

Improving Census occupancy determination and person imputation



The potential of administrative data for the Census

Australian Bureau of Statistics
Informing Australia's important decisions



What is the Census?



- ▶ The Census of Population and Housing is a rich snapshot of all people in the country on Census night and is the leading source of information for small population groups and areas.
- ▶ The Census underpins data that informs the planning and delivery of Government and community services, business decisions and is a key source for important academic research.

Administrative data research for the 2021 Census



The ABS is researching the potential of admin data to:

- ▶ Improve Census data quality
- ▶ Add new information to the Census
- ▶ Reduce the cost and burden of the Census



Administrative data research for the 2021 Census



This research investigates how we could **improve the Census count** by using administrative data to:

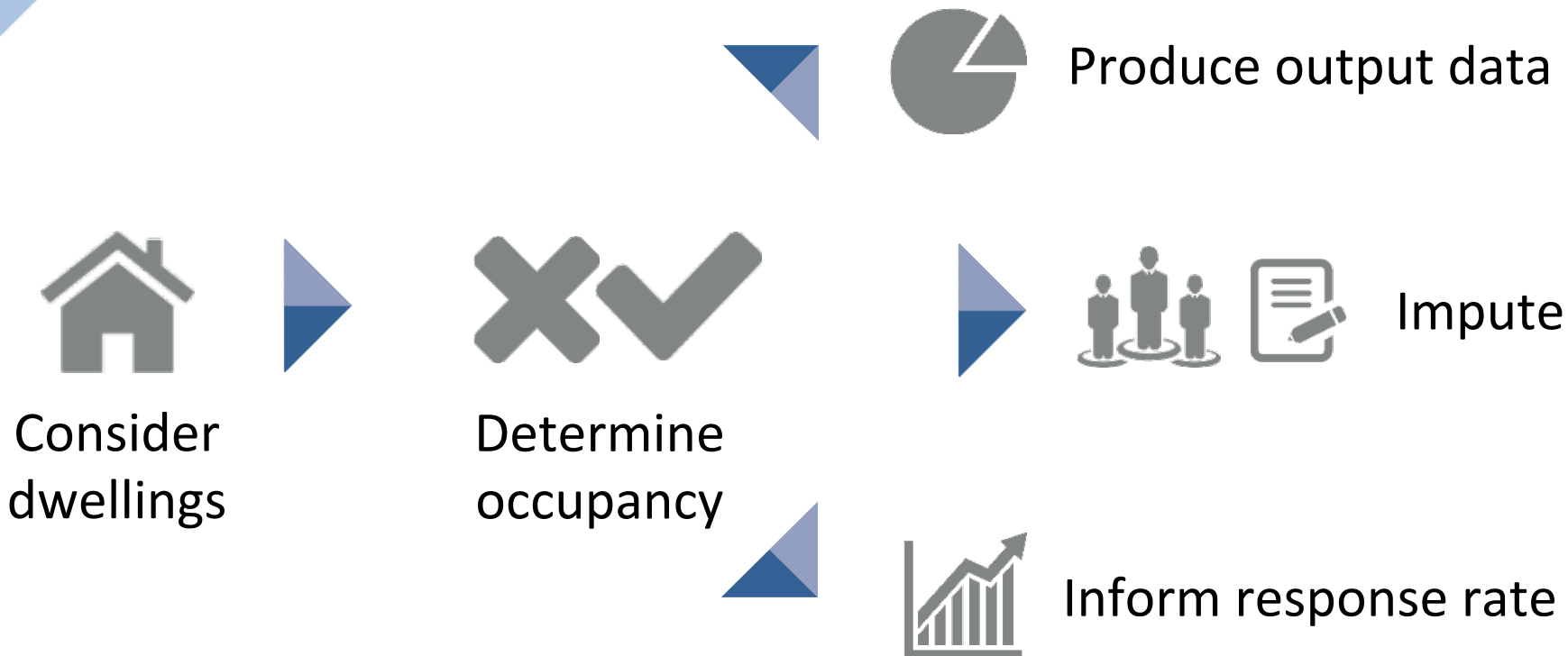
1. Improve our occupancy determination
2. Improve our imputation for non-responding dwellings

Public Trust and Privacy



- ▶ Building public trust and ensuring privacy are important pieces of this work
- ▶ The 2021 Census will be conducting a Privacy Impact Assessment which will address any plans to use administrative data such as those discussed here
- ▶ We will be updating our website as we progress (see links at end of presentation)

What are occupancy and imputation?



