PGx in Real-World Data Studies

Professor Daniel Prieto-Alhambra
Dr Junqing (Frank) Xie
University of Oxford
24/09/2024

DARWIN EU® Real-World data/study/evidence

Catalogue of Standard Data Analyses

Off-the-shelf studies

These are mainly characterisation questions that can be executed with a generic protocol. This includes disease epidemiology, for example the estimation of the prevalence, incidence of health outcomes in defined time periods and population groups, or drug utilization studies at the population or patient level.

- Patient-level characterisation
- Patient-level DUS analyses
- Population-level DUS analyses
- Population-level descriptive epidemiology

Complex

These are studies requiring development or customisation of specific study designs, protocols, analytics, phenotypes. This includes studies on the safety and effectiveness of medicines and vaccines.

- Prevalent user active comparator cohort studies
- New user active comparator cohort
- Self-controlled case risk interval
- Self-controlled case series
- Time series analyses and Difference-in-difference studies
- RMM effectiveness

Real-world data

Complex

These are studies requiring development or customisation of specific study designs, protocols, analytics, phenotypes. This includes studies on the safety and effectiveness of medicines and vaccines.

Population genetics

Biobanks:
Smaller, richer, linked

- Lifestyle factors
- Clinical measurements
- **Genomics***
- Proteomics
- PROMs

Biobanks:
Smaller, richer, linked

- Better causal inference
  - Confounders
  - Mendelian Randomisation
- Mechanistic studies
- Pharmacogenomics ★
- Vaccinomics ★

Support growing regulatory needs for drugs/devices

Existing and ongoing global biobanks

biobank uk
Enabling scientific discoveries that improve human health

Table 1. Comparison of the Self-Reported Ethnic Origins of UK Biobank Participants (Recruited in 2006–2010) With Census Data for the Age Group 40–69 Years in England, Wales, and Scotland in 2001 and 2011[a]

Ethnicity[b]	UK Biobank (n = 499,877)		2001 UK Census (n = 20,198,307)		2011 UK Census (n = 23,146,612)	
	No. of Persons	%	No. of Persons	%	No. of Persons	%
White[c]	472,837	94.6	19,085,322	94.5	21,133,317	91.3
Black or black British[d]	8,066	1.6	302,073	1.5	565,777	2.4
Mixed[e]	2,958	0.6	82,389	0.4	191,085	0.8
Indian	5,951	1.2	325,651	1.6	442,338	1.9
Pakistani	1,837	0.4	147,695	0.7	239,166	1.0
Bangladeshi	236	0.0	46,220	0.2	75,919	0.3
Chinese	1,574	0.3	70,572	0.3	109,412	0.5
Other Asian	1,858	0.4	73,917	0.4	240,324	1.0
Other ethnic group	4,560	0.9	64,468	0.3	149,274	0.6

