

National Cohorts Office (NCO)

Xueling SIM

Director, NCO

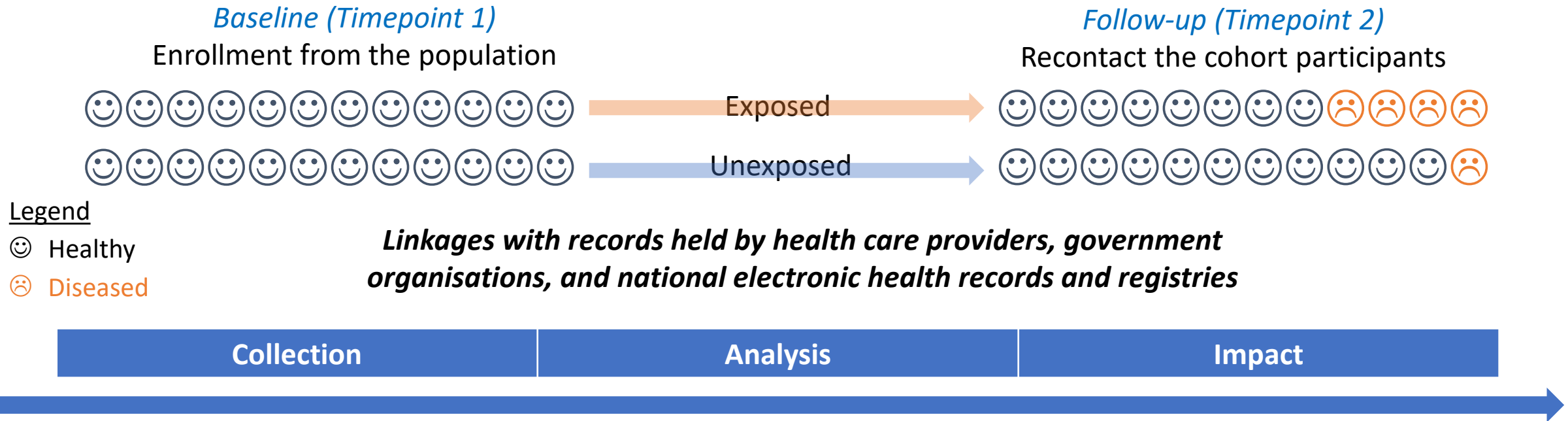
Associate Professor

Saw Swee Hock School of Public Health, National University of Singapore

for SingHealth Duke-NUS AMC Research Council Meeting

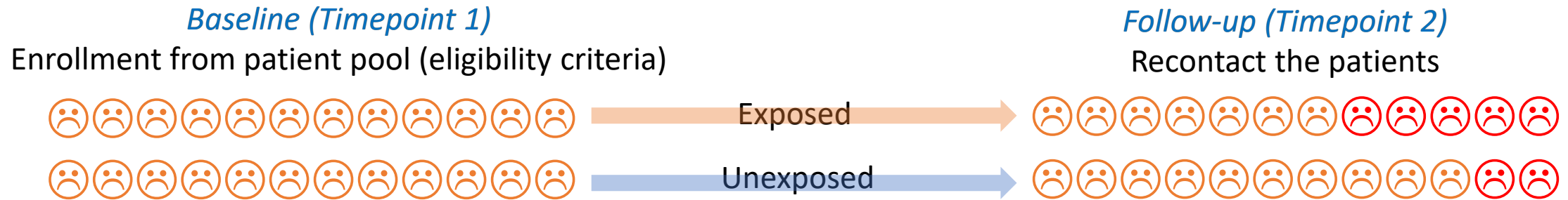
26 Feb 2026

Population Cohorts



- Identify risk factors leading to multiple health status and outcomes (at baseline, they are outcome-free)
- Can study the effects of one exposure on multiple health status and outcomes
- Collect "talking", "tissue", and "tracking" data

Disease Cohorts



Legend

☹ Diseased

☹ Complications

Linkages with records held by health care providers, government organisations, and national electronic health records and registries

Collection

Analysis

Impact

- Identify risk factors leading to disease progression and complications (at baseline, they have the disease)
- Can study the effects of multi-comorbidities
- Collect "talking", "tissue", and "tracking" data

Cohorts and Biobanks

Cohorts

- Epidemiological research design
- Individuals are observed over time for the outcome(s) of interest
- Tracks the natural progression of risk factors