What happened in 2016?



- ▶ The Census Independent Panel said that the quality of the 2016 Census was comparable to that of 2011
- ▶ While the quality was high overall, there were...
 - Too many dwellings determined to be occupied
 - Too many people imputed
 - Not the right distribution of people imputed
- ▶ Response rate was “actually” higher than reported



- ▶ Census and related data
 - e.g. 2016 Census and Post Enumeration Survey
- ▶ Government data
 - e.g. Taxation, Social services and Medicare enrolments
- ▶ Other data
 - e.g. rental data, utility connections

1. Occupancy determination

- ▶ Historically, occupancy determination has relied on intelligence from field staff
- ▶ As our population changes (accommodation types, increasing mobility), the Census moves online, and the number of field staff reduces, determining occupancy becomes more challenging
- ▶ Admin data provides another source of information to provide insight into how likely a dwelling is to have been occupied at the time of Census

1. Occupancy determination

- ▶ Admin data enables us to produce a score (between 0 and 1) of how likely it is that a dwelling was occupied at the time of Census



e.g. **0.8**

Inform field
operations



Improve
output data



1. Occupancy determination

- ▶ The Post Enumeration Survey (PES) estimated that we incorrectly determined 44% of non-responding dwellings to be occupied (representing about 500,000 people)
- ▶ This hasn't affected our population estimates because we adjust for any over or under-counting using the PES
- ▶ Most small area Census counts aren't much affected either, but a number of areas, particularly with secure apartments, have higher counts than they should

1. Occupancy determination



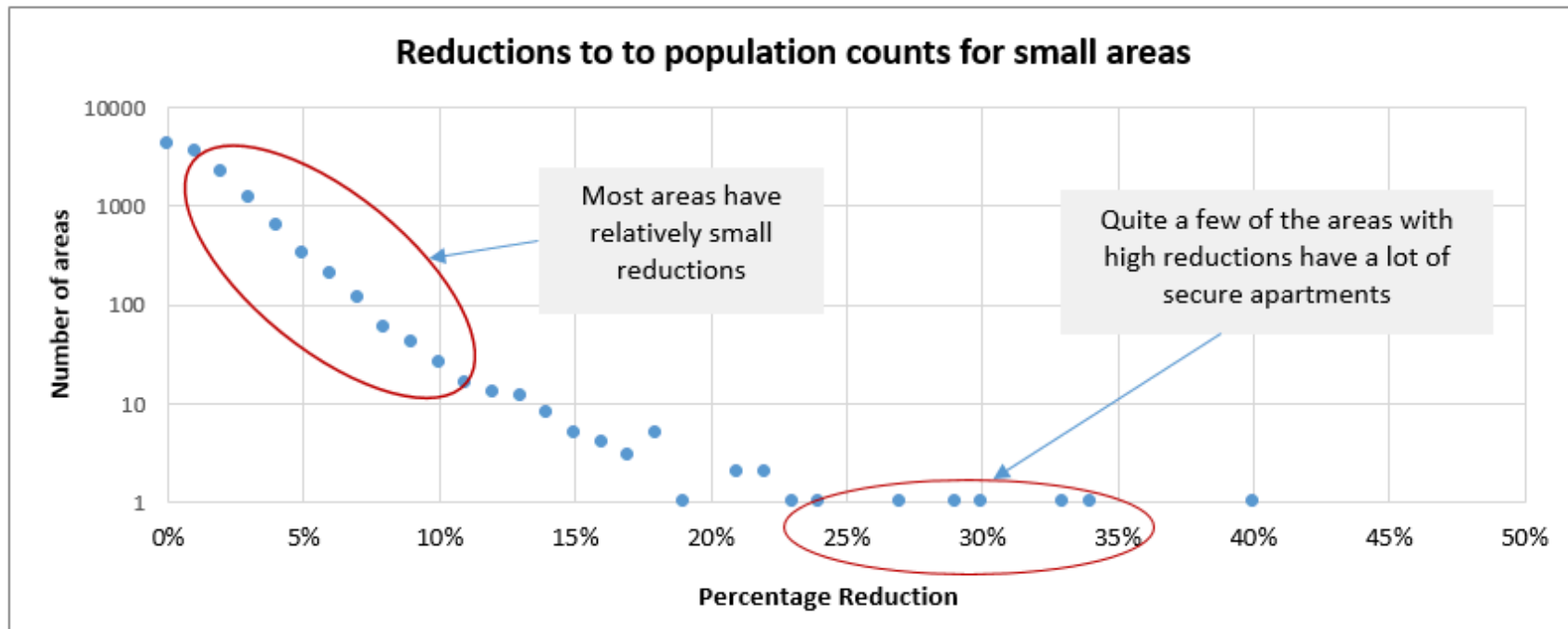
- ▶ We can reduce the number of dwellings we determine as occupied by any amount we want by setting a particular threshold for the occupancy score
- ▶ A key methodological question we are working through is what threshold to use

