Can’t live with ‘em, can’t live without ‘em

Assumptions – the necessary evil

International Total Survey Error Workshop

Bergamo, Italy

June 2019

Abby Morgan – abby.morgan@stats.govt.nz,

Rebecca Green

Patrick Graham

Overview

- Intro to the problem
- Basic theory
- Our application
- Challenging the assumptions
- Conclusions

Intro to the problem

- Coverage estimation has traditionally focused on under-coverage
- Increased admin data = increased risk of over-coverage
- Ideal solution:
  - Mechanism for measuring both under and over-coverage without additional data collection

Intro to the problem

- Dual Systems Estimation

		List 2		
		1	0	
List 1	1	n_{11}	n_{10}	N_{1+}
	0	n_{01}	$n_{00}?$	
		N_{+1}		

- Closed target population
- No erroneous inclusions
- Homogeneity of capture
- Causal independence
- Perfect linking

Intro to the problem

Stats NZ
Tatauranga Aotearoa

- Dual Systems Estimation

		List 2		
		1	0	
List 1	1	n_{11}	n_{10}	N_{1+}
	0	n_{01}	$n_{00}?$	
		N_{+1}		

- Closed target population
- No erroneous inclusions
- Homogeneity of capture
- Causal independence
- Perfect linking

Theory

We start with a cross-tabulation of the Target and the List

		List		
		1	0	
Target	1	n_{11}	n_{10}	N_T
	0	n_{01}	0	
		N_L		

If we can estimate the under- and over-coverage of the list, then we can estimate the total target population as:

$$\widehat{N}_T = N_L - \hat{n}_{01} + \hat{n}_{10}$$

Theory

We represent the three cell probabilities:

		List	
		1	0
Target	1	ϕ_{11}	ϕ_{10}
	0	ϕ_{01}	0

We can use the cell probabilities to work out under and over coverage probabilities

$$\Pr(not\ on\ list|\ in\ Target) = \phi^{under} = \frac{\phi_{10}}{\phi_{11} + \phi_{10}}$$

$$\Pr(not\ in\ Target|\ on\ List) = \phi^{over} = \frac{\phi_{01}}{\phi_{11} + \phi_{01}}$$

$$\widehat{N}_T = N_L \left(\frac{1 - \phi^{over}}{1 - \phi^{under}} \right)$$

Our application: dwelling coverage

- Consider the Census list of dwellings as the list
- Conduct an independent enumeration for small areas
  - Establish it as the ‘source of truth’ = target
    - Assume target is achieved

		List (Census)	
		1	0
Target (Ind. Enum)	1	ϕ_{11}	ϕ_{10}
	0	ϕ_{01}	0

Challenging assumptions $\lambda \equiv 1$

		List (Census)	
		1	0
Sample (Ind. Enum)	1	$\lambda\phi_{11}$	$\lambda\phi_{10}$
	0	$(1-\lambda)\phi_{11} + \phi_{01}$	$(1-\lambda)\phi_{10}$

$$\Pr(not\ on\ list \mid in\ Target) = \phi^{under} = \frac{\lambda\phi_{10}}{\lambda\phi_{11} + \lambda\phi_{10}} = \frac{\phi_{10}}{\phi_{11} + \phi_{10}}$$

$$\Pr(not\ in\ Target \mid on\ List) = \hat{\phi}^{over} = \frac{(1-\lambda)\phi_{11} + \phi_{01}}{\phi_{11} + \phi_{01}} \neq \frac{\phi_{01}}{\phi_{11} + \phi_{01}}$$

