

Administrative and field paradata sources for quality assurance and contingency approaches, Census 2021, England & Wales

Cal Ghee
Office for National Statistics



Overview

- Plans v reality
- Summary of 3 areas of administrative and paradata usage
 - Babies and young children (administrative)
 - 3-15 year olds (administrative)
 - Alternative Household Estimate (AHE) (administrative and paradata)
- Lessons learned

Notes:

- UK census results are estimates, not counts. Estimates are calculated using Dual System Estimation (DSE) capture-recapture method, using post-enumeration survey
- UK census is self-enumeration, with field follow-up of non-responding addresses
- This presentation covers England & Wales only

Plans v reality

- Examples of uses in the standard design
 - preparing for collection ✓
 - quality assurance/validation of final estimates ✓ ✓
 - estimating the population in communal establishments ✓
- Preparation to use in contingency situations
 - including scenarios of low response overall ✗
 - in particular areas ✗
 - for certain population groups ✓
- Being prepared to use for other contingencies ✓

Examples of uses of administrative data

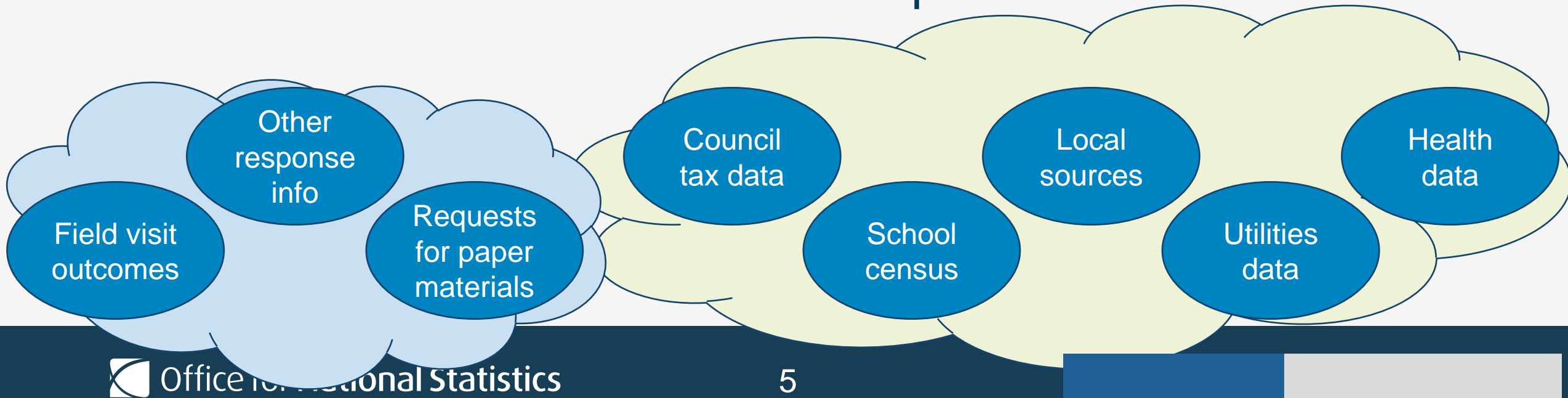
- QA showed estimates to be implausible
- Used administrative sources to calibrate DSE estimates
 - Ages 0-2, every area
 - Ages 3-15, only certain areas
 - Had assured quality administrative data for Wales due to different administrative uses
 - England regions data not of same quality, so used calibration levels from Wales corrections to apply to the one English region where the need was demonstrated

Births and deaths data,
health data used as
upper limit

Combined school
census and private
school data

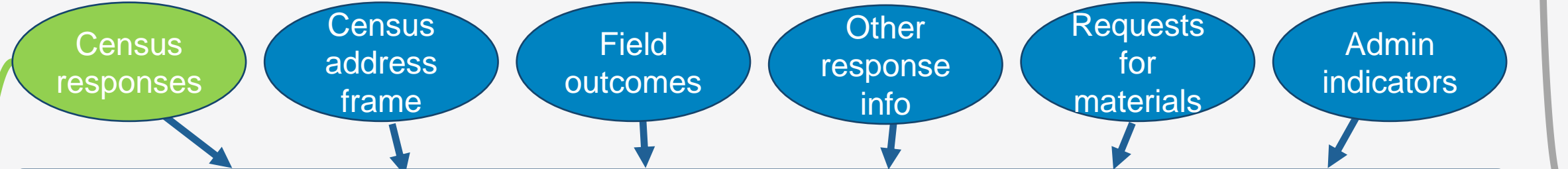
Alternative household estimate (AHE)

- Census estimation method (DSE) has underlying assumptions: risk bias in results if assumptions not met
- AHE is an independent calculation of occupied households to check DSE assumptions



Responses to address frame: 94% (91pp with usual residents, 3pp confirm no usual residents) **6% NR**

Responses with usual residents: 91% **6% NR** 3% vac



Mix of **deterministic rules** and calculation of observed **occupancy rates** in similar responding cases, including using administrative data indicators of occupancy and vacancy

Assessed to be occupied by usual residents (44%)		Assessed to be vacant of usual residents (56%)	
13% rules	87% occupancy rates	68% deterministic rules	32% occupancy rates

Response rate (occupied by usual residents): 97% **3%** vac

Whole address frame, occupied by usual residents: 94% Vac: **6%**

AHE v DSE

- Was used to alter the final household estimates for a few areas
 - For 95% of Local Authorities (NUTS 3), our separate DSE and AHE estimates were comparable
- Very useful tool QA tool for reassuring users

Learning from Census 2021 (1)

- Be prepared
 - Get data ready in advance
 - Don't underestimate preparation time – requirements, acquisition, engineering, platforms, access
 - Don't underestimate data complexities – georeferencing, matching and linkage
 - Prioritising which sources and aspects of sources are most useful is difficult

Learning from Census 2021 (2)

- Circular use: if use in preparing data, beware how you use it to validate the same data
- The importance of accurate linkage
- Covid impact
- Admin data quality caveats
 - Currency
 - Coherence
 - Addressing and georeferencing

Be prepared:
including
understanding and
explaining
apparent
contradictions in
admin data
indicators

Thank you

Cal Ghee

Office for National Statistics
2021CensusStatsDesign@ons.gov.uk