1. Occupancy determination



Occupancy score threshold	Reduction in imputed dwellings relative to census 2016	Estimated 2016 response rate (2016 response rate was 94.8%)	Estimated 2016 net undercount (2016 net undercount was 1%)
0.5	24%	95.8 %	2%
0.74	44%	96.7 %	3%

1. Occupancy determination



2. Imputation for non-responding dwellings



- ▶ Historically, we have used a hot-decking, nearest neighbour donor approach

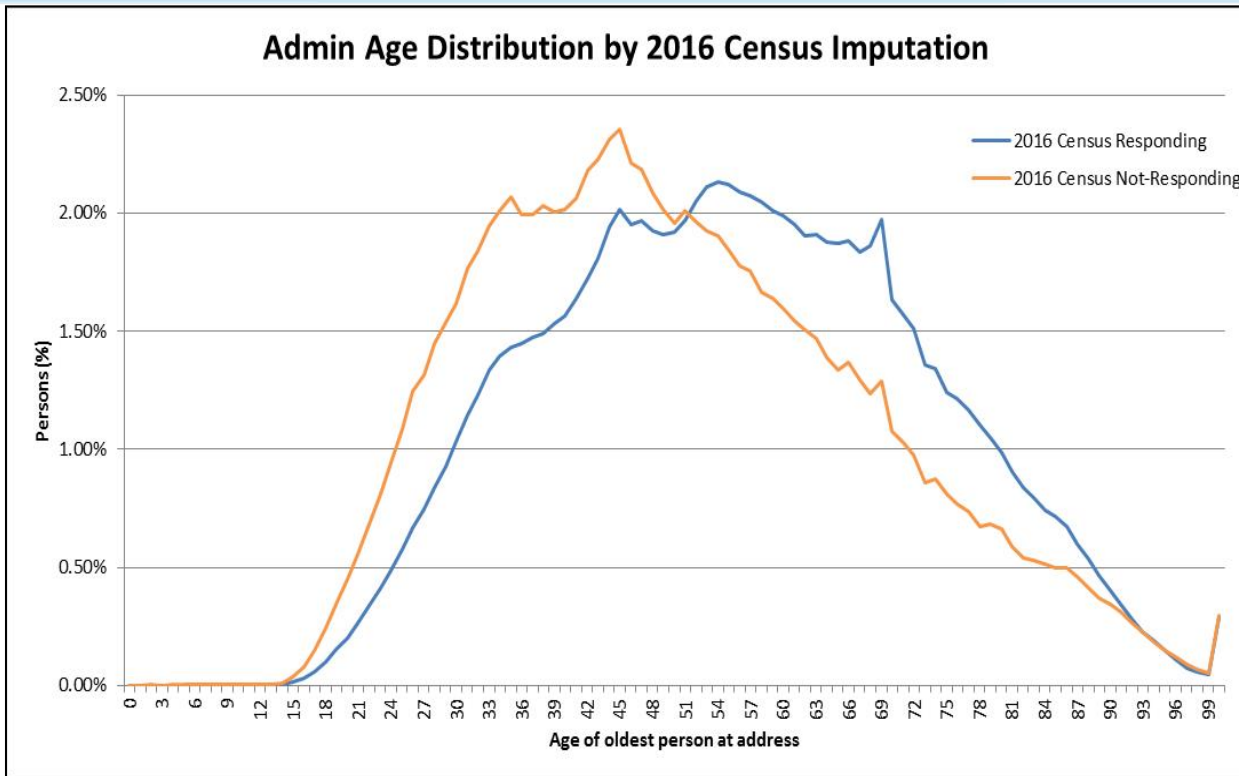


2. Imputation for non-responding dwellings



- ▶ This relies on the assumption that the responding population is the same (statistically) as the non-responding population
- ▶ Admin data now suggests that the distribution of the responding population is different to that of the non-responding population

