Estimation in the Australian Census Post-Enumeration Survey

Philip Bell

philip.bell@abs.gov.au

Research Paper: An Estimating Equation Approach to Census Coverage Adjustment, May 2007 (cat no. 1351.0.55.019)
http://www.abs.gov.au/ausstats/abs@.nsf/mf/1351.0.55.019
by Philip A. Bell, Claire F. Clarke and Julian P. Whiting

Information Paper: Measuring Net Undercount in the 2006 Population Census, 2007 (cat no. 2940.0.55.001)
http://www.abs.gov.au/ausstats/abs@.nsf/mf/2940.0.55.001

What I plan to cover

- How a post-enumeration survey (PES) can measure Census undercount
- The DSE for estimating missed dwellings
- The PREG for estimating missed persons
- Some results

The population Census

- a huge operation!
- coverage: persons in Australia on Census night
- fine geographic detail available
- *usual residence* available (& actual location)
- a key input to ERP (estimated resident population)

BUT Census misses some people

- (and counts some more than once!)

SO an estimate of Census undercount is required

A survey to check on the Census

Idea: run a PES to find the missing people

Issues:

- independence of survey from Census
  - selection in the survey may prompt people to fill out their Census form
- surveys also miss some people
  - often the same types of people as the Census does!
- survey won't exactly replicate Census concepts

