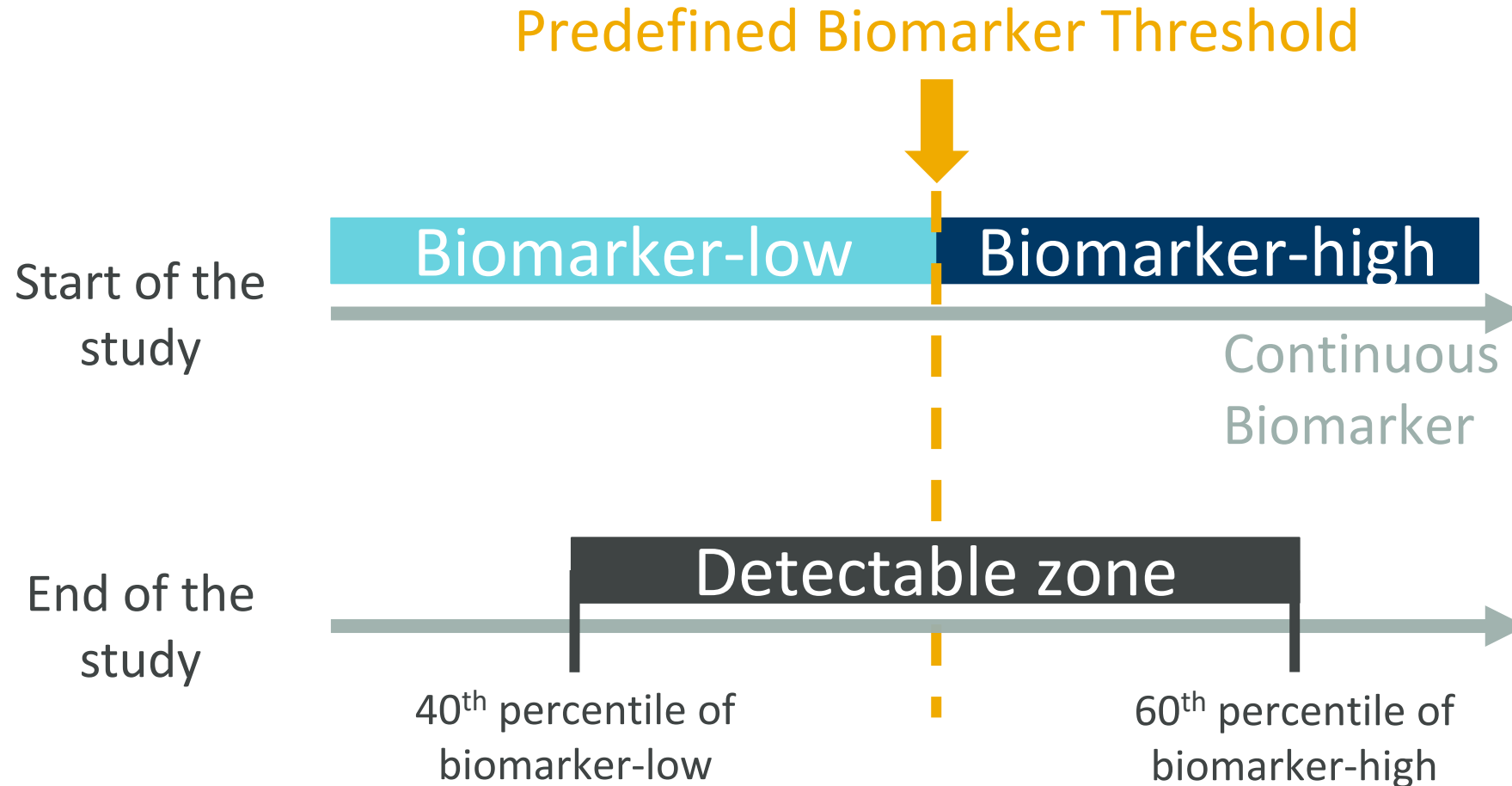
Predefined Biomarker Threshold

Start of the study

Biomarker-low

Biomarker-high

Continuous Biomarker

End of the study

Detectable zone

40th percentile of biomarker-low

60th percentile of biomarker-high

Combining scientific computing and machine learning techniques to model logitudinal outcomes

- An efficacy endpoint changes over time.

- An assumption made*: there exists a dynamical system that decribes the changes*.

- Goal: *given a set of subjects' longitudinal observations, apply machine learning algorithms to learn a governing equation (in the form of differential equations).*
  - *There exist successful applications in experimental physics*.