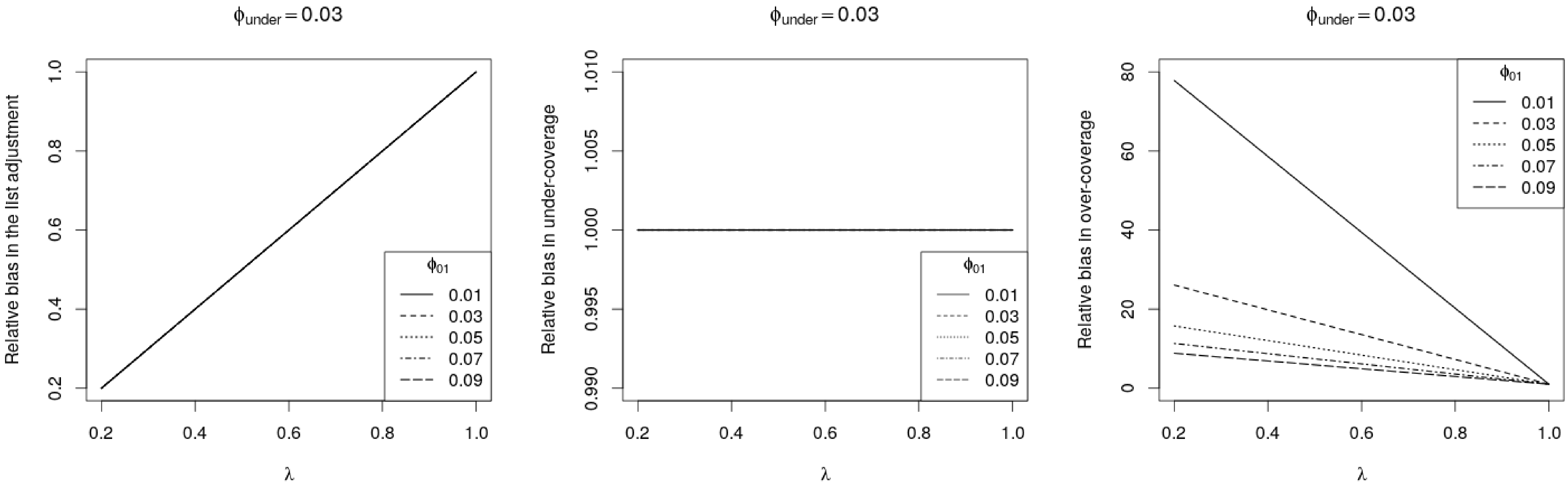
Relative Bias calculations

$$Relative\ bias\ in\ the\ list\ adjustment = \left(\frac{1-\hat{\phi}^{over}}{1-\hat{\phi}^{under}}\right) \div \left(\frac{1-\phi^{over}}{1-\phi^{under}}\right)$$

$$Relative\ bias\ in\ the\ undercoverage\ estimate = \frac{\hat{\phi}^{under}}{\phi^{under}}$$

$$Relative\ bias\ in\ the\ overcoverage\ estimate = \frac{\hat{\phi}^{over}}{\phi^{over}}$$

Challenging assumptions $\lambda \equiv 1$



Challenging assumptions $\lambda_1 = 1, \lambda_2 \equiv 1$

- Separate λ into λ_1 and λ_2
- λ_1 as sample inclusion given found in Census
- λ_2 as sample inclusion given not found in Census
- Establish λ_1 =1
  - Review all records in 01 cell
  - Resolve to truth and restore to 11 cell when in target pop

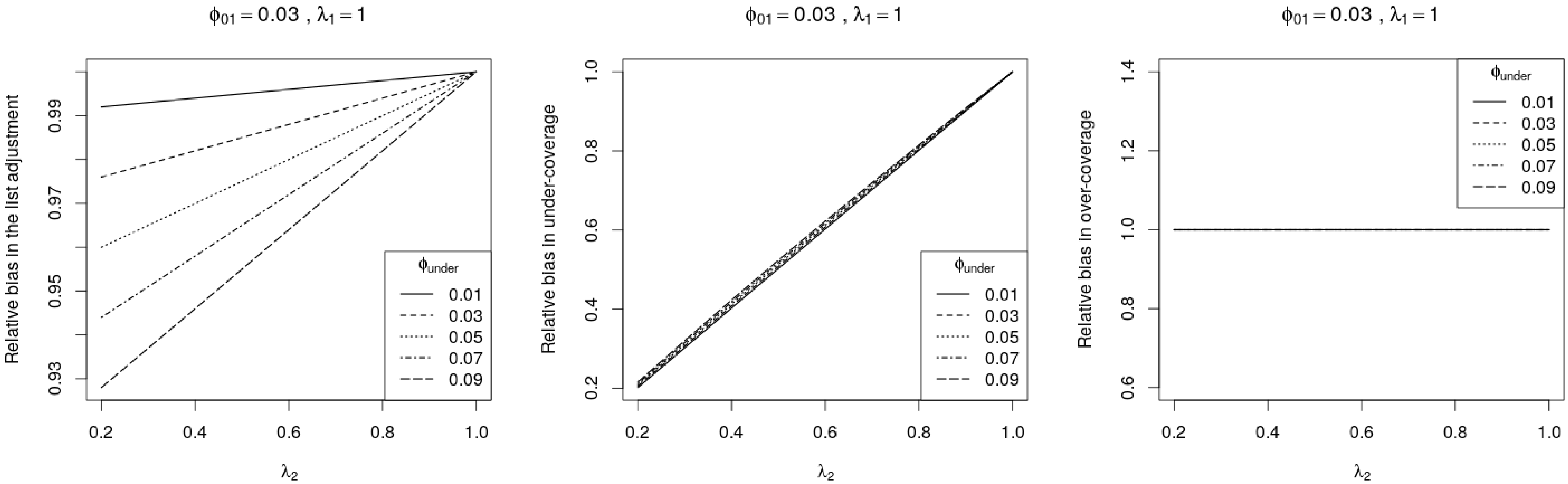
Challenging assumptions $\lambda_1 = 1, \lambda_2 \equiv 1$

		List (Census)	
		1	0
Sample (Ind. Enum)	1	$\lambda_1\phi_{11}$	$\lambda_2\phi_{10}$
	0	$(1-\lambda_1)\phi_{11} + \phi_{01}$	$(1-\lambda_2)\phi_{10}$

$$\Pr(not\ in\ list|in\ target) = \hat{\phi}^{under} = \frac{\lambda_2\phi_{10}}{\lambda_1\phi_{11} + \lambda_2\phi_{10}} \neq \frac{\phi_{10}}{\phi_{11} + \phi_{10}}$$

$$\Pr(not\ in\ target|in\ list) = \phi^{over} = \frac{(1-\lambda_1)\phi_{11} + \phi_{01}}{\phi_{11} + \phi_{01}} = \frac{\phi_{01}}{\phi_{11} + \phi_{01}}$$

Challenging assumptions $\lambda_1 = \lambda_2$



Conclusions

- Incorrect assumptions can have material impacts
- Checking and quantifying assumptions provides the opportunity to adjust for them
- Be proactive not reactive

Discussion questions

- Should we be checking our assumptions if we have no way of mitigating violations?
- Should we be incorporating uncertainty about our assumptions in our inferences?
  - If so, how?