2021 Census Cohort Study

Louisa Blackwell & Nicky Rogers

Presentation to NSDEC 27 April 2022

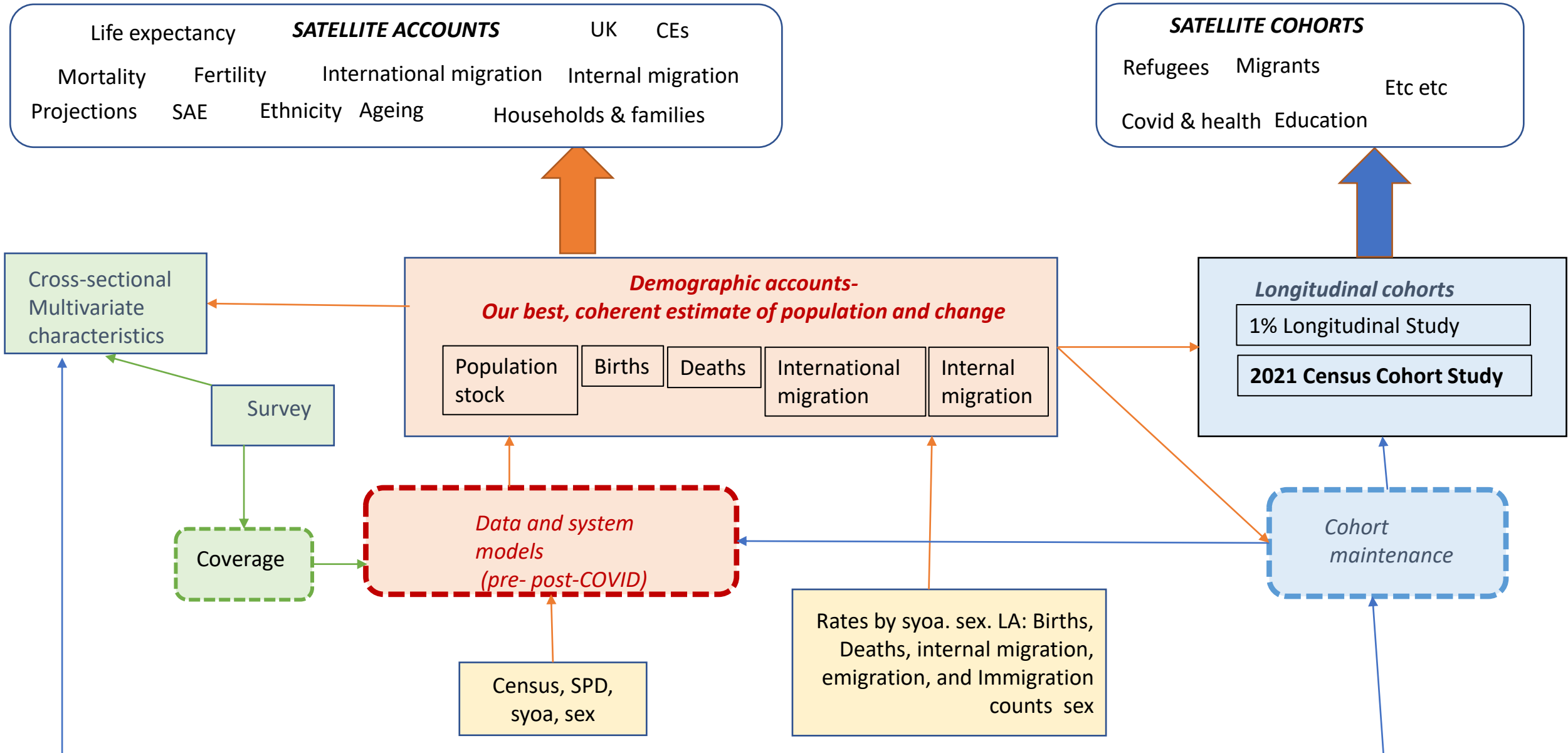Office for **National Statistics**

Outline

- Overview and Timeline

- Context and rationale for 2021 Census Cohort Study

- Proof of Concept

- Ethical principles

- Feedback and guidance

Overview and Timeline

2022

2023

Apr   Jul   Dec   Jan   Apr   Jun

Introduce to NSDEC

PoC to NSDEC

NSDEC Autumn Update

PoC E&W Population

NSDEC Full Cohort

Spring publication – PoC decision

Prepare for 2023 Recommendation

Incorporate NSDEC Feedback

Develop user acceptability strategy and engage
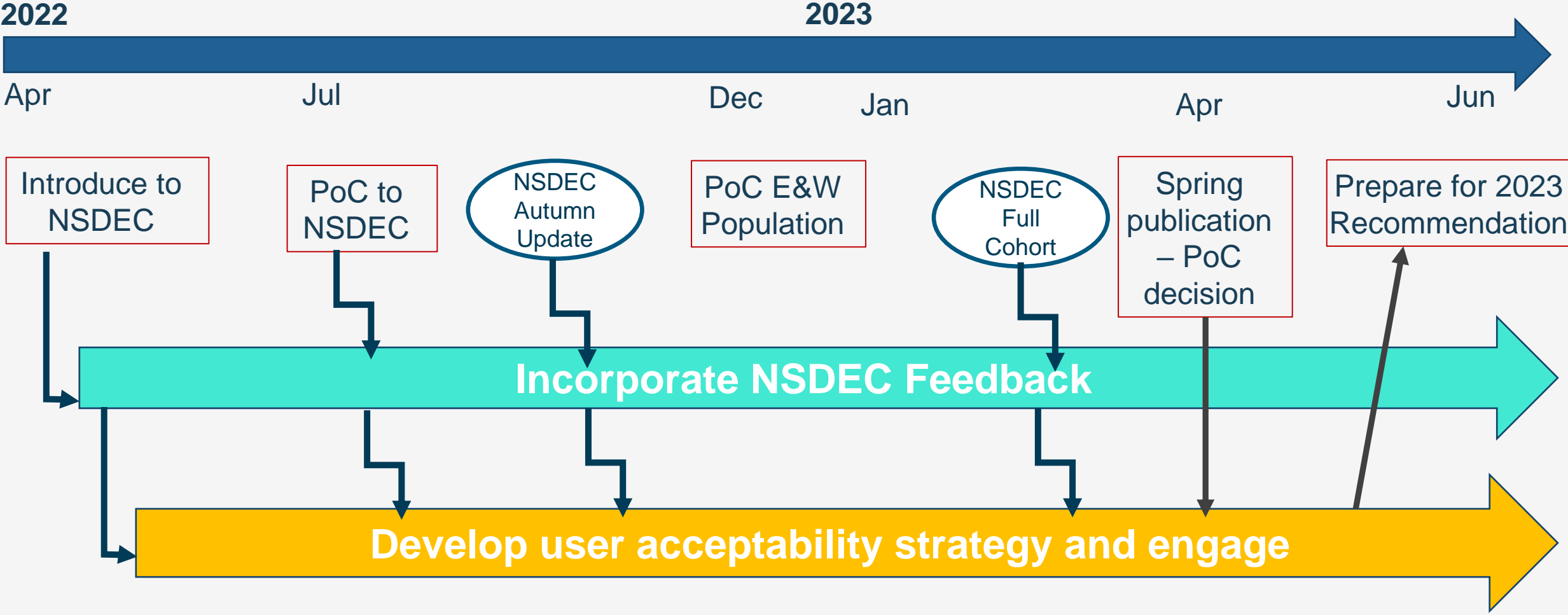
Hybrid Statistics System

SATELLITE ACCOUNTS

Life expectancy · Mortality · Fertility · International migration · UK · CEs · Internal migration · Projections · SAE · Ethnicity · Ageing · Households & families

SATELLITE COHORTS

Refugees · Migrants · Covid & health · Education · Etc etc

Cross-sectional Multivariate characteristics

Survey

Coverage

Demographic accounts-
Our best, coherent estimate of population and change

Population stock · Births · Deaths · International migration · Internal migration

Data and system models (pre- post-COVID)

Census, SPD, syoa, sex

Rates by syoa. sex. LA: Births, Deaths, internal migration, emigration, and Immigration counts sex

Longitudinal cohorts

1% Longitudinal Study

2021 Census Cohort Study

Cohort maintenance

Coherence and benchmarking: the virtuous statistical triangle

- Need all three to provide robust solution to population estimation
- Benchmarking happens between these statistical products

Behaviour through time
- Demographic trends and behaviours
- Data and system models

Dynamic Population Model (DPM) – coherence

Longitudinal microdata
- Census, admin and survey-based prospective admin data linkage

Cross-sectional microdata
- Statistical Population Dataset (SPD)
- Survey data

2021 Census Cohort: Rolled forward population

Aim: To give the best 'microdata' representation of the E&W population over time.