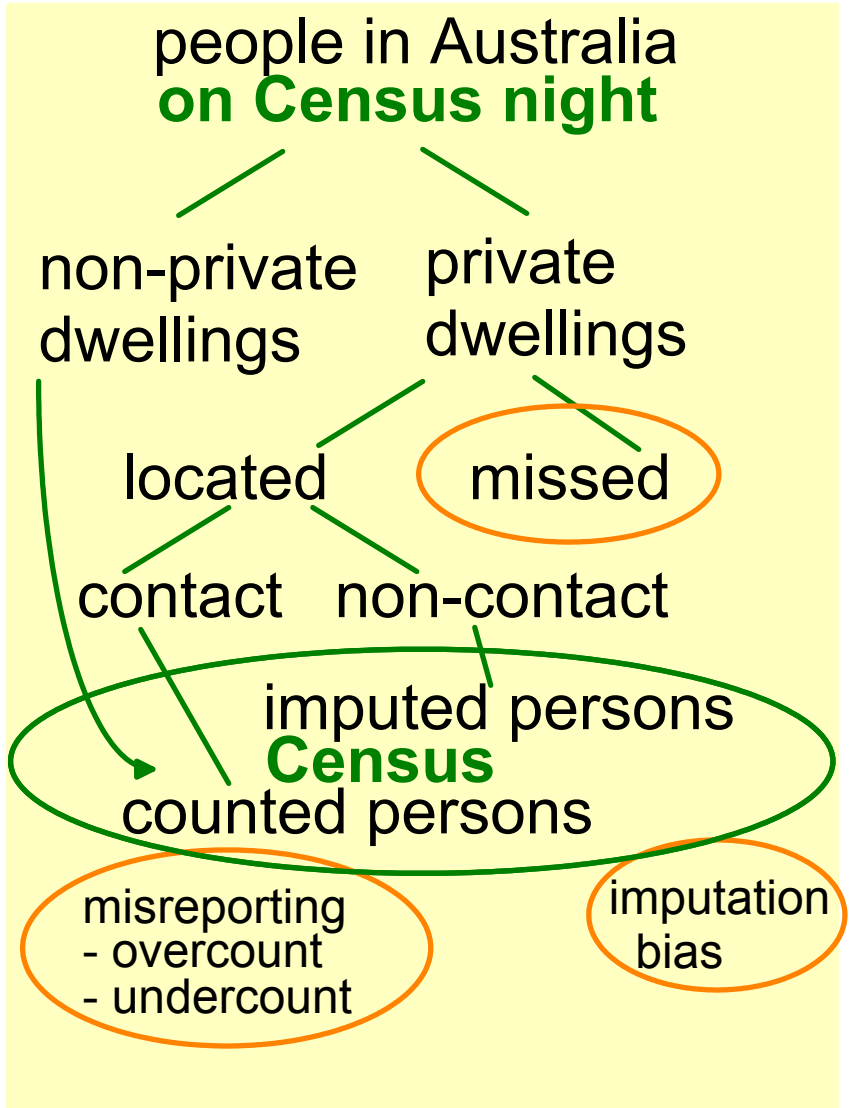
Census Post-enumeration Survey

- uses the Monthly Population Survey (MPS) frame
  - selects "blocks" from collector's districts (CDs)
  - list dwellings and skip through list to select dwellings
  - probability of selection constant within state/territory
  - standard rules ensure a single chance of selection
- around a month after the Census
  - to avoid influencing Census results in the selected CDs
  - for PES, forms arriving later classed with "non-contacts"
- **PES coverage** is persons in Australia at PES time
  - Private Dwellings (PDs) only
  - use this to represent whole Census population