500 K participants, recruited between 2006-2010

Scotland *~36,000 participants*

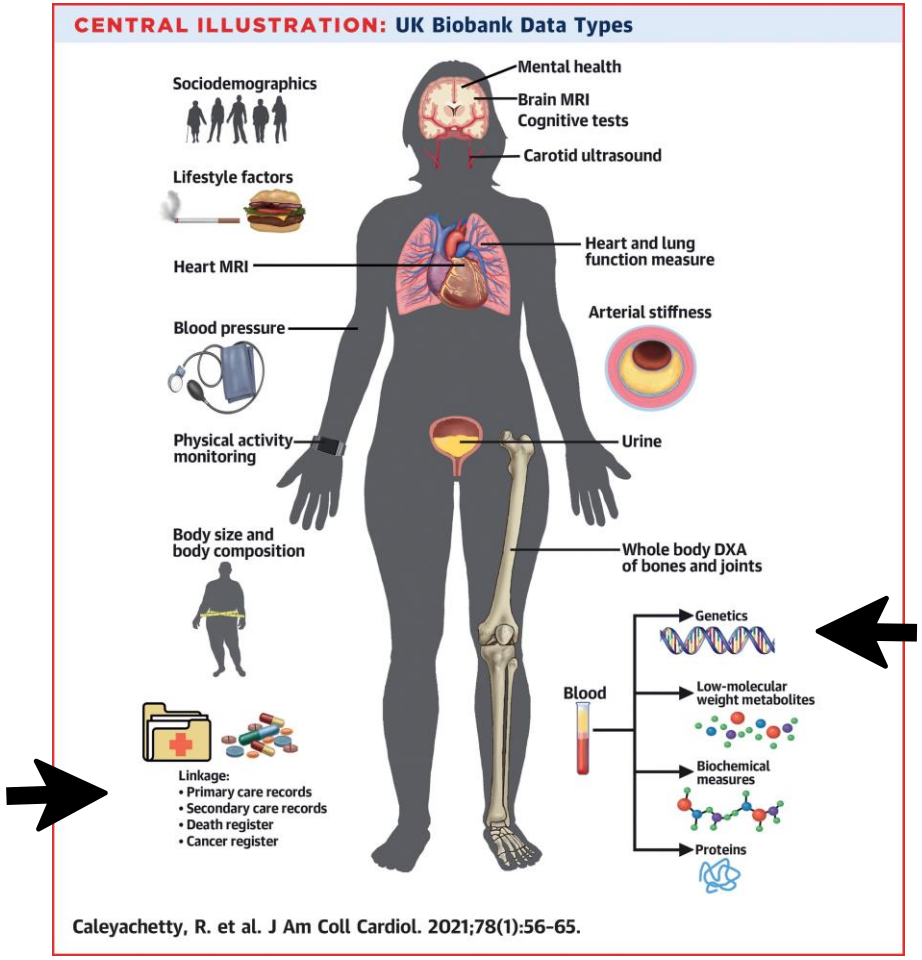
England *~450,000 participants*

~Wales *20,000 participants*

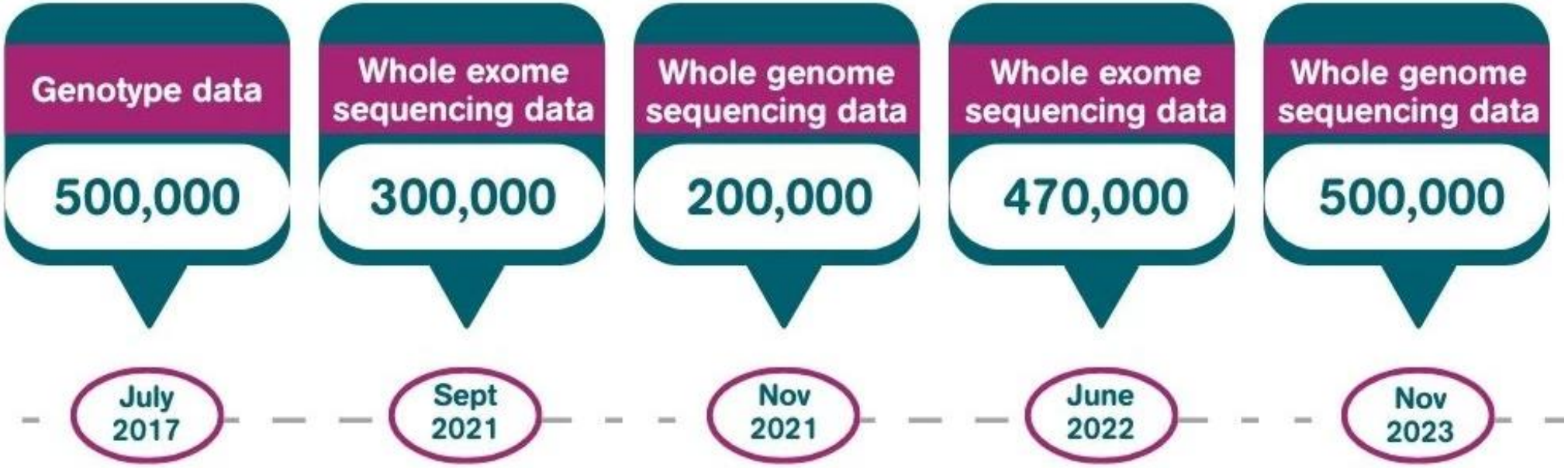
biobank uk

Enabling scientific discoveries that improve human health

Touchscreen questionnaire and computer-assisted verbal interview	
Sociodemographic	Ethnicity, education, employment, household information, Townsend deprivation index (socioeconomic status)
Lifestyle	Smoking; alcohol consumption; physical activity; diet; sleep
Environmental factors	Current address; current (or last) occupation; domestic heating and cooking fuel; housing; means of travel; shift work; mobile phone use; sun exposure
Early life factors	Birthplace, birth weight, breastfed, childhood body size and height, maternal smoking, handedness, adopted, and part of multiple birth
Family history	Illnesses of father/mother/siblings, age of parents, age parents died, and number of siblings
Psychosocial factors	Social support, bipolar/major depression, anxiety, nerves, psychological traits, and mood
Health and medical history	Medical conditions, medications, operations, cancer screening, pain, oral health, eyesight, hearing, and general health
Sex-specific factors	Male specific—first facial hair, age voice broke, hair/balding pattern, children fathered; female specific—hormone replacement therapy, contraception, pregnancy, menstruation, menopause, and cervical test
Cognitive function	Pairs matching; reaction time; prospective memory[a]; fluid intelligence[a]; numeric memory[b]
Hearing tests	Speech reception threshold[a]



Granular genetic data



https://www.ukbiobank.ac.uk/enable-your-research/about-our-data/genetic-data

Existing and ongoing global biobanks

Our Future Health

~5 million, 10x bigger than UKBB !!!

Current data release

989,132
participants with completed baseline health questionnaires

330,069
participants with genotype array data

644,119
participants with linked health record data

Date of latest release:
26 June 2024

Release documentation:

Our Future Health data dictionary (Excel file)

Participant and Questionnaire coding file (CSV file)

NHS England linked health records coding file (CSV file)

Genotyping array CPRA ("Chrom:Pos:Ref:Alt") variant list (CSV file)

Table 2: Demographic characteristics

	N	%
Baseline questionnaire data	989,132	100
Sex [a]		
Female	560,995	56.7
Male	427,643	43.2
Intersex / Other / Prefer not to answer	494	0.0
Age (years) [b]		
18-29	82,998	8.4
30-39	140,730	14.2
40-49	156,404	15.8
50-59	210,580	21.3
60-69	234,258	23.7
70-79	140,231	14.2
80+	23,931	2.4

Existing and ongoing global biobanks

All of Us RESEARCH PROGRAM

~ 1 million, US

Diversity

Includes racial and ethnic minorities as well as sexual and gender minorities, people with low income or limited education, and other groups.

~45% Racial and Ethnic Minorities

80+% Underrepresented in Biomedical Research

Participants at a Glance

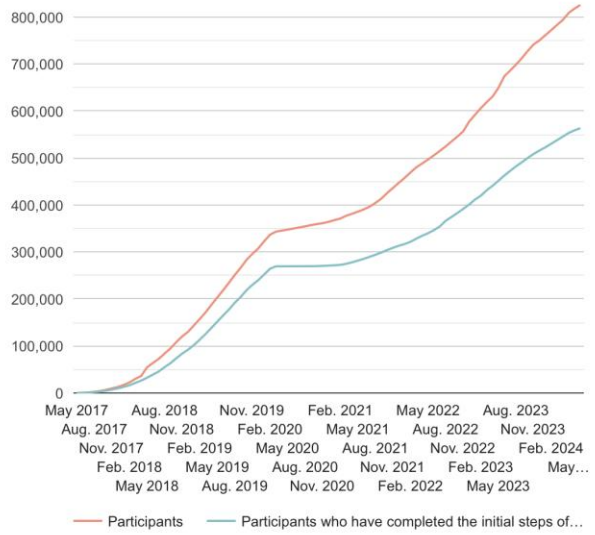
832,000+
Participants

566,000+
Participants who have completed initial steps of the program

Enrollment Numbers

This graph represents participants who have consented to join the program and those who have completed all initial steps of the program. The initial steps are consenting, agreeing to share electronic health records, completing the first three surveys, providing physical measurements, and donating at least one biospecimen to be stored at the biobank.