Biobanks

- Large collections of human biological materials, linked to health information, for research
- Legal entities or as part of a legal entity
- With governance mechanisms to allow for third party access

Using cohort data for research

Start with best practices

Participants consent checklist:

- Participants' personal identifiers are retained
- Recontact: Consent is obtained to allow recontact of participants
- Data linkage: Consent is obtained to allow data linkage to other data sources
- Secondary use: Collected data and biological materials may be used for purposes beyond the initial research scope
- Sharing: Data and biological materials may be shared with third partners including commercial companies

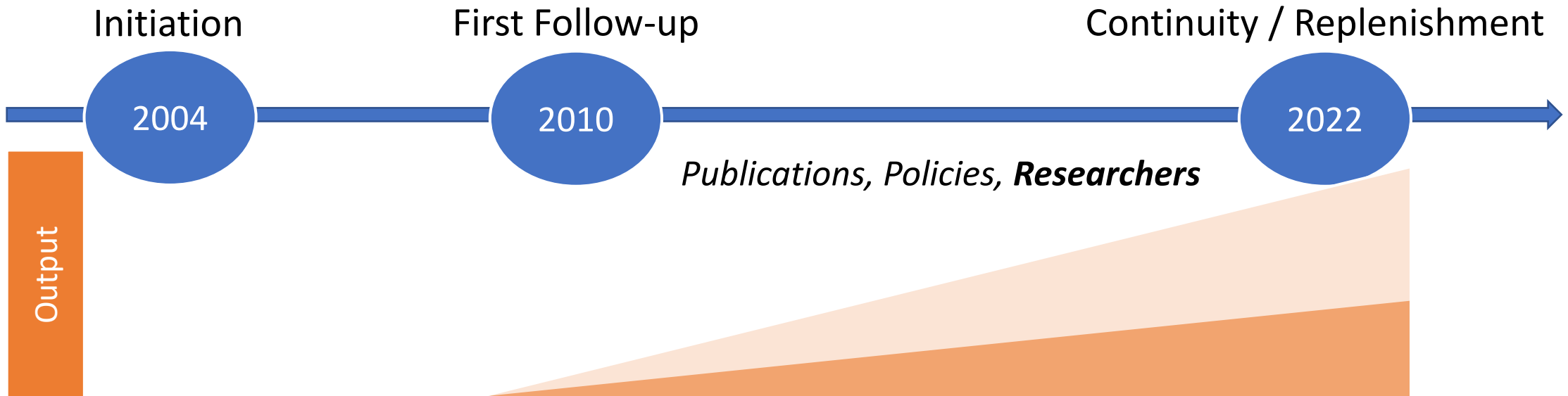
Process Checklist:

- Dataset(s) has the relevant data of interest
- Appropriate consent governing secondary use of data and/or biological materials
- Legal agreements on the use of data and/or biological materials
- Minimise possibility of re-identification of participants through the data
- If linked to external datasets, de-identified data to be provision for research use with the appropriate approvals and security controls
- Cost related to access and analysis

Can we shorten access time & increase output?

Singapore Population Health Studies (SPHS) since 2004

- *Conceptualise cohort aims*
- *Establish sampling frame*
- *Setup operations for recruitment, data collection, biological specimens collection*
- *Follow-up participants*
- *Set up data sharing governance and processes*
- *Generate new data through new methods or stored specimens*
- *Collaborate with other cohorts to expand scope and size*
- *Mentor new generation of researchers; recognition of cohort creators*



Versatile in health outcomes BUT require **long planning time horizon and sustained funding**
Under-utilized due to lack of appropriate infrastructures to maximize the value and benefits derived

Before RIE2025 and establishment of NCO



NO CORE FUNDING

Cohorts require *sustained and significant funding* to conduct recruitment, follow-ups and store biological samples



LACK OF INTEROPERABILITY

Management and harmonisation of data from different cohorts and timepoints, with heterogenous infrastructure/management



LIMITED DATA SHARING

Data is *not easy to find nor well documented* to allow secondary use.
Access process varies and collaborations are hindered by *lack of interoperability*.



