Government Statistical Service
Gwasanaeth Ystadegol y Llywodraeth

Measuring Uncertainty in ONS population estimates: Tools, techniques and outcomes

Demographic Methods Centre, ONS

Introduction

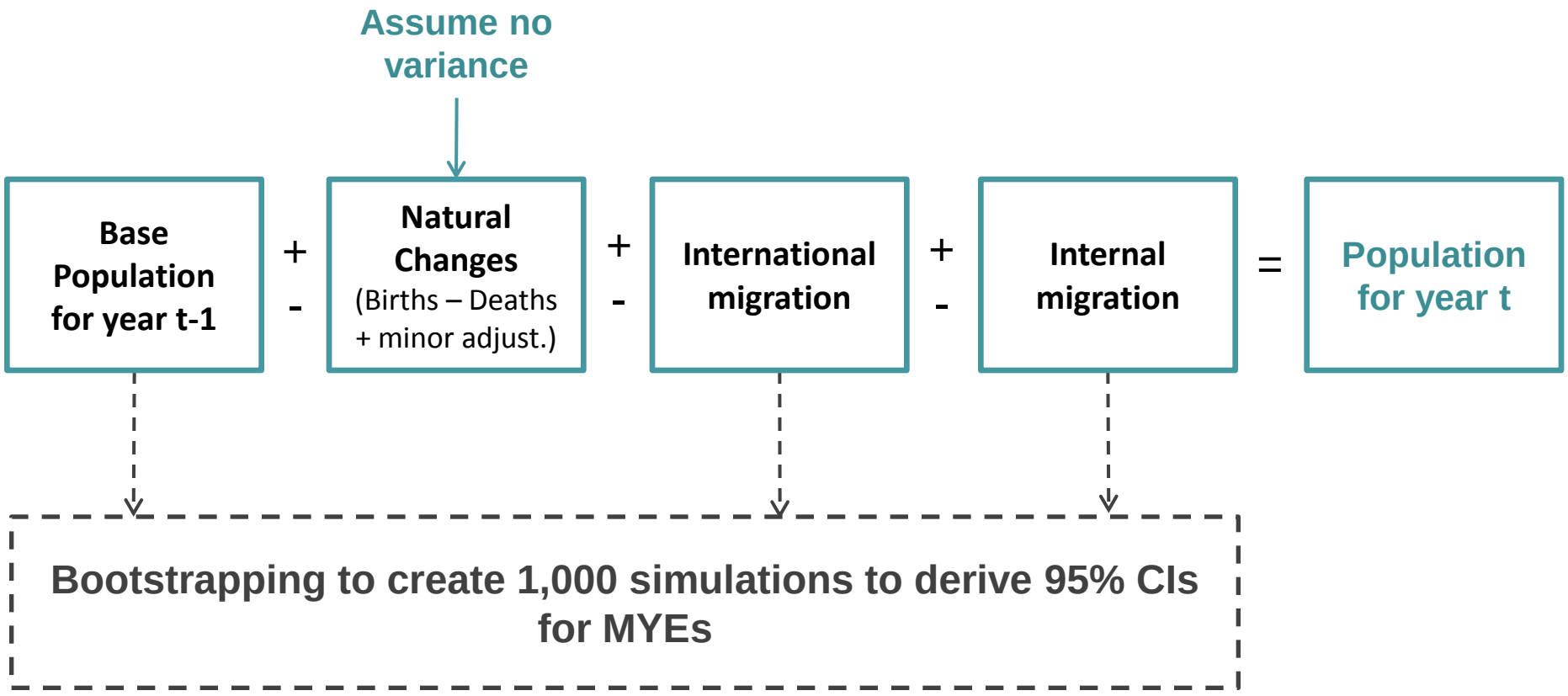
- Mid-year estimates- why measure uncertainty?
- MYE Uncertainty- bootstrapping methods
- Results
- SPDs- what are they?
- Current research

MYEs: why measure uncertainty?

- Mid-year estimates (MYEs) are annual estimates of the UK population who are usually resident on 30 June.
- 65,648,000 in 2016
- Age, sex, local authority, components of change
- Primary and secondary uses
- Resource allocation for local government, health,
- Policy areas: education planning and provision, ageing and pension planning, housing demand and planning &&&
- Code of Practice for Official Statistics: levels of quality are measured and reported, including main sources of bias and other errors.
- MYE: geographic and demographic patterns in the errors