The following numbers are approximated to protect participants' privacy. Numbers are updated as of August 26, 2024.



https://allofus.nih.gov/get-involved/opportunities-researchers

EHR Domains

Conditions

25,638
medical concepts

254,700 participants

View Conditions

Drug Exposures

29,865
medical concepts

239,740 participants

View Drug Exposures

Labs & Measurements

16,618
medical concepts

255,640 participants

View Labs & Measurements

Procedures

30,328
medical concepts

242,580 participants

View Procedures

Genomics

SNV/Indel Variants

245,400
Participants in Short-Read Whole Genome Sequencing (WGS) dataset

1,074,881,214
SNV/Indel Variants

View SNV/Indel Variants

Genomic data only in Researcher Workbench

1,040 participants in the Long-Read WGS dataset

97,060 participants in the Short-Read WGS Structural Variants dataset

312,940 participants in the Genotyping Arrays dataset

Register for access

Measurements and Wearables

Physical Measurements

8
Physical Measurements

337,540 participants

Participants have the option to provide a standard set of physical measurements.

View Physical Measurements

Fitbit

6
Fitbit Measurements

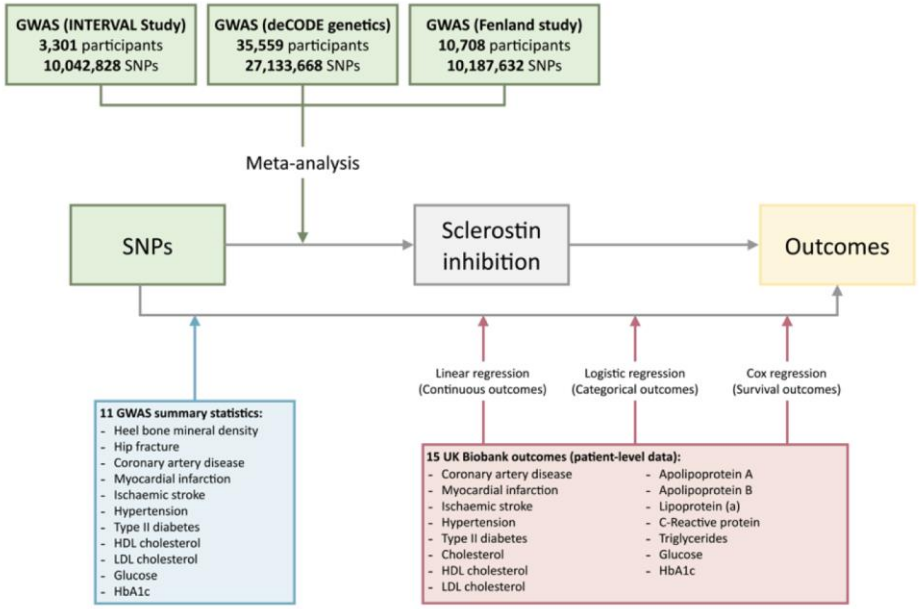
15,620 participants

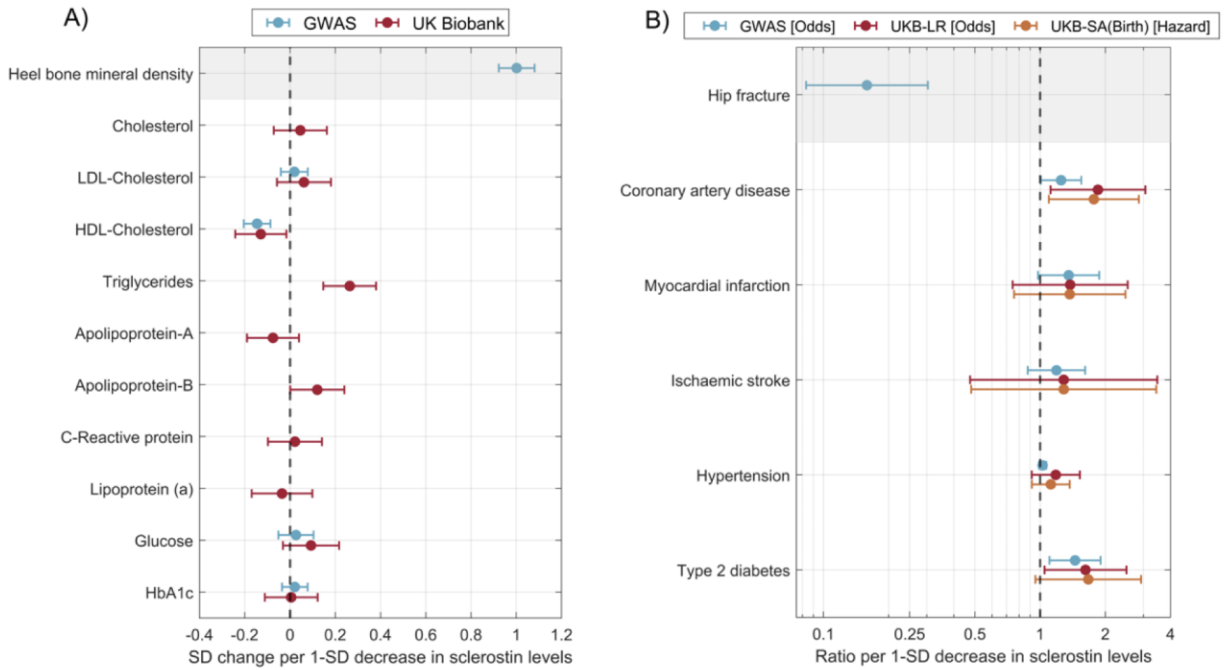
Fitbit data includes heart rate and activity summaries.