LIMITED SUSTAINABILITY AND USE

National Cohorts Office (NCO)

National initiative set up to advocate for and facilitate competency building among cohorts to enhance the long-term biomedical research ecosystem in Singapore

Mission: To maximise utility, value and impact of cohorts

The National Cohorts Office (since Dec 2021)

Ecosystem Facilitator	Cohort Advocator	Competency Building	Methodological Innovation	Funding Administration
Connect across cohorts and other stakeholders to build a collaborative ecosystem	Highlight needs of cohorts at the national level	Establish and encourage best practices	Promote development, evaluation and dissemination of relevant methods	Launch and administer Strategic Cohorts RFP

Build Community. Increase Visibility. Enable Sharing. Maximise Utility

Putting FAIR principles into practice

Principles	Guidelines	Output
Findable	Data and supplementary materials to have identifiers that are findable by human and computational resources	Online data dictionary
Accessible	Data to be deposited in a trusted repository that is accessible and shared	Data access governance
Interoperable	Data to use a formal, accessible and broadly applicable language for knowledge representation	Consensus standards aligned to data dictionary
Reusable	Data and collections have a clear usage licences and provide accurate information on provenance	Cohorts with consent for secondary use and traceable record linkage

GUSTO since 2009

<https://gustodativault.sg/about/gusto>

Navigation: About the Cohort | **Data Catalogue** | Search Topics/Variables | Planned Visit for Future Timepoints

View by Timepoints | View by Curation Status

GUSTO Timepoints
 PW: Pregnancy Week DEL: Delivery Visit DAY1: First Day of Measurements after Delivery
 WK: Postnatal Week M: Postnatal Month YR: Postnatal Year

- Curated, Raw/In progress/Hardcopy
- * Data collection in progress and may require longer processing time
- ** Infant and Child's 3-day Food Diary shares the same dataset as 24-hour Food Recall at M6, M9 & M12

Demography/Cohort Profile

Theme	Topic	Curated	Raw/In progress/Hardcopy
Maternal Demographics	Ethnicity	PW11	
	Education, Work, Income	PW11, PW26, YR5, YR8	YR12, YR13*, YR14*
	Housing, Household members	PW11, PW26, YR5, YR8	YR12, YR13*, YR14*
Paternal Demographics	Marital status, Religion, Citizenship	PW11, PW26, YR5, YR8, YR12	YR13*, YR14*
	Education, Work, Income	PW11, M24, M36	
	Island-wide Daily Average PM2.5 (2009-2013)	PW11	
	Island-wide Daily Average PSI (2009-2013)	PW11	

Keywords
Search across themes and topics

Subject Types
 Mother Child Father

Domains
 Demography/Cohort Profile × Women's Health ×
 Maternal Metabolic and Body Composition ×
 Maternal Sleep, Personality and Mental Health ×
 Children's Health ×
 Child Metabolic and Body Composition ×
 Child Neurodevelopment ×
 Life Events & Social Relationships ×
 Paternal Factors × Imaging × Omics (Mother) ×
 Omics (Child) × Omics (Father) ×

Week 19-21: Fetal Anomaly Scan

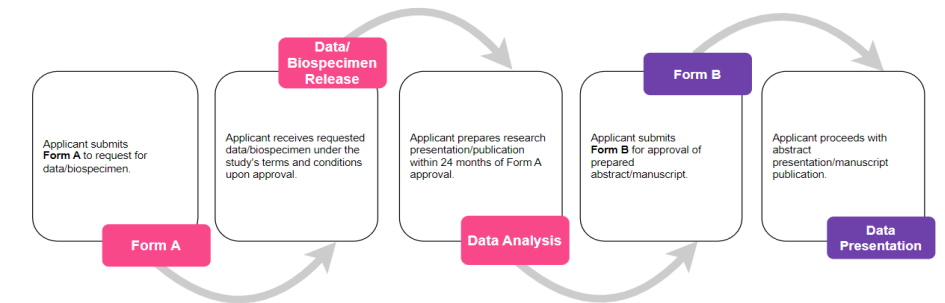
Mother (n=1278)	Child (n=1288)
Main GUSTO (n=1182)	Main GUSTO (n=1182)
IVF (n=86)	IVF (n=86)
10 Pairs of Twins (n=10)	10 Pairs of Twins (n=20)

Lost to follow up

Mother (n=172)	Child (n=172)
Main GUSTO (n=162)	Main GUSTO (n=162)
IVF (n=10)	IVF (n=10)
No Twins (n=0)	No Twins (n=0)