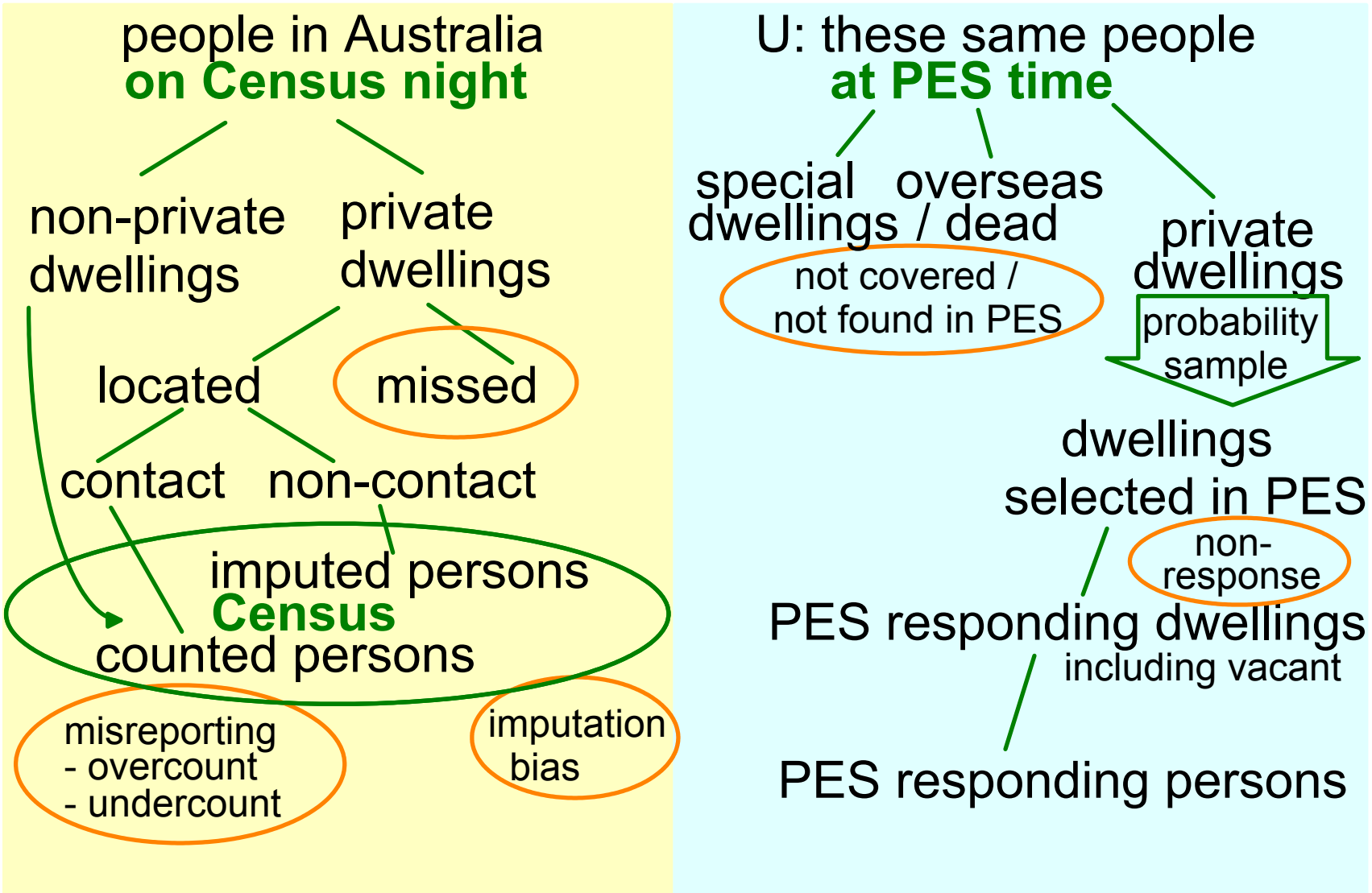
Overview of estimation

- For each person in PES coverage, check
  should the person have been counted in Census (**0 or 1**)
  was the person counted in Census (**how many times**)
  - it is critical that these values are accurate; they are obtained from **matching** the PES units to the Census
- Weight PES sample to represent the population
  estimate = weighted sum of person values
  - adjust PES weights for response and coverage by making estimates of ***was*** equal the ***Census count***
- Estimate the undercount using PES sample
  undercount = PES estimate of ***should*** minus ***was***