View Fitbit

Use cases (1): Drug safety

What are effects of sclerostin inhibition on CVD risk factors and outcomes?

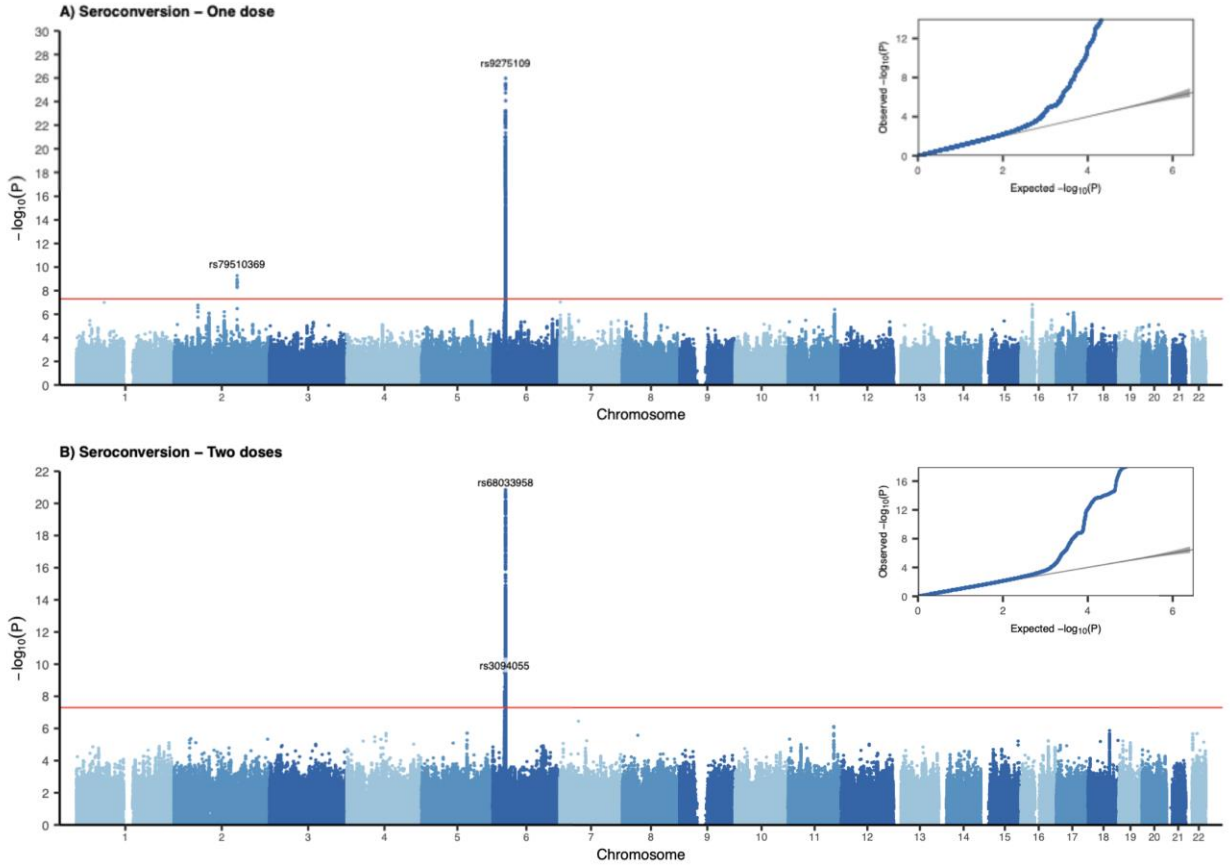




Accepted in Nature Communications

Use cases (2): Mechanistic

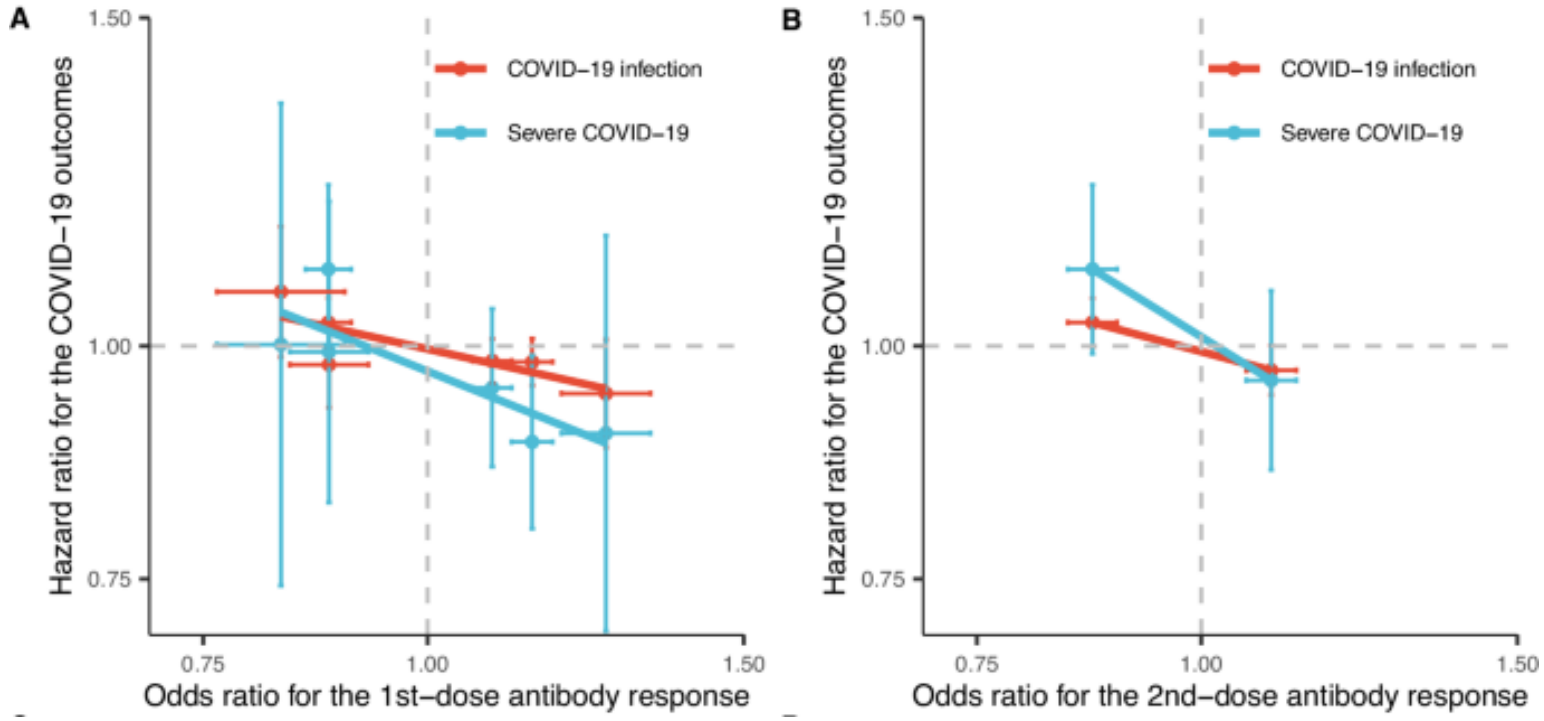
What are the genetic determinants for COVID-19 vaccines antibody responses?