Objectives:

- **Statistically controlled dataset – feedback loop with DPM (signals + weights) that acts as a benchmark study**

- Use 2021 Census as a starting base – good coverage and high response

- Allow cohort replenishment through inclusion of births and immigrations + flagged deaths and emigrations

- Report more granular analyses than is possible with 1% LS

- Only high quality admin data used to create new records for immigrants

- Ensure data inclusivity of marginal or 'invisible' population groups e.g. refugees, NEETS

Focus on design principles: cohort maintenance

1. **Why?** Importance of cohort maintenance

2. **What? When?** Maintaining the representativeness of the 2021 Census Cohort

3. **How?** Capturing migration

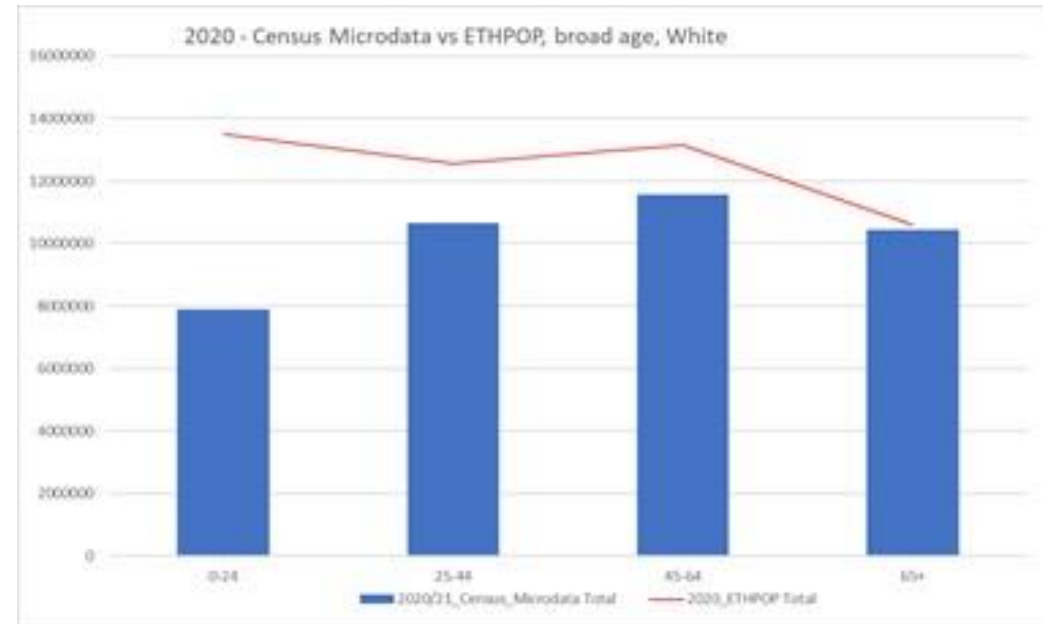# Non-replenished cohort by broad age; White
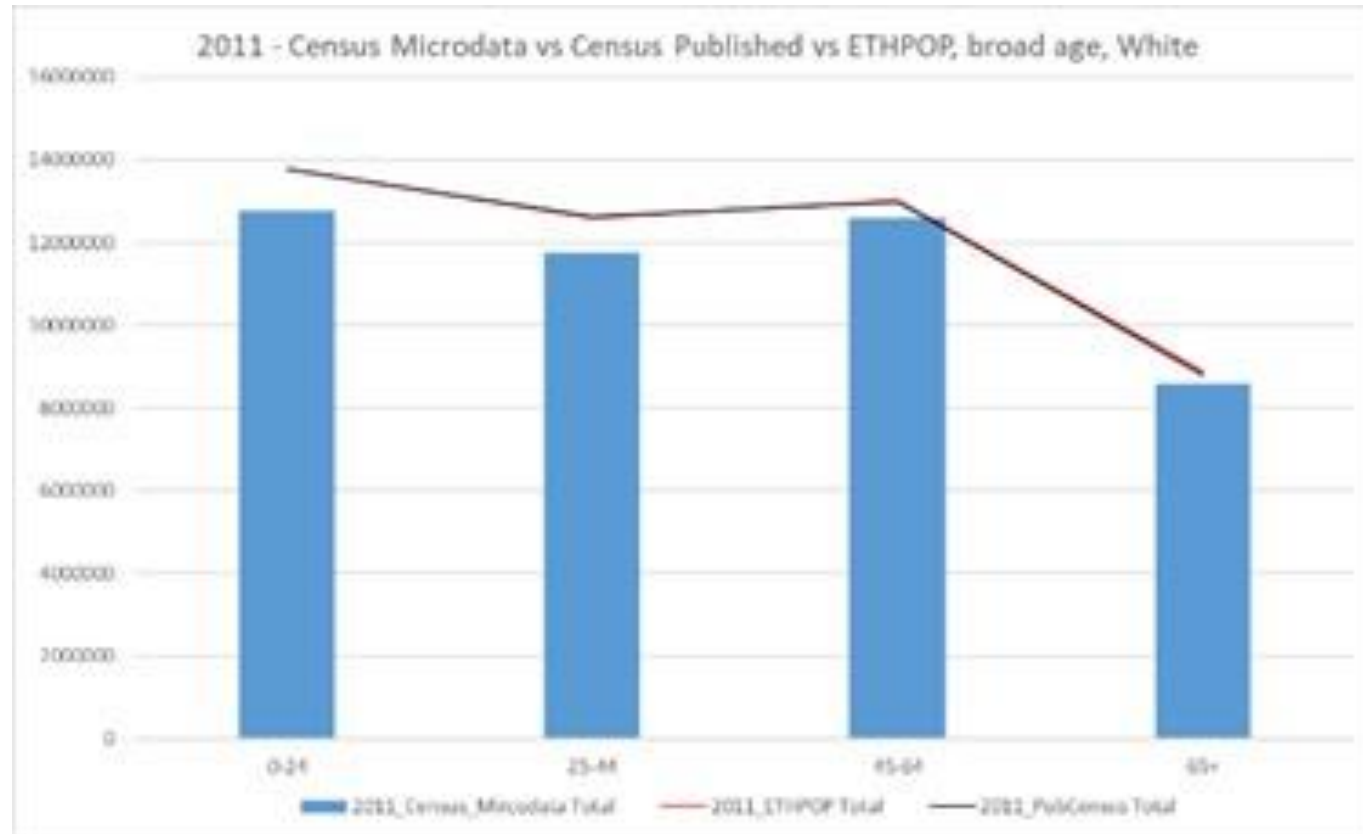
2011