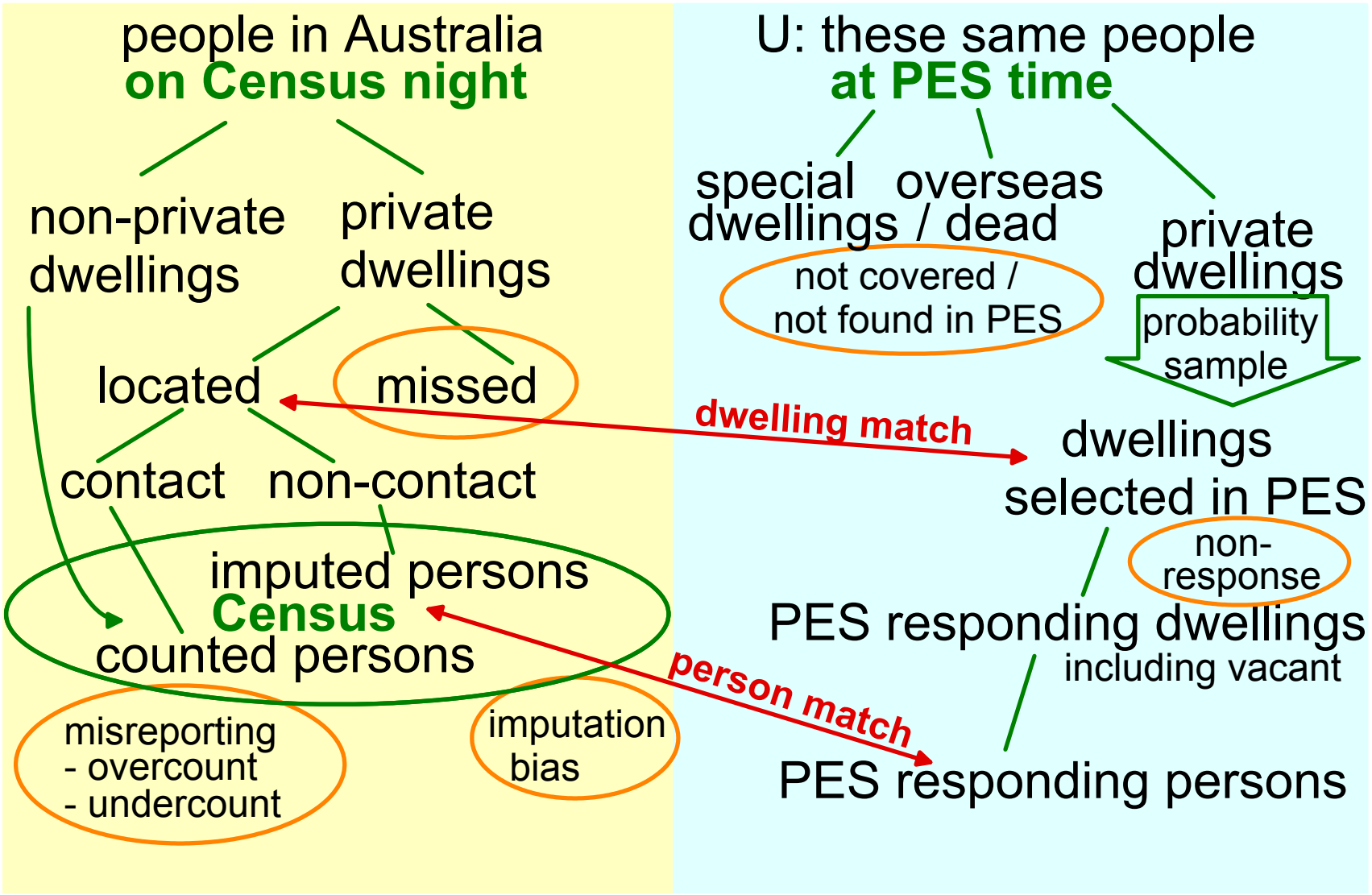
Census overview



PES overview



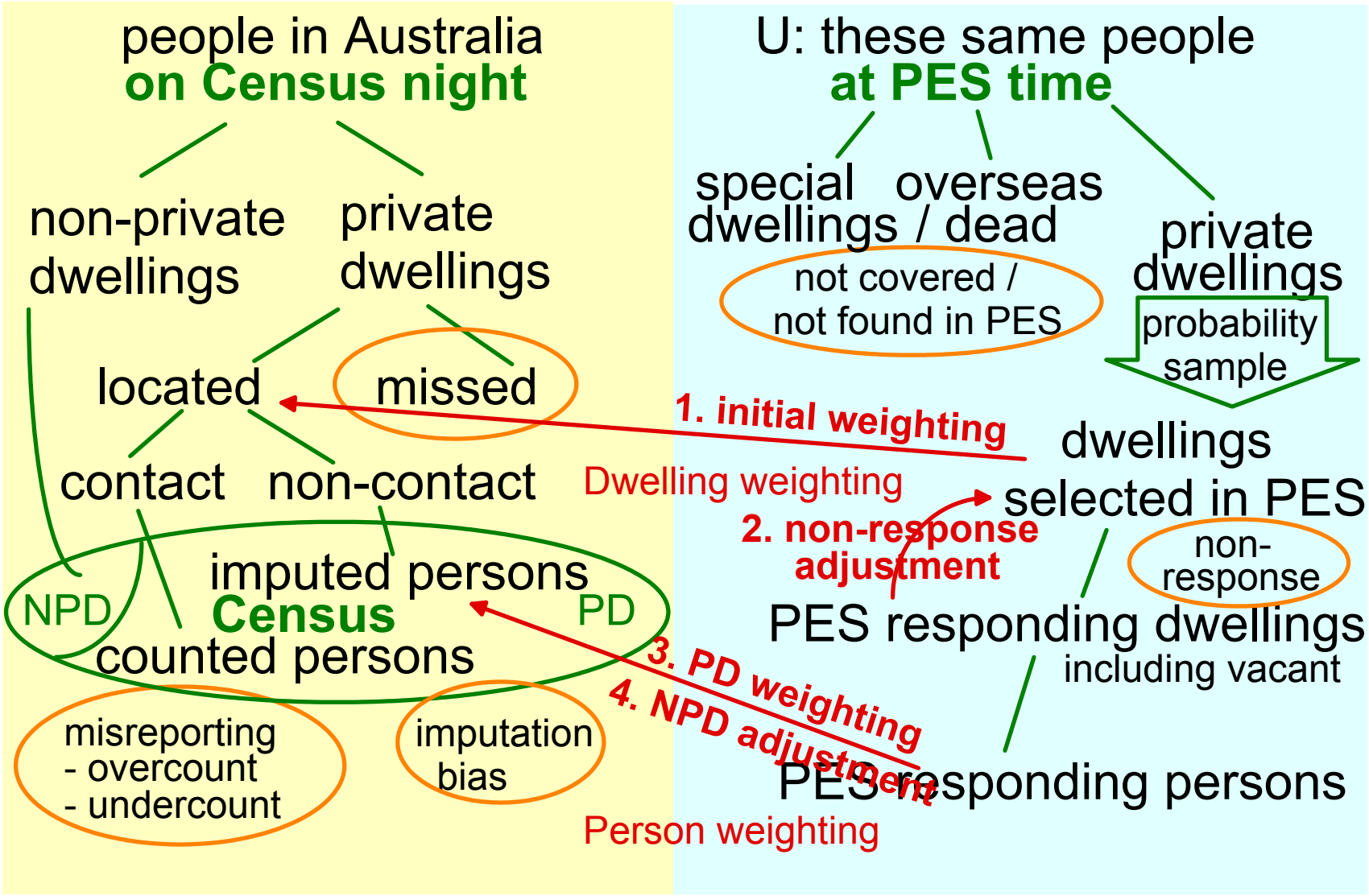
PES match and search



Weights produced in PES

- Dwelling weighting: represents all private dwellings
  - PES dwellings **matching** to Census represent the dwellings located by Census
  - PES dwellings **not matching** to Census represent the dwellings missing in Census
- Person weighting: represents all persons in Australia on Census night (PD and NPD)
  - PES persons who **were counted** in the Census represent the persons counted in the Census
  - PES persons **not counted** in the Census represent persons undercounted in the Census
  - can measure persons in **non-contact** dwellings too

Steps of PES weighting



Principle underlying weighting

- **Identical** weight changes should apply to all units with the same characteristics
  WHETHER OR NOT
  they were counted in the Census.
- So the **weight adjustment**
  for units **counted** in the Census
  is also used for similar units **not counted**

Step 1: Initial dwelling weighting

- Gives a weight to all dwellings selected in PES
  - each gets a weight based on their selection probability:
    321 in NSW (sampled 1 in 321 dwellings in this big state)
    ... 37 in NT (sampled 1 in 37 to improve NT estimates)
  - each was either counted or missed in the Census

1. estimate the counted dwellings
2. adjustment = $\frac{\text{actual Census count of dwellings}}{\text{estimated number of counted dwellings}}$
3. initial weight = adjustment x selection weight

- for 15 regions (state/territory by capital city) (non-ICF)
- and also four states by indigenous communities (ICF)

- This estimator is called the dual system estimator

Dual system estimator (DSE)