2. Imputation for non-responding dwellings

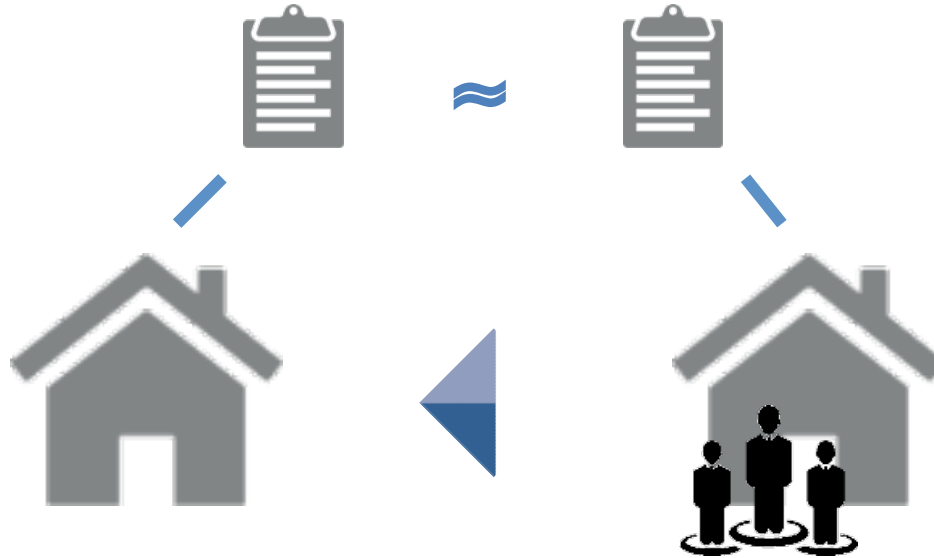


2. Imputation for non-responding dwellings

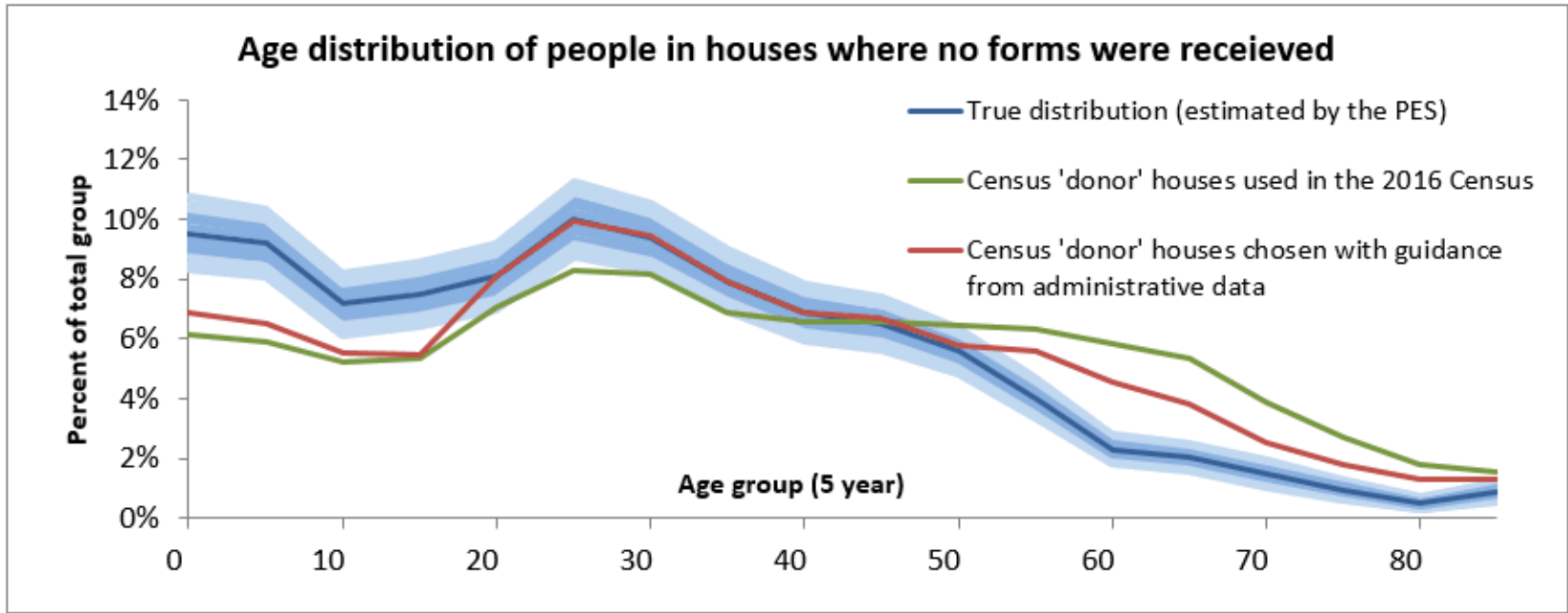


- ▶ In 2016, the Post Enumeration Survey indicated that we didn't impute the right distribution of people. Specifically, not enough young people and not enough males
- ▶ By choosing hot-decking donor households based on administrative data