Accepted in Nature Communications

Use cases (3): Vaccine effectiveness

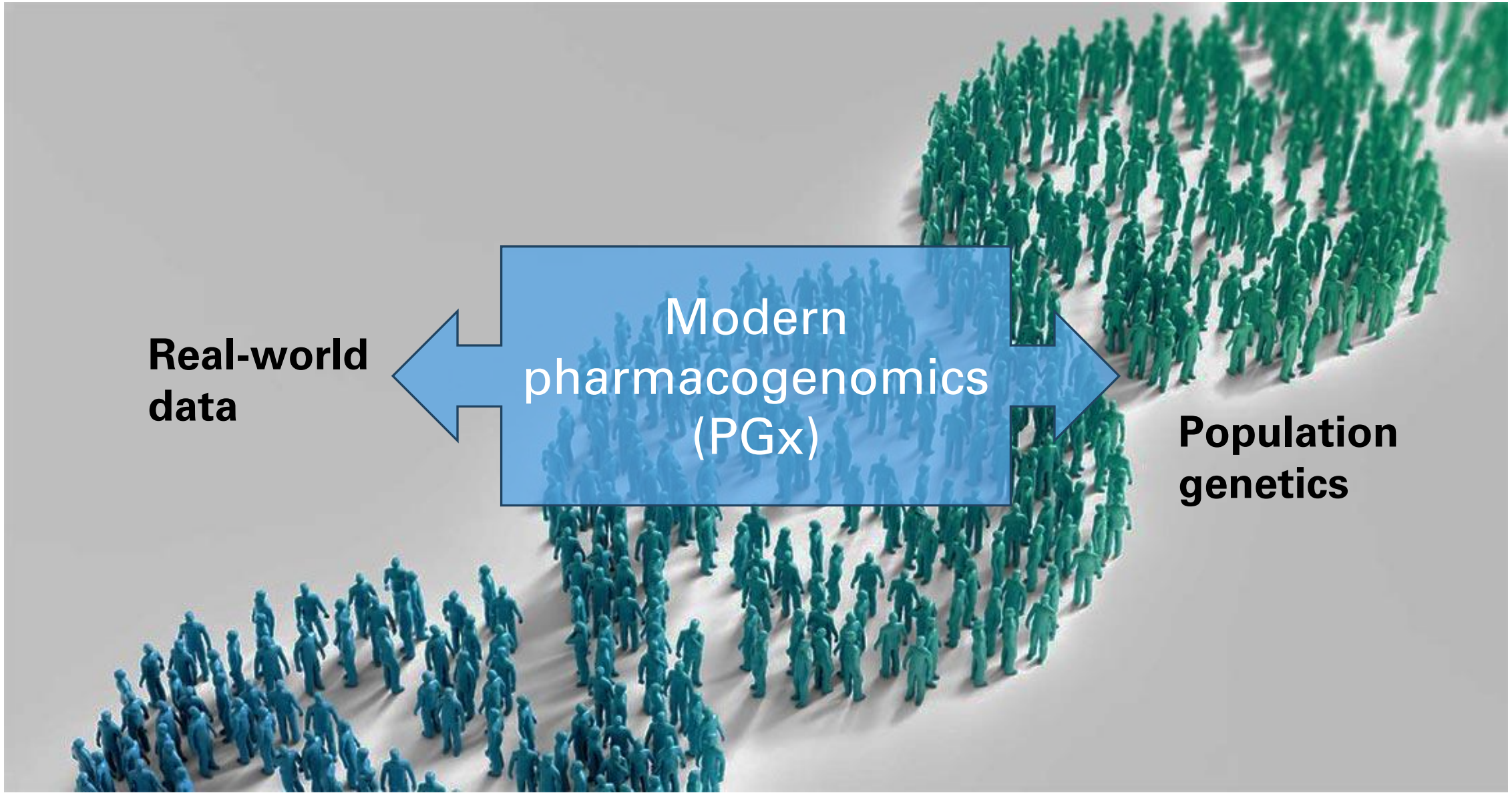
(more details covered by Frank)