- Challenges in clinical trials applications:
  - *Betwen-subject variability.*
  - *Observations are collected less frequenty.*
  - *Derivatives are not observed.*

Prey-predator model: Lotke-Volterra equations

$$\frac{d(\text{🐀})}{dt} = \alpha \cdot \text{🐀} - \beta \cdot \text{🐀} \cdot \text{🦉}$$

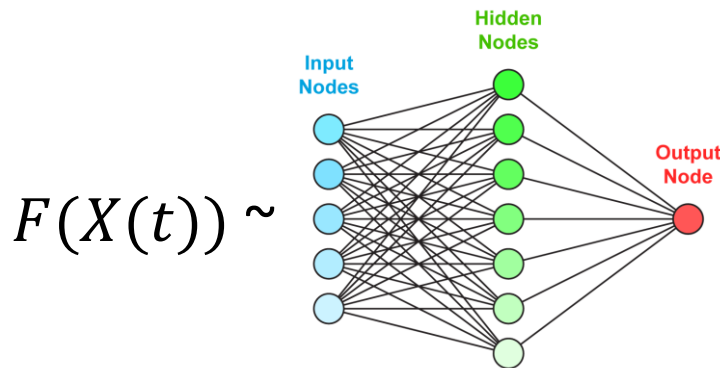$$\frac{d(\text{🦉})}{dt} = \delta \cdot \text{🐀} \cdot \text{🦉} - \gamma \cdot \text{🦉}$$

The example is taken from http://chrisrackauckas.com/research.html

Combining scientific computing and machine learning techniques to model logitudinal outcomes

The dynamics is decribed by a (system of) ordinary differential equation(s).
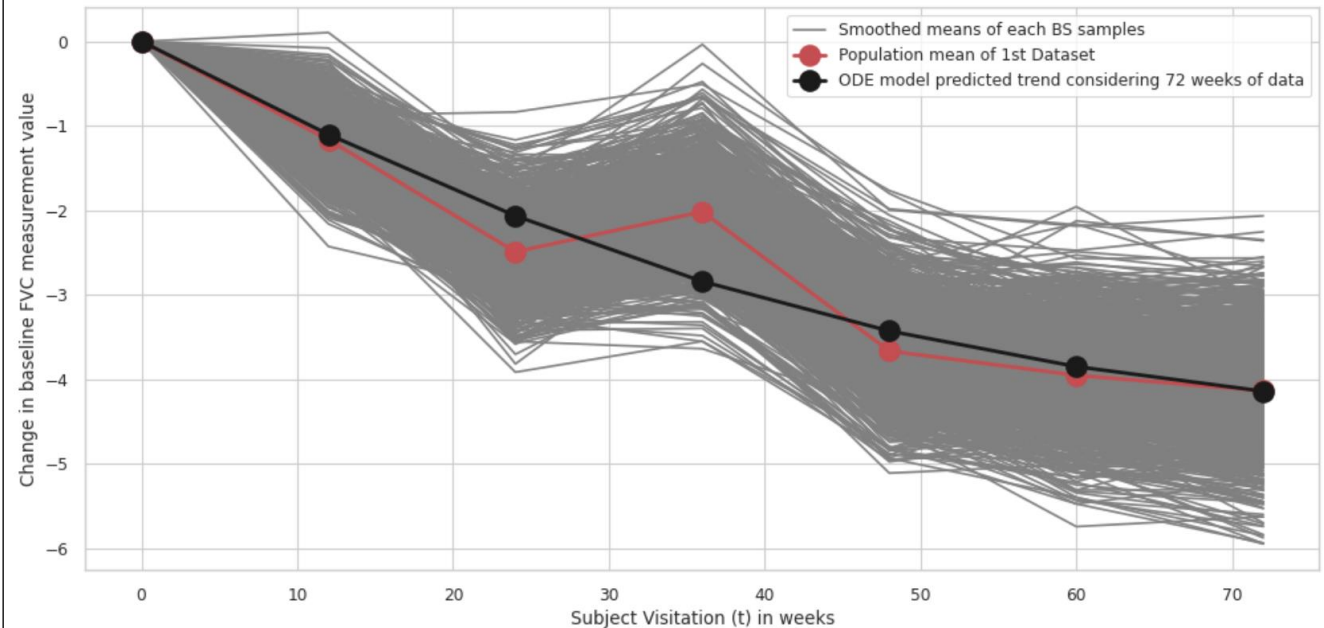
$$\frac{dX(t)}{dt} = F(X(t))$$

Unknown term can be learned with a NN trained on observed data or approximated by a nonlinear function.