variables, we can make the imputed population closer to reality

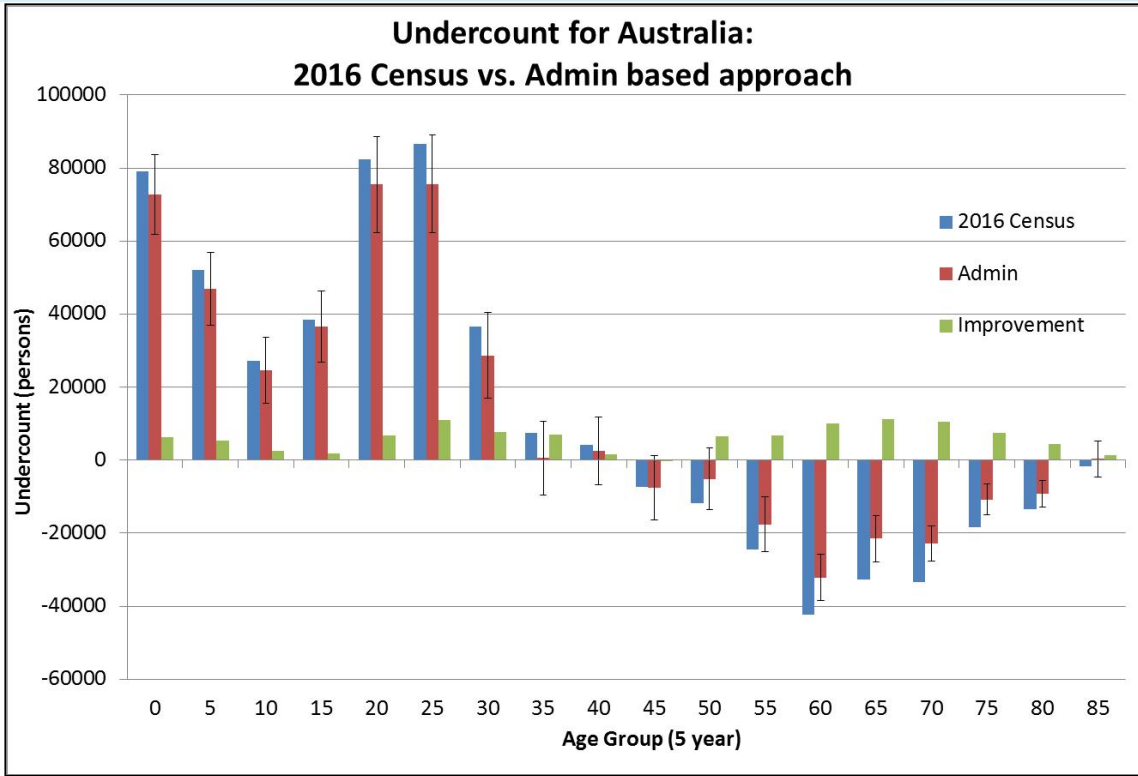
2. Imputation for non-responding dwellings



2. Imputation for non-responding dwellings



2. Imputation for non-responding dwellings



Research information on our website



- ▶ [Administrative data research for the 2021 Census](#)
- ▶ [Can administrative data help to improve the Census count?](#)

Questions?