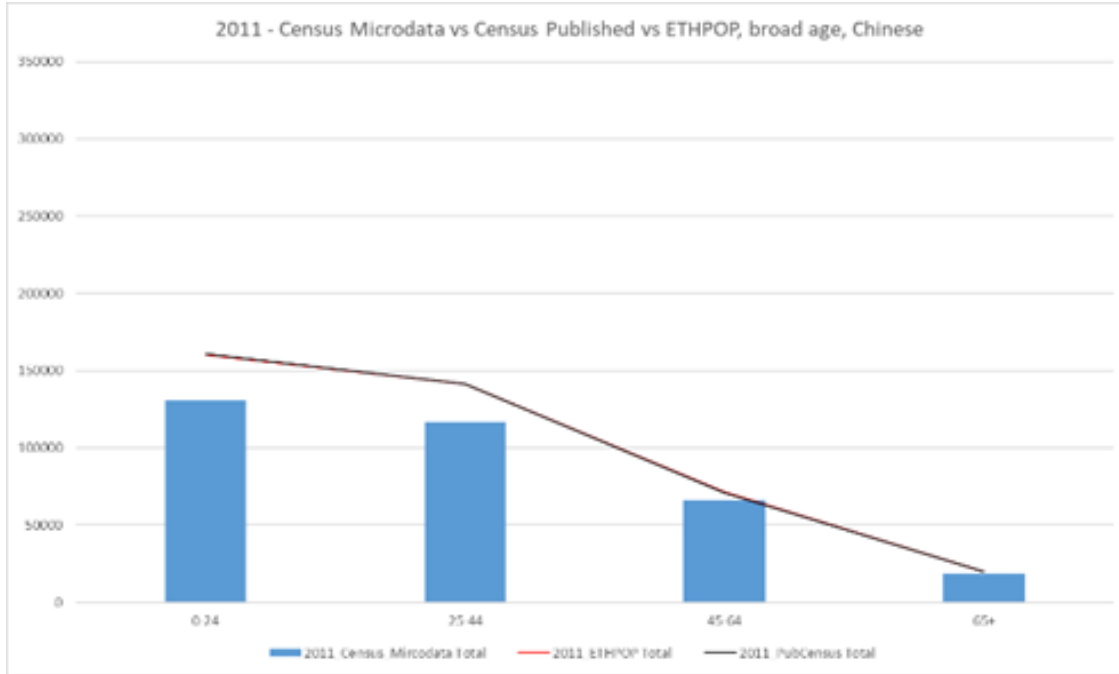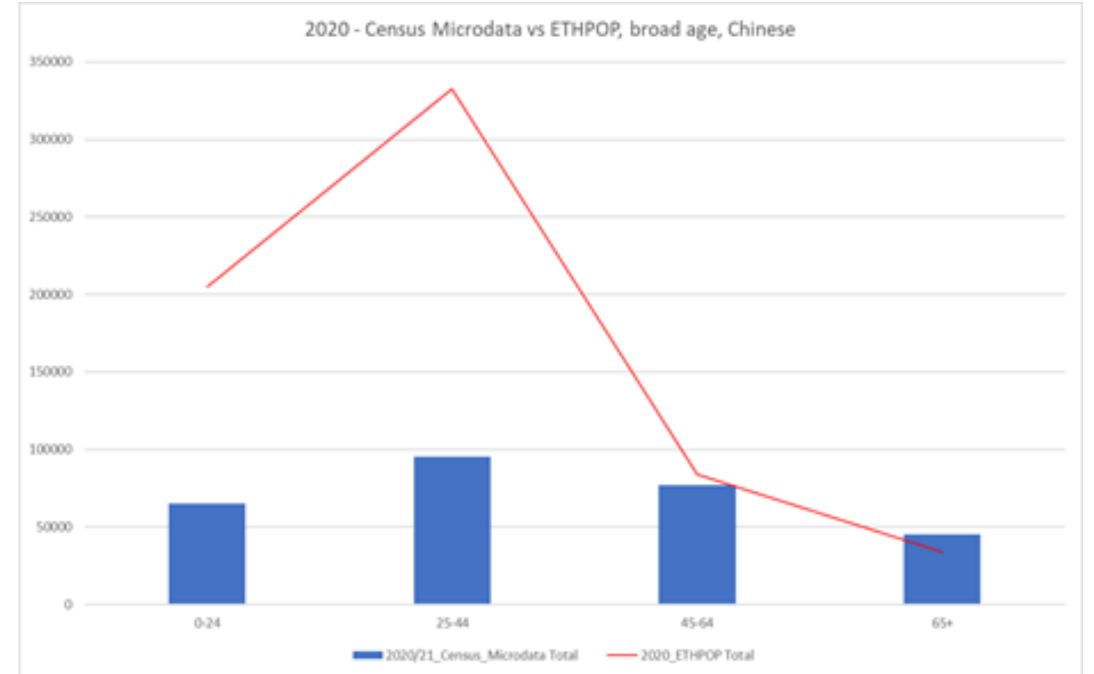# Non-replenished