Accessible
Findable, Interoperable

Collaboration Guidelines



Navigation: Catalogues | GUSTO OMOP | Biospecimens | Omics | Data Visualisations | **Publications** | Request for Data | Search Data

8 **GUSTO** Agarwal, P. K.; Xie, H.; Sathyapalan Rema, A. S.; Meaney, M. J.; Godfrey, K. M.; Rajadurai, V. S.; Daniel, L. M. - *Pediatrics and neonatology* - 2024
 BACKGROUND: With increasing acceptance of universal developmental screening in primary care, it is essential to evaluate the local validity and psychometric properties of commonly used questionnaires like the parent-completed Ages and Stages Questionnaires, 3rd Edition (ASQ-3) in identifying developmental delays. The aim of this study is to assess the convergent validity of the ASQ-3 with the Bayley Scales of Infant Development-3rd edition (Bayley-III) in identifying developmental delay in a low...

71 **GUSTO** Nguyen, T. T.; Qian, X.; Ng, E. K. K.; Ong, M. Q. W.; Ngho, Z. M.; Yeo, S. S. P.; Lau, J. M.; Tan, A. P.; Broekman, B. F. P.; Law, E. C.; Gluckman, P. D.; Chong, Y.; Cortese, S.; Meaney, M. J.; Zhou, J. H. - *Journal of the American Academy of Child and Adolescent Psychiatry* - 2024
 OBJECTIVE: It is unclear how the functional brain hierarchy is organized in preschool-aged children, and whether alterations in the brain organization are linked to mental health in this age group. Here, we assessed whether preschool-aged children exhibit a brain organizational structure similar to that of older children, how this structure might change over time, and whether it might reflect mental health. METHOD: This study derived functional gradients using diffusion embedding from resting sta...

71 **GUSTO** Low, P. H. X. - *Journal of Abnormal Psychology* - 2024
 Parenting by lying: The role of lying in children's beliefs.
 How are children socialized to lie? The role of lying given that parents tell various lies to their children for parenting purposes, which is a... how children perceive and interpret the lying behavior around them may be crucial to how they then learn to lie. Yet, we do not know how children's perceptions of different types of parental lies drive this socialization. In a comprehensive birth cohort of parent-child dyads (N = ...)

20 **GUSTO** Alex, A. M.; Aguarte, F.; Botteron, K.; Buss, C.; Chong, Y.; Dager, S. R.; Donald, K. A.; Entringer, S.; Fair, D. A.; Fortier, M. V.; Gilmore, J. H.; Girault, J. B.; Graham, A. M.; Groenewold, N. A.; Hazlett, H.; Lin, W.; Meaney, M. J.; Piven, J.; Qiu, A.; Rasmussen, J. M.; Roos, A.; Schultz, R. T.; Skeide, M. A.; Stein, D. J.; Styner, M.; Thompson, P. M.; Turesky, T. K.; Wadwa, P. D.; Zar, H. J.; Zöllner, L.; de Los Campos, G.; Knickmeyer, R. C.; ENIGMA ORIGINS group - *Nature neuroscience* - 2024

Reusable

Major milestone for health research as UK Government decision enables access to UK Biobank volunteers' GP patient data



[home](#) / [news](#) / major milestone for health research as uk government decision enables access to uk biobank volunteers' gp patient data /

The power of UK Biobank's dataset to advance the diagnosis, treatment and management of many conditions will be increased overnight following this long-awaited decision.

Published:
10 February 2026

Today, 10 February 2026, the UK Government has published a data provision notice, which paves the way for coded GP patient data in England, also known as primary care data, to be shared with consented cohorts like UK Biobank. NHS England will now be responsible, and legally liable, for these data, removing the burden of responsibility and liability from busy GPs.

Having GP data for all of UK Biobank participants will double the number of recorded cases of some health conditions, commonly handled by our GPs. These include arthritis, asthma, dementia, depression, eczema, heart failure and impaired hearing or vision, all of which reduce quality of life, lead to time off work and put significant pressure on the NHS.