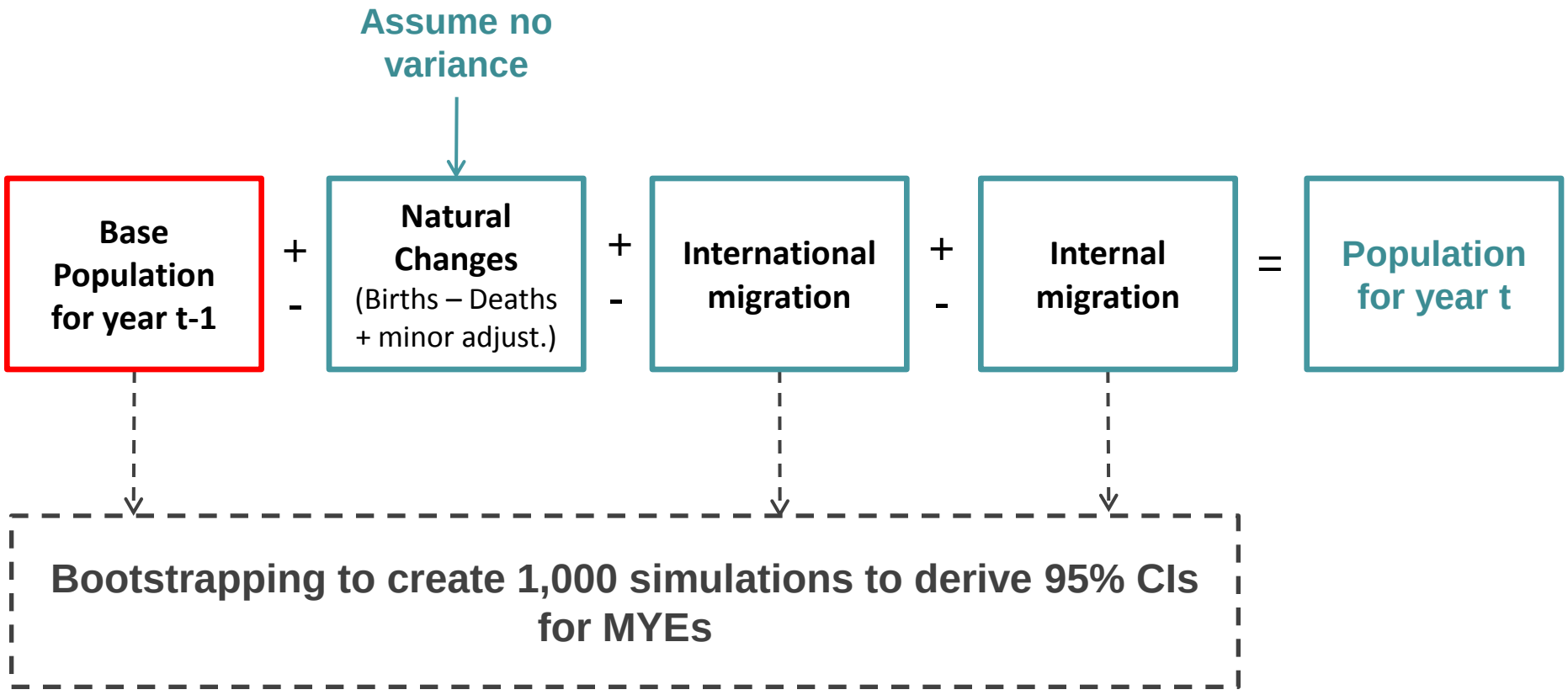
Cohort component method for MYEs

Uncertainty Estimates =



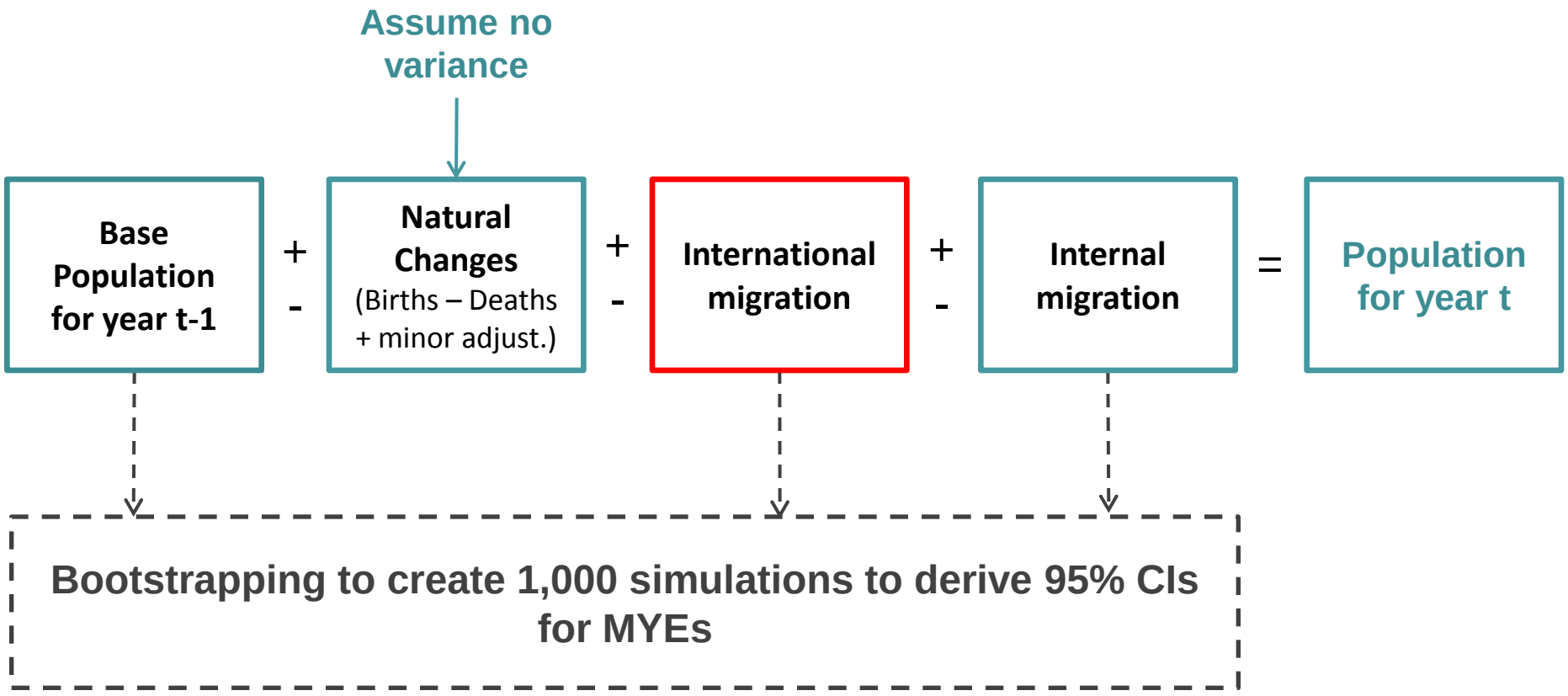
Cohort component method for MYEs

Uncertainty Estimates =