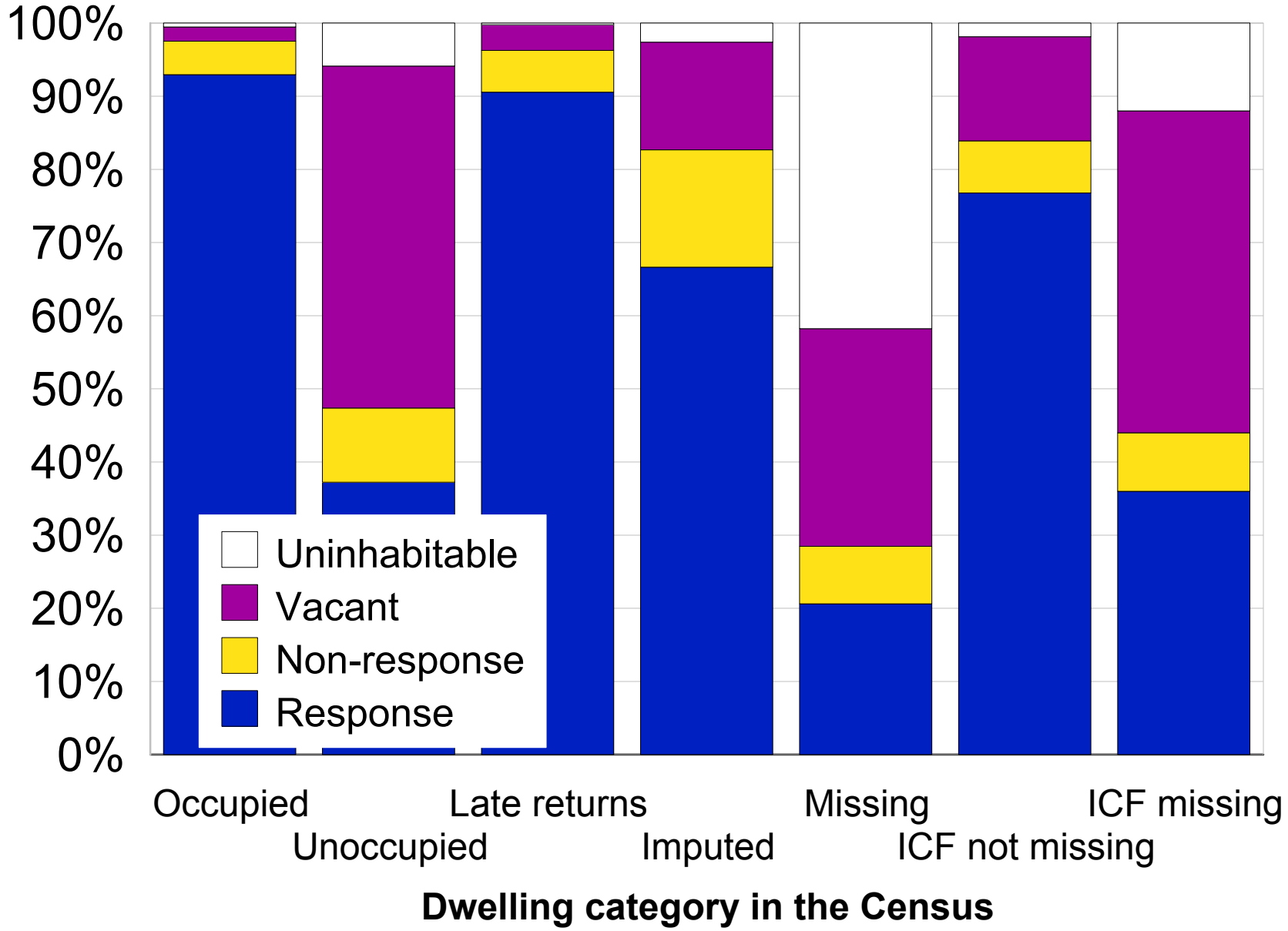
Census outcome	PES outcome		Total
	counted	missed	
counted	PES **estimate** of counted dwellings	**evaluate & apply ratio**	Actual counted dwellings
missed	PES **estimate** of missed dwellings	**adjustment** **apply**	Dwellings missed in the Census
Total	**original** weights only represent **this part**	ratio adjustment accounts for this part	**adjusted** weights represent **both parts**

- "Dual system" refers to having
  - two collections of the same population
  - each with a probability of missing units

Step 2: Non-response adjustment

- Not all selected dwellings respond to PES
- Within adjustment categories, adjust weights of:
  - **Vacant** dwellings at PES time, and
  - **Responding** dwellings
- To represent other dwellings: either
  - **F**ull non-contact (could potentially be vacant)
    - inflate weights of both **Vacant** and **Responding**
  - **Refusals** etc. (won't be vacant)
    - inflate weights of **Responding** dwellings only