This story was covered in:

The Times – [Decades of medical insight unlocked as UK Biobank gets GP records](#)



BBC Radio 4 – [Today Programme](#)

The Mirror – [GPs hand over patient data in move that will 'transform our understanding of disease'](#)

Research Professional News – [Major UK research resources get long-awaited access to GP data](#)

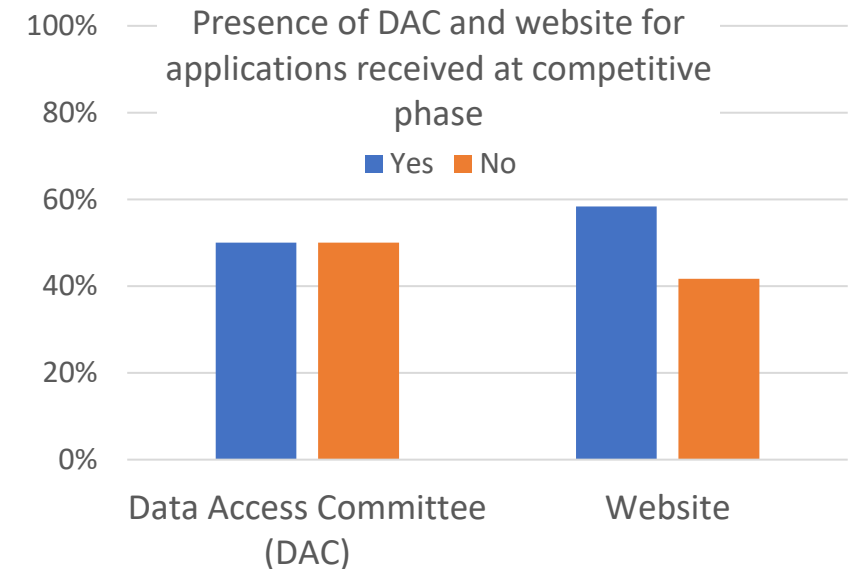
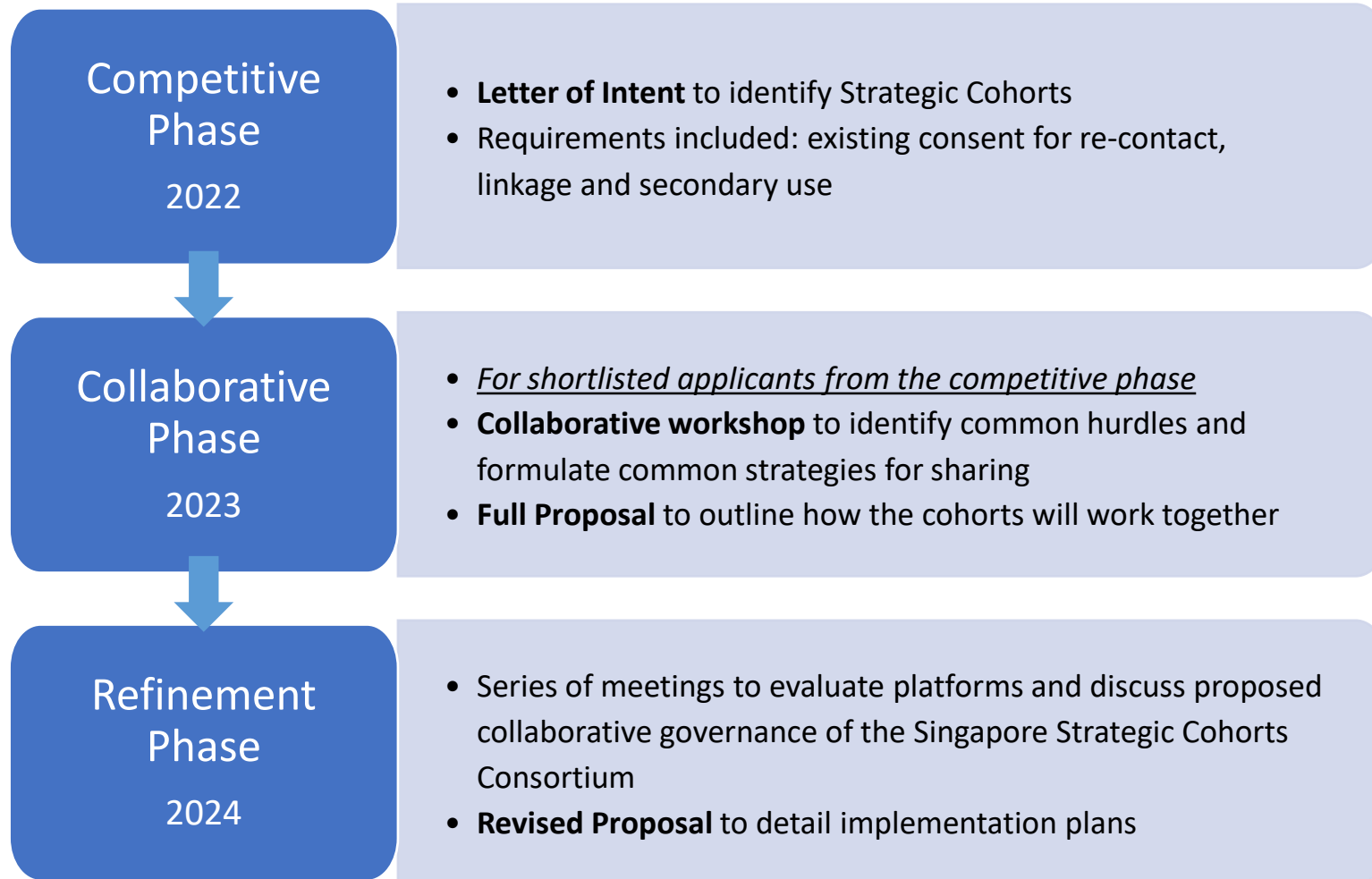
“ Access to our participants' coded GP patient data will be a game changer for research into some of the most common and debilitating diseases. There's no doubt this will revolutionise the power of UK Biobank.