cohort by broad age; White
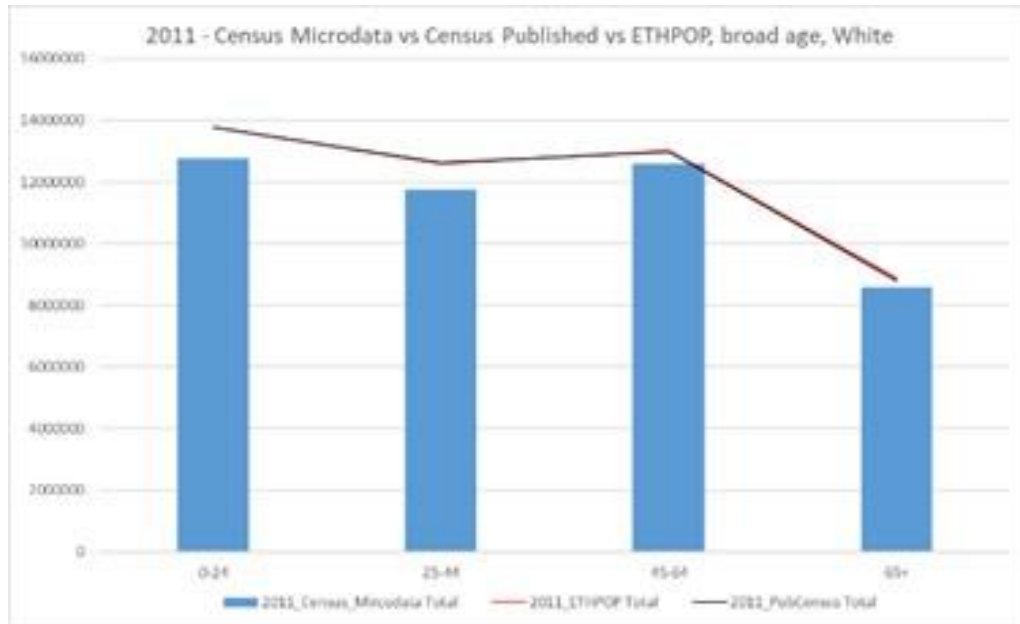
2011