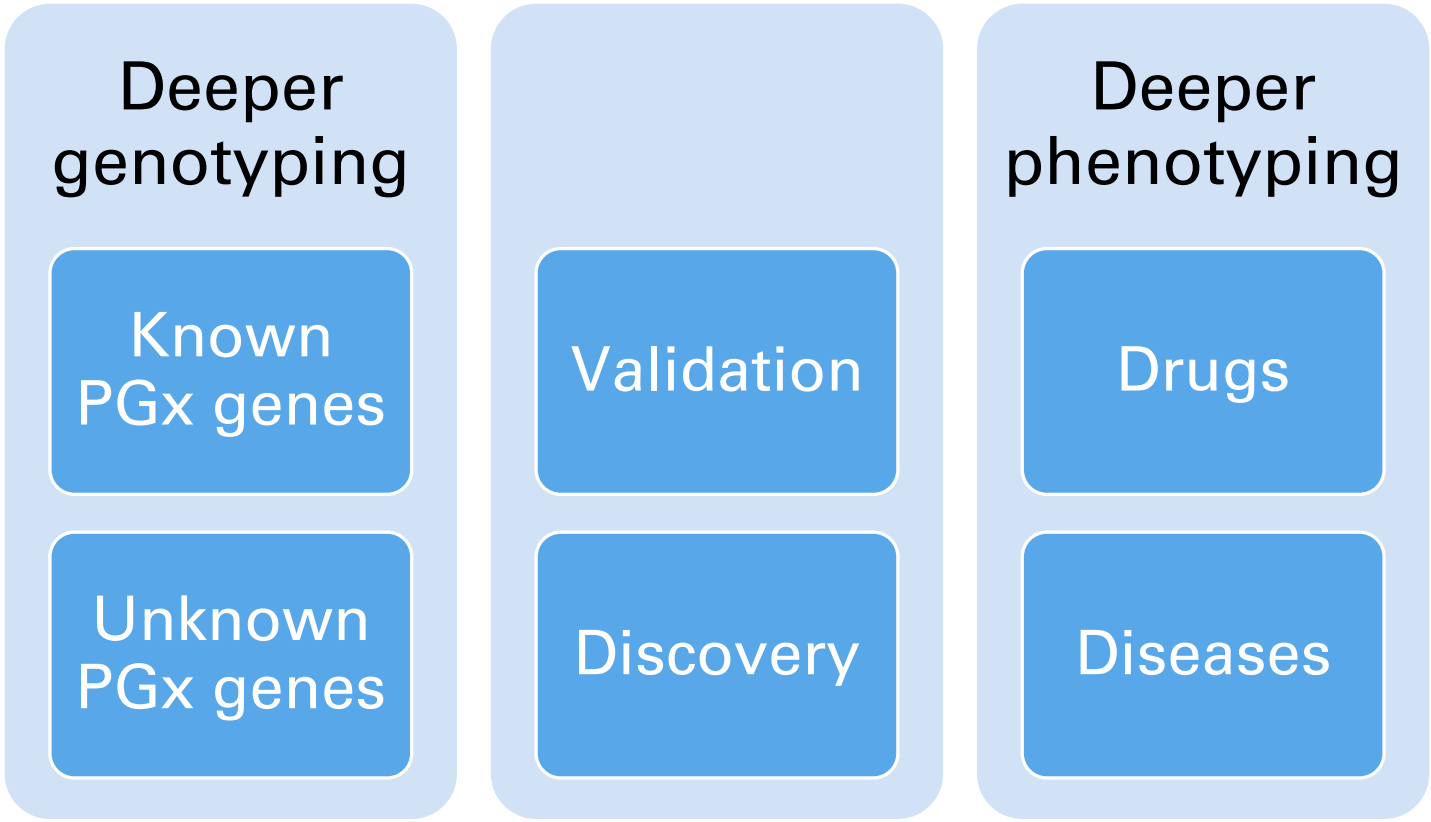
Published in Nature Communications

Junqing Xie (Frank)

One recent example for potential public health interest:


Real-world data
Modern pharmacogenomics (PGx)
Population genetics

Maximizing drugs efficacy while minimizing unwanted effects in real-world settings



How can we make more personalized recommendations about the number of doses people may need?