Professor Sir Rory Collins, Principal Investigator and Chief Executive, UK Biobank

“ This is great news! I joined UK Biobank in the hope it may find answers to the causes of disease and so improve the health of future generations. I'm so pleased that my coded GP data will soon be available to UK Biobank approved scientists. With GP care grounded in the community, these data will undoubtedly offer fresh insights and detail for further research into many types of illnesses, all contributing to the aim of improving health outcomes for everyone.

Karen, UK Biobank Participant Advisory Group member

Multi-phased Request for Proposals to identify Strategic Cohorts incentivised for FAIR principles



- **Governance:** only half of the cohorts had a Data Access Committee
- **Visibility:** only 60% of cohorts had a website

Singapore Strategic Cohorts Consortium (SSCC)

NCO has funded and is co-creating an open sharing ecosystem with these cohorts for high-quality data access and utility for health value outcomes at scale and establish Singapore as a data and translation hub

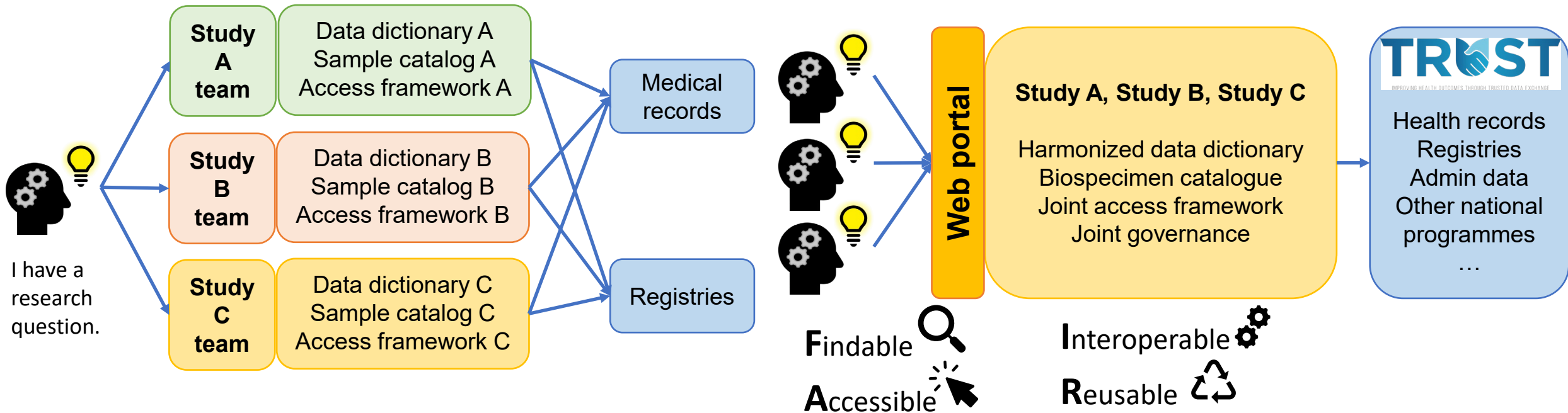
Goal: Single shop-front to query data and biospecimen catalogue to maximise potential for health data research