2020

Non-replenished cohort by broad age; Chinese
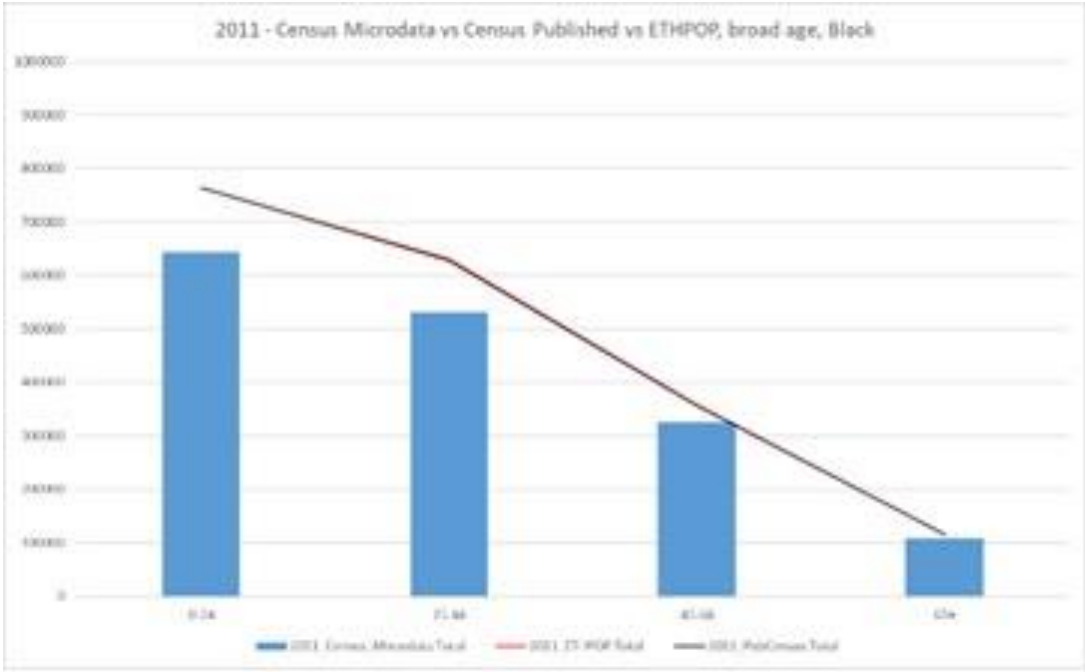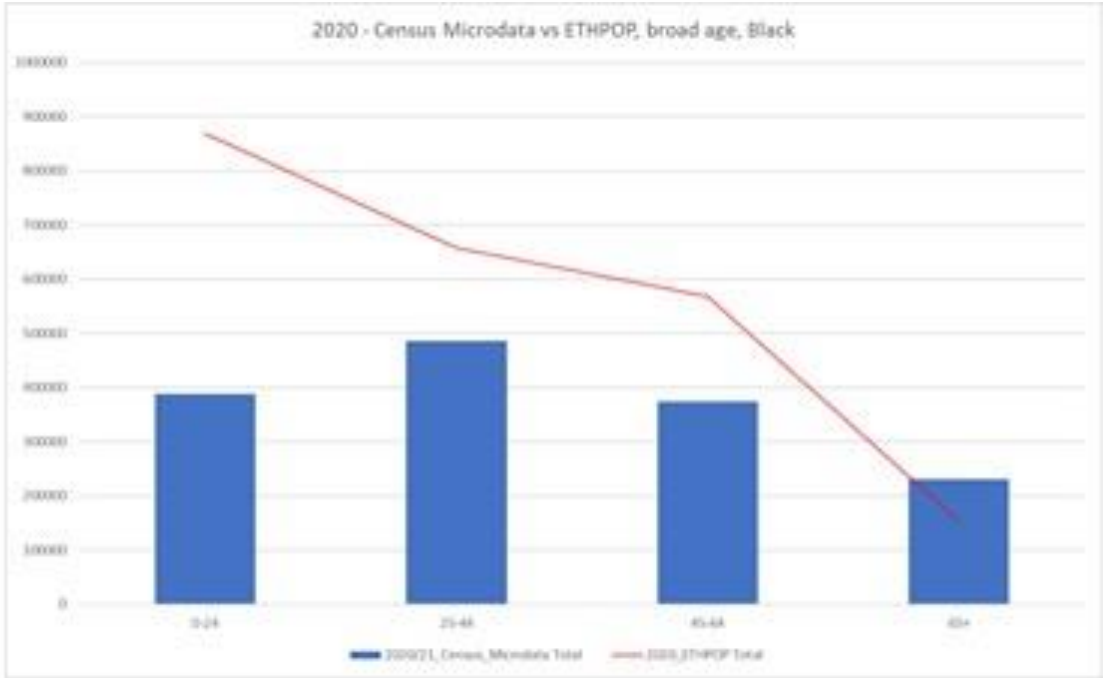