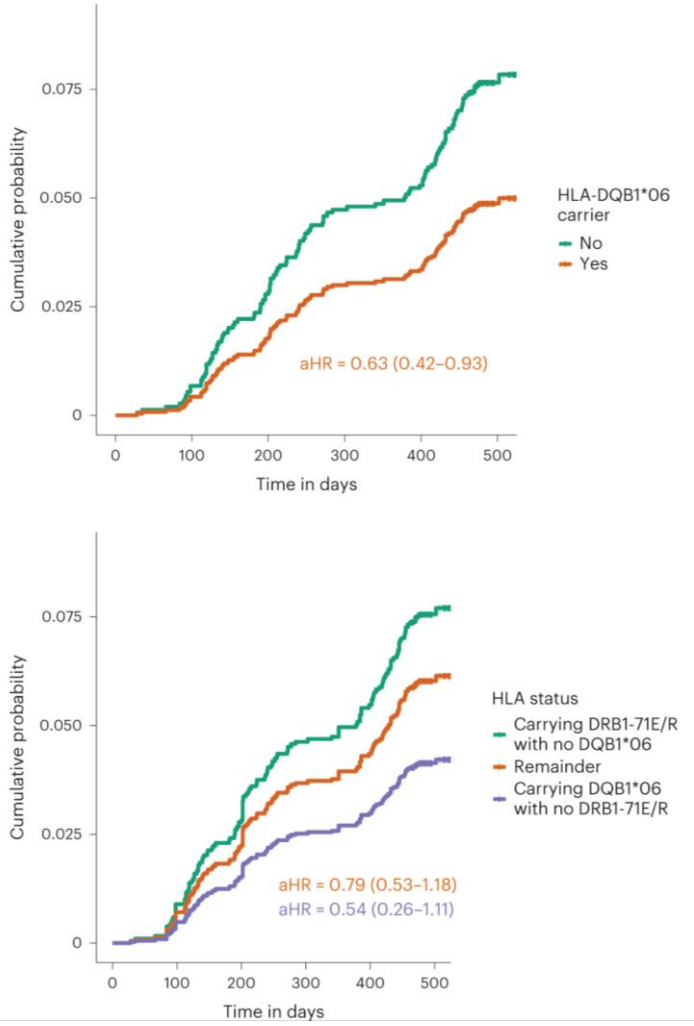
nature medicine

Article

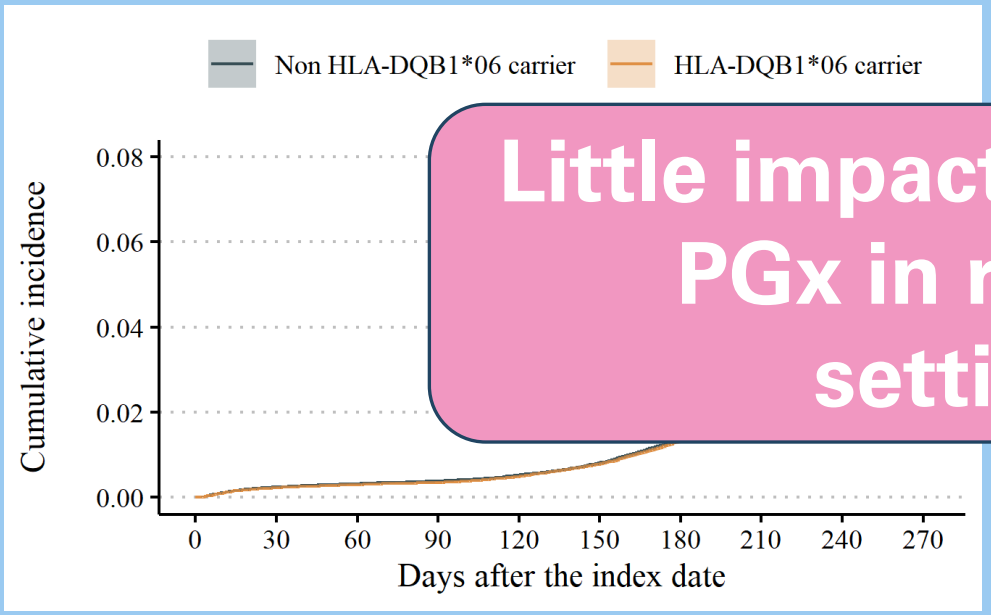
https://doi.org/10.1038/s41591-022-02078-6

Human leukocyte antigen alleles associate with COVID-19 vaccine immunogenicity and risk of breakthrough infection

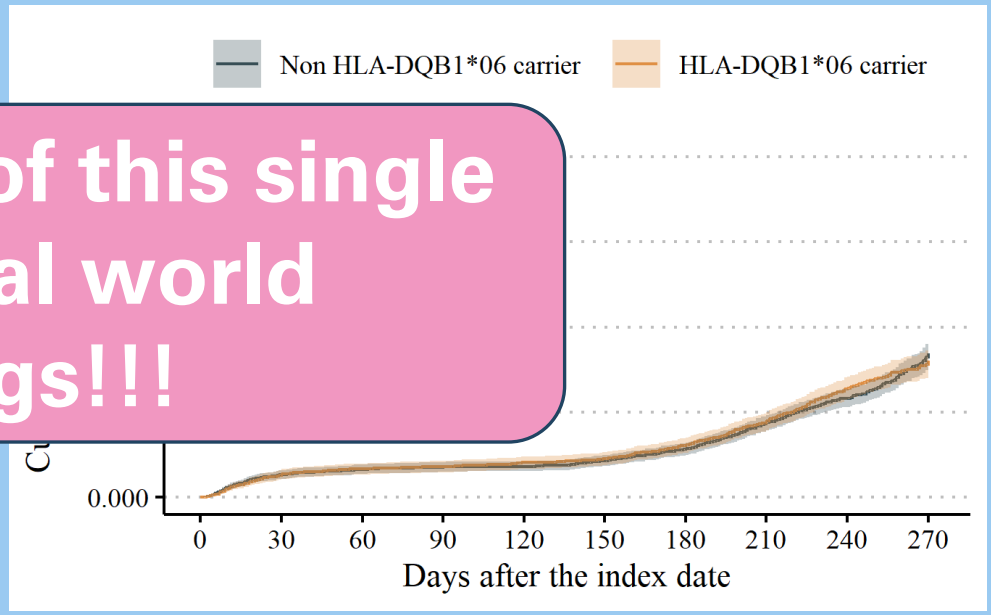
Findings:
Trial participants with HLA-DQB1*06 gene alleles had **37%** lower risk of breakthrough COVID-19 infection.



We conducted a similar study linking PGx with RWD in UK Biobank.