Strategic Cohort	Cohort PI	Host Institution	Sample size
SG100K	Prof John Chambers	Nanyang Technological University	T1: 100,000
The Singapore Chinese Health Study (SCHS)	Prof Koh Woon Puay	National University of Singapore	T1: 63,000 T2: 52,000 T3: 39,000
Singapore Epidemiology of Eye Diseases Cohort (SEED)	Prof Cheng Ching-Yu	Singapore Eye Research Institute	T1: 10,000 T2: 6,700
Singapore Longitudinal Ageing Studies and Yishun Cohort Study (SLAS/YS)	Prof Roger Ho Chun Man	National University of Singapore	SLAS-1: 2,804 SLAS-2: 3,270 YS: 542

T1 refers to the first timepoint, or baseline. T2 refers to the second timepoint and so on.

Harmonisation and single access framework

Findable, Accessible, Interoperable, Reuseable (FAIR)



Current Model

Lack of core funding, maintenance embedded in competitive funding
Tedious and long process to access and understand data structure



Aspirant Model

Infrastructure funding for maintenance and uplifting to **FAIR principles**
Standardised curation and quality to maximise value faster

NCO Cohort Engagements

March 2024 – May 2025

HELIOS & SG100K



SLAS



NHCS Biobank



SINGHEART

PIONEER



SLP



PHASE & SIGNS



SCHS



YISHUN STUDY



PREDICT



HEALTH@NUS & HCI



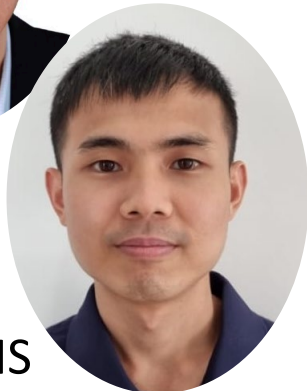
SCORM



GUSTO & S-PRESTO



SEED



SPHS



BREATHE



IWHP



SG-LEADS



Sustaining and maximizing cohorts

- What the custodians were worried about
 - Funding sources and continuity
 - Biological specimen storage
 - Participant recruitment and engagement
 - Cohort longevity – continual growth and succession planning
- What the custodians want to do
 - Generating research outcomes through collaboration, visibility
 - Data enrichment through linkage
 - Making the cohort FAIR

2025 Cohort Enrichment and Use Cases through Linkages Request for Proposals (RFP) Objectives

1

Maximise utility, value and impact of existing population cohorts by uplifting to FAIR¹ principles
(Findable, Accessible, Interoperable and Reusable)

2

Support existing population cohorts to prepare their data and processes to enable linkage and secondary use

3

Enrich and generate new insights from use cases through data linkage of existing population cohorts with other datasets on TRUST²

¹FAIR principles: Findable, Accessible, Interoperable and Reusable. Wilkinson, Mark D., et al. "The FAIR Guiding Principles for scientific data management and stewardship." Scientific data 3.1 (2016): 1-9.

²Trusted Research and Real world-data Utilisation and Sharing Tech (TRUST) is a national linkage platform for health-related research.

Cohorts supported by NCO

Cohort	Cohort Custodian	Host Institution	Baseline Sample size
SG100K	John Chambers	Nanyang Technological University	100,000
The Singapore Chinese Health Study (SCHS)	Koh Woon Puay	National University of Singapore	63,000
Singapore Epidemiology of Eye Diseases Cohort (SEED)	Cheng Ching-Yu	Singapore Eye Research Institute	10,000
Singapore Longitudinal Ageing Studies and Yishun Cohort Study (SLAS/YS)	Roger Ho Chun Man	National University of Singapore	SLAS-1: 2,804 SLAS-2: 3,270 YS: 542
Panel on Health and Ageing of Singaporean Elderly (PHASE) Transitions in Health, Employment, Social Engagement and Inter-generational Transfers in Singapore Study (THE SIGNS Study)	Rahul Malhotra	Duke-NUS	4,990 4,549
The PopulatiON HEalth and Eye Disease PRofile in Elderly Singaporeans (PIONEER) study	Ecosse Lamoureux	Singapore Eye Research Institute	2,643
Singapore Life Panel (SLP)	Paulin Straughan	Singapore Management University	14,400
Health@ NUS Prospective Cohort Initiative	Falk Müller-Riemenschneider	National University of Singapore	3,806 (on-going)
SingHEART	Yeo Khung Keong	National Heart Centre Singapore	1,129 (on-going)
Singapore Translational Cancer Consortium (STCC)	Chng Wee Joo	Consortium for Clinical Research and Innovation, Singapore	NA (on-going)
Understanding Tumour Predisposition in Asia (UTOPIA)	Joanne Ngeow	National Cancer Centre Singapore	5,129 (on-going)