PES response categories



Step 3: PD person weighting

- each responding person gets their dwelling weight
- this dwelling-weighted person file **represents**:
  - persons in private dwellings (PDs) **at PES time**
  - who don't get missed when the dwelling responds to PES
- but **fails to represent**:
  - persons in **non-private dwellings** at PES time
  - persons **moving overseas** or dying since PES
  - persons **missed** by PES within the responding dwellings
- The PD person weights aim to make the person file represent all persons in private dwellings at Census time (including non-contact dwellings)

A simple approach: use the DSE

- DSE would apply a constant ratio adjustment within non-overlapping **benchmark categories**
  1. estimate the persons counted in the Census (sum the dwelling weight times the ***was*** value)
  2. adjustment = $\frac{\text{actual Census count of persons}}{\text{estimated number of counted persons}}$
  3. person weight = adjustment x dwelling weight (for **all persons**, not just those counted in the Census)
- For a benchmark category (e.g. NT young males):
  - adjust for observed under-representation of counted
  - assume the same under-representation of uncounted
  - thus NT young males get the same *coverage-response adjustment*, regardless of their Census response

DSE for persons

Northern Territory young males in private dwellings

Census outcome	PES outcome		Total
	counted	missed	
counted	PES **estimate** of counted persons	**evaluate & apply ratio**	Actual counted persons
missed	PES **estimate** of missed persons	**adjustment** **apply**	Persons missed in the Census
Total	**original** weights only represent **this part**	ratio adjustment accounts for this part	**adjusted** weights represent **both parts**

Issues with the DSE

- Census categories may not agree with PES
  - actual Census counts are based on *Census region-age-sex variables*
  - but PES persons not matching to Census only have *PES region-age-sex variables*
  - so Census counts are not exactly for PES post-strata
  - more of a problem if benchmark categories are based on less solid items (e.g. indigenous status)
- biased if benchmark categories not homogeneous
  - if a subset (e.g. indigenous) need a different adjustment
  - sticking to non-overlapping categories restricts the room for extra variables (small sample in very fine categories)

Prediction REGression estimation

- new method used instead of DSE
  - uses *instrumental variables regression*
- the PREG extends the DSE to:
  - non-overlapping categories
  - that can be reported differently in PES and Census
- assumes the **response-coverage adjustment** depends only on a person's PES responses
- evaluates this adjustment based on Census counts
  - makes **PES estimates** of **was** for each **Census category** exactly reproduce the **actual Census counts**

Step 4: NPD-adjusted person weight

- The PD weight should represent persons in private dwellings at time of the Census
- The NPD adjustment applies PREG to adjust to PD + NPD counts for the contact sector
  - so persons in NPDs are represented by persons in PDs with the same region, age and sex
- Persons missed at non-contact dwellings don't get adjusted
  - since all non-contacts were private dwellings
  - want unbiased measure of "non-contacts" to show the effectiveness of imputation

Key improvements in 2006 PES

- Covers remote areas and indigenous communities
- Improves adjustment for non-responding dwellings
- Measures non-contact sector
  - ie late returns plus imputed returns
- Measures differences in reporting (Census vs PES)
- Benchmarks to many more Census counts
  - region, sex, age, indigenous, indigenous community, marital status, country of birth
  - lower bias, but higher SEs!
    encompassing more aspects of error

Components of undercount

PES category	Undercount in contact sector: Persons not counted	Net difference between PES and Census category	Persons with Census category not-stated	Undercount in non-contact sector (after imputation)	Net undercount
Australia	799,967	0	0	-250,481	549,486
State	...		**too much imputed!**		
Sex					
male	455,725	8,929	0	-135,130	329,524
female	344,242	-8,929	0	-115,351	219,962
Age group	...	**Indigeneous status was not imputed!**			
Indigenous				↓	↓
yes	54,056	-30,797	7,989	27,930	59,178
no	745,911	30,797	367,343	479,551	1,623,602
Marital status	...				
Birthplace	...				

Thank you