2011

2020

Non-replenished cohort by broad age; Black

2011

2020

Cohort development

Flag potential emigrants

Adjust using weights

Admin Microdata

Undercount
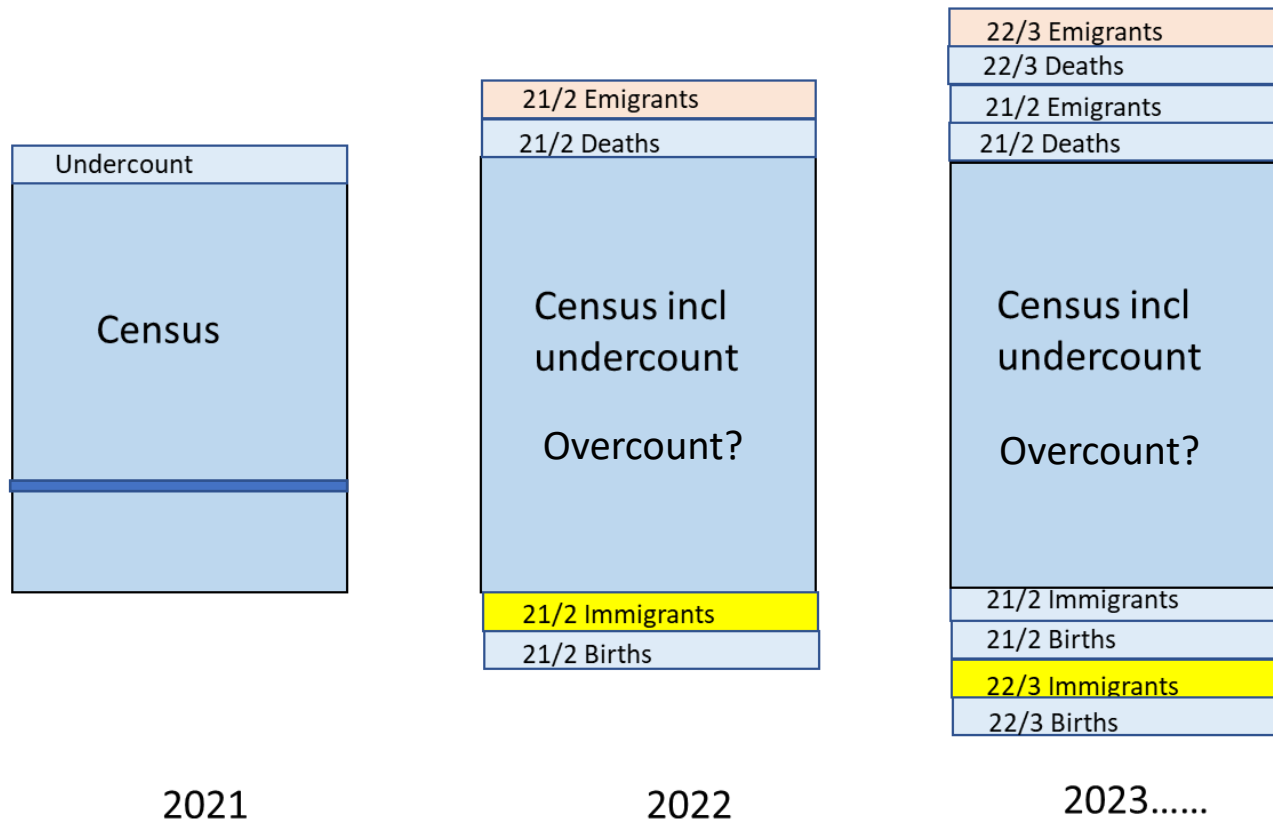
Over count?