Build community, enable sharing, increase visibility, maximise utility

Website

Increase visibility and signpost researchers to cohort resources. Create platform to share findings.

Landscape scan

Continue engaging with more cohorts to discover strengths and needs while promulgating best practices



Slides credit: Slidesgo

Harmonisation

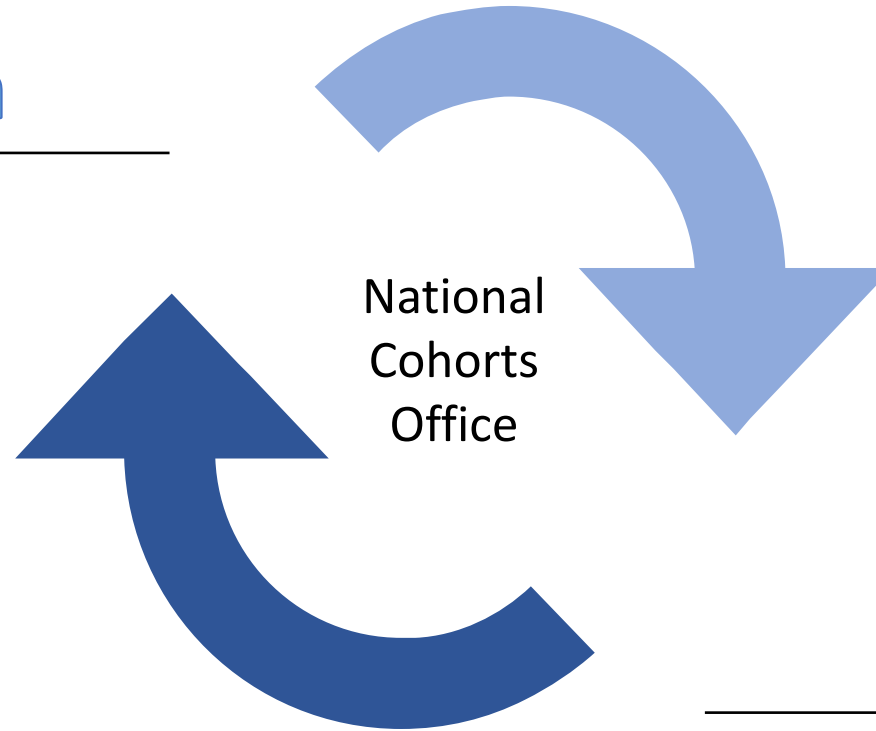
Facilitate harmonisation via sharing of tools and standards to open up more data for secondary use

Build user base

Support more use cases to demonstrate the value of cohort studies and build critical mass of researchers

Core Collection

- Motivate custodian to organise and uplift the cohort for future secondary research use
- Uplift to FAIR principles to increase utility and enhance collaborations through open sharing mechanism
- Increase diversity and multi-domain data



- Enhance existing datasets through linkage
- Coordinate with funders to align new datasets to FAIR principles (effective data management and data sharing)
- Establish Community of Practice to entrench best practices
- Encourage and incentivise greater Participant and Public Involvement and Engagement (PPIE)

Enhanced Data Use

Drive community-consensus data standards to increase interoperability

Consolidate data dictionaries into a Question Bank with harmonized codes via a national Question Bank.

Promote stakeholder (researchers, policy makers, industry and public) engagement

Summary

- Long term investment
 - It takes time for cohorts to mature as research infrastructure; FAIR principles to shorten time to impact
- National service
 - It is a lot of work to make cohorts FAIR; custodians need incentivisation and recognition
- Reciprocity between custodians and users
 - Increase user base (secondary use) to enrich/enhance the cohorts for greater and more diverse impact