$$F(X(t)) \sim$$

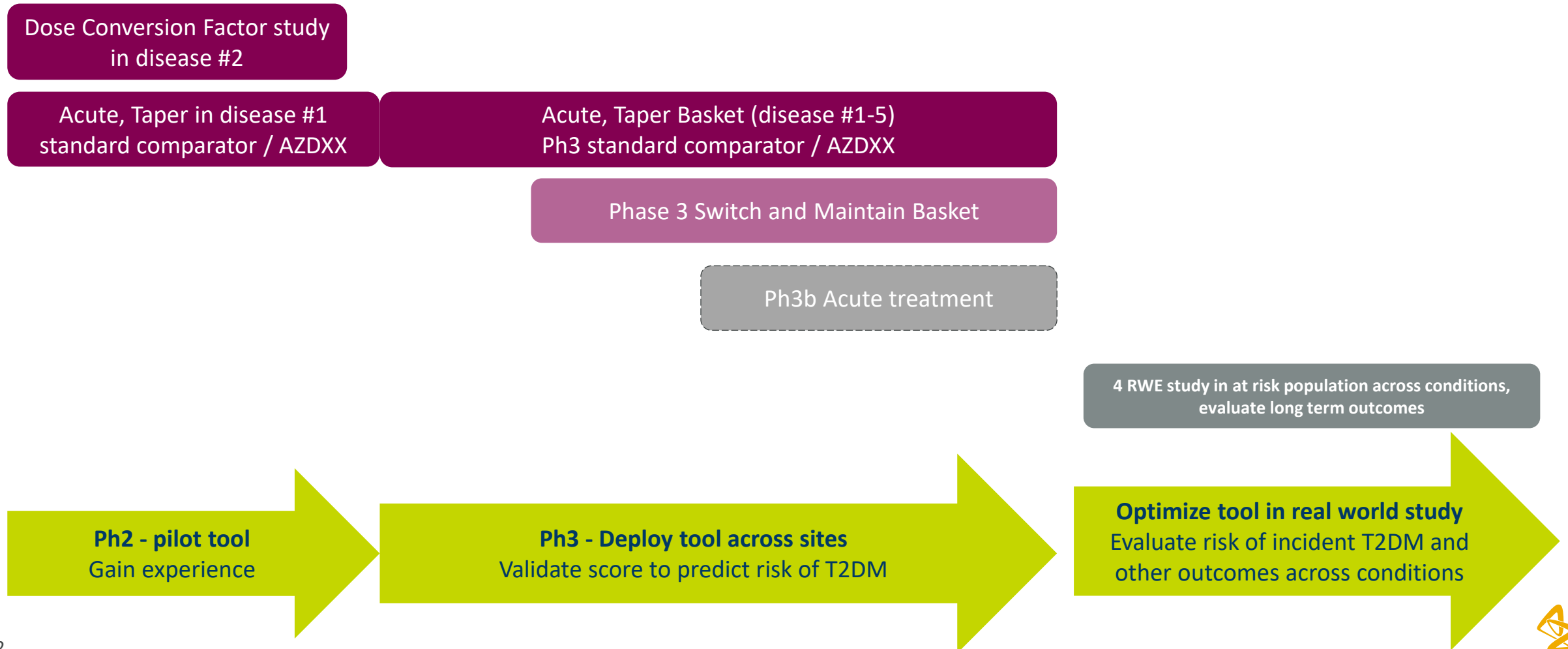Application to IPF placebo data (modeling mean CHG from BL in pFVC).

Discovered dynamics:

$$\frac{dCHG(t)}{dt} = a_0 + a_1 CHG(t)(1 - CHG(t))$$

Master student project performed at AZ: https://www.diva-portal.org/smash/record.jsf?pid=diva2%3A1569638&dswid=-924

Development of a T2DM risk score through a digital application

Aim: To use real world data to develop a T2DM risk score as screening tool to support patient selection for AZDXX clinical trials. During phase 3 the tool will be deployed, and risk score validated. A phase 4 RWE study will enable further optimization to determine the right patient who can benefit from AZDXX

Dose Conversion Factor study in disease #2

Acute, Taper in disease #1 standard comparator / AZDXX

Acute, Taper Basket (disease #1-5) Ph3 standard comparator / AZDXX

Phase 3 Switch and Maintain Basket

Ph3b Acute treatment

4 RWE study in at risk population across conditions, evaluate long term outcomes

Ph2 - pilot tool
Gain experience

Ph3 - Deploy tool across sites
Validate score to predict risk of T2DM

Optimize tool in real world study
Evaluate risk of incident T2DM and other outcomes across conditions

19 March 2021

Definition of success: further use case criteria to be agreed upon

Benchmark

- The QDiabetes®-2018 tool can be used as a 'roadmap' for payer adoption

- With a more refined target population, we can target a similar or better predictive capacity

  - C-statistic >0.8 in external validation
  - Brier score <0.34 in external validation

Strategies for Use

- Outline strategies currently in use for identifying and treating patients at high risk of T2DM and proposed changes in strategy using the tool

- Define the minimum and desired sensitivity in those scenarios

Transportability

- Simpler input variables, such as a web-form, can be used by anyone anytime; however, loss of precision/prediction

- Longitudinal variables (e.g. cumulative exposure to standard comparator instead of a binary, change in laboratory measures instead of last measure) reduce transportability but potentially enhance prediction

https://www.bmj.com/content/359/bmj.j5019

Acknowledgements

Jason Cooper

Guillaume Desachy

Mahdi Hashemi

Alexandra Jauhiainen

Magnus Kjaer

Marcus Millegård

Rachel Moate

Karin Nelander

Julia Niewczas

Fredrik Öhrn

Chris Rackauckas

Yevgen Ryeznik

José Sánchez

Jennifer Schumi

Sofia Tapani

John Tillinghast

Lan-Feng Tsai

Jim Weatherall

David Wright

Wenjing Xin

Thank you.

Questions
& Answers

Name | 20 October 2021