2021

Cohort development

Flag potential emigrants

Adjust using weights

Admin Microdata

Undercount

Census

2021

21/2 Emigrants
21/2 Deaths

Census incl undercount

Overcount?

21/2 Immigrants
21/2 Births

2022

Cohort development

2021

Undercount

Census

2022

21/2 Emigrants
21/2 Deaths

Census incl undercount

Overcount?

21/2 Immigrants
21/2 Births

2023……

22/3 Emigrants
22/3 Deaths
21/2 Emigrants
21/2 Deaths

Census incl undercount

Overcount?

21/2 Immigrants
21/2 Births
22/3 Immigrants
22/3 Births

Flag potential emigrants

Adjust using weights

Admin Microdata

Cohort maintenance function. Weights then adds immigrants and flags candidate then actual emigrants, controlled to:
1. Provisional monthly estimates from Demographic Accounts
2. Finalised estimates from Demographic Accounts
Adds births, flags deaths
Cross-border flows tba

Public benefit in providing evidence on much granular basis

Potential to contribute towards informing policy, planning and delivery of local services. Identify associated impacts on society, businesses, local economies and the environment not possible through existing data sources

- Identifying populations at risk to estimate life expectancy and healthy life expectancy estimates by characteristics and occupation.

- Changes in morbidity and mortality at a local area across population groups e.g. by ethnicity, occupation, industry, children in care

- Experiences in life, set against characteristics, lived experiences and behaviours at local level levels.

- Build capability to produce fully inclusive statistics on the whole population, not just private households, but people living in communal establishments, travellers, circular migrants, homeless, children living in care etc

- Levelling Up – Targeted policies and resources to support communities and economic growth - Insight into where local and national services and infrastructure (transport and construction) are needed most

Confidentiality/Data Security

- Draw lessons from ONS Longitudinal Study
- Use for research and statistical purposes only. Estimation used as there will be under-coverage
- Study is not concerned with individuals; identifying information is being used for linking only
- Details of the data shares published on ONS website with information on how we use data and how we keep it secure
- Separation of roles for access to identifiable data and de-identified data
- Analysis dataset will be anonymised
- NHS data ONS receives is not subject to opt-outs and consent of individuals is not required as the data being used has already been collected
- Data held within the secure DAP project area for as long as they are needed for ONS research and statistical purposes
- Published research will have been through strict disclosure control procedures
- Full Cohort: Project Board to endorse research proposals. User support team to apply disclosure control and guidance on use of data. Only approved researchers able to access data

Public views and engagement

- Learn from ONS & HO engagement on public acceptability of data linkage

- Learn from 2021 Census and engage with community groups on public and statistical benefits of such a study

- Communicate outcomes from proof of concept and engage with key audiences and community groups

- Respond to feedback transparently by publishing outcomes from any engagement for the full cohort study

Legal compliance

- Undertaking a DPIA for proof of concept

- Ensure data and methods we employ are consistent with legal requirements under data protection legislation

- Legal gateways exist to receive and process the data (SRSA, Census Act 1920)

- Engage with legal team and data suppliers to ensure data sharing agreements reflect planned use of data

Feedback from NSDEC

- We welcome feedback and guidance at this early stage

- We plan to bring an application to NSDEC in July for the Proof of Concept (PoC)

- We plan to share outcomes from the PoC with the Committee

- And also regularly engage to the develop of the full 2021 Census Cohort Study, if the PoC is successful