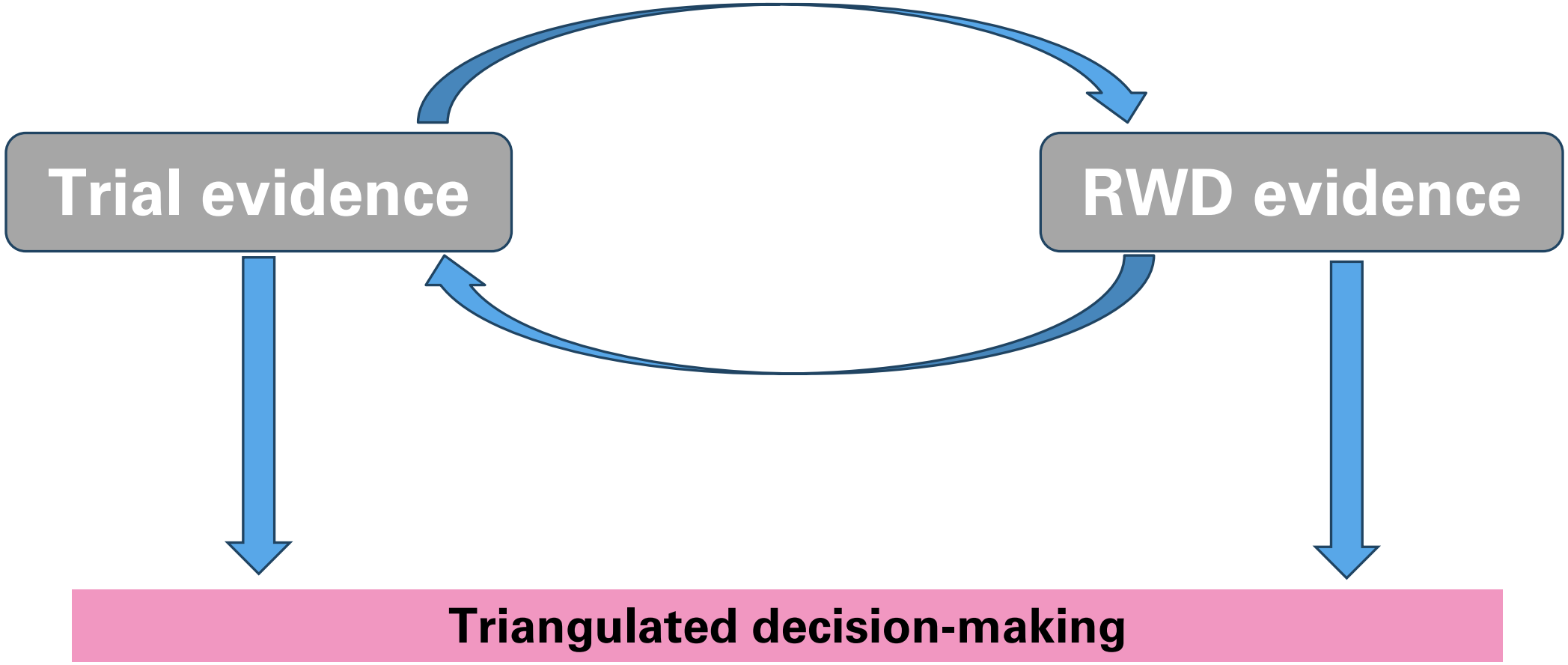
Infection: HR 0.98 (0.94 to 1.02)



Hospitalisation: HR 1.01 (0.89 to 1.14)

Little impact of this single PGx in real world settings!!!

Why & Lessons


Trial evidence
RWD evidence
Triangulated decision-making

This is not the end of our story!

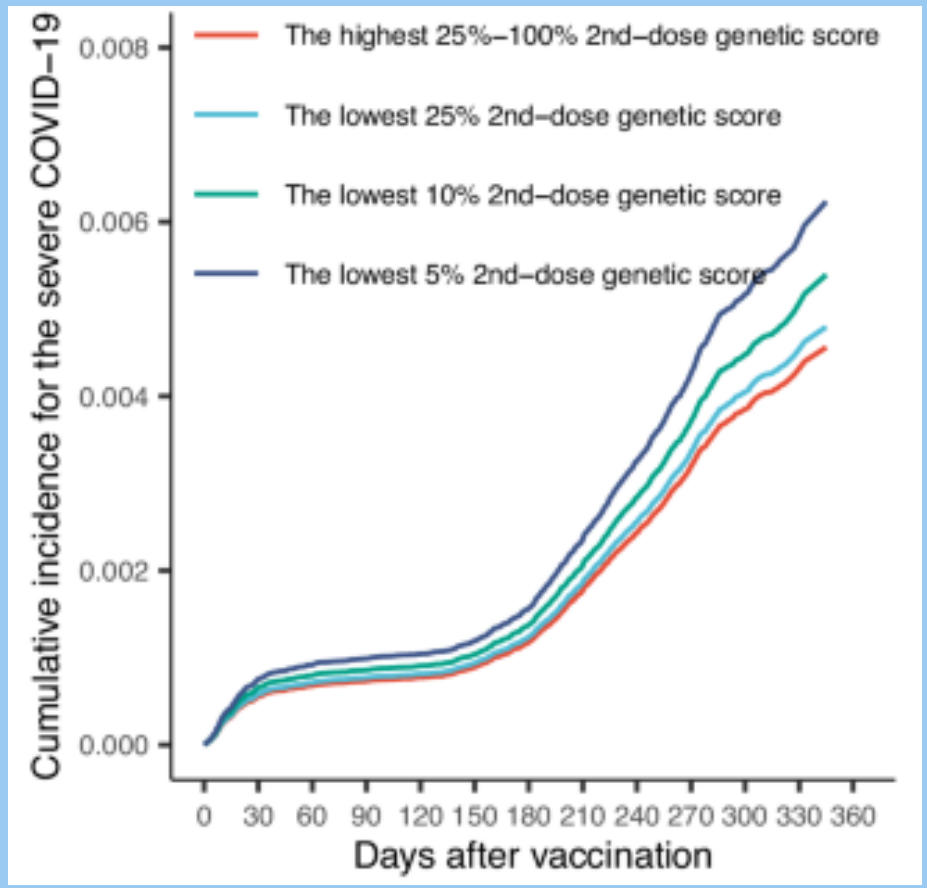
nature communications

Article

https://doi.org/10.1038/s41467-024-48339-5

Relationship between HLA genetic variations, COVID-19 vaccine antibody response, and risk of breakthrough outcomes

But, combined variations in multiple PGx did play an important role in real world settings!!!



Thanks for your listening

PGx in Real-World Data Studies

Professor Daniel Prieto-Alhambra
Dr Junqing (Frank) Xie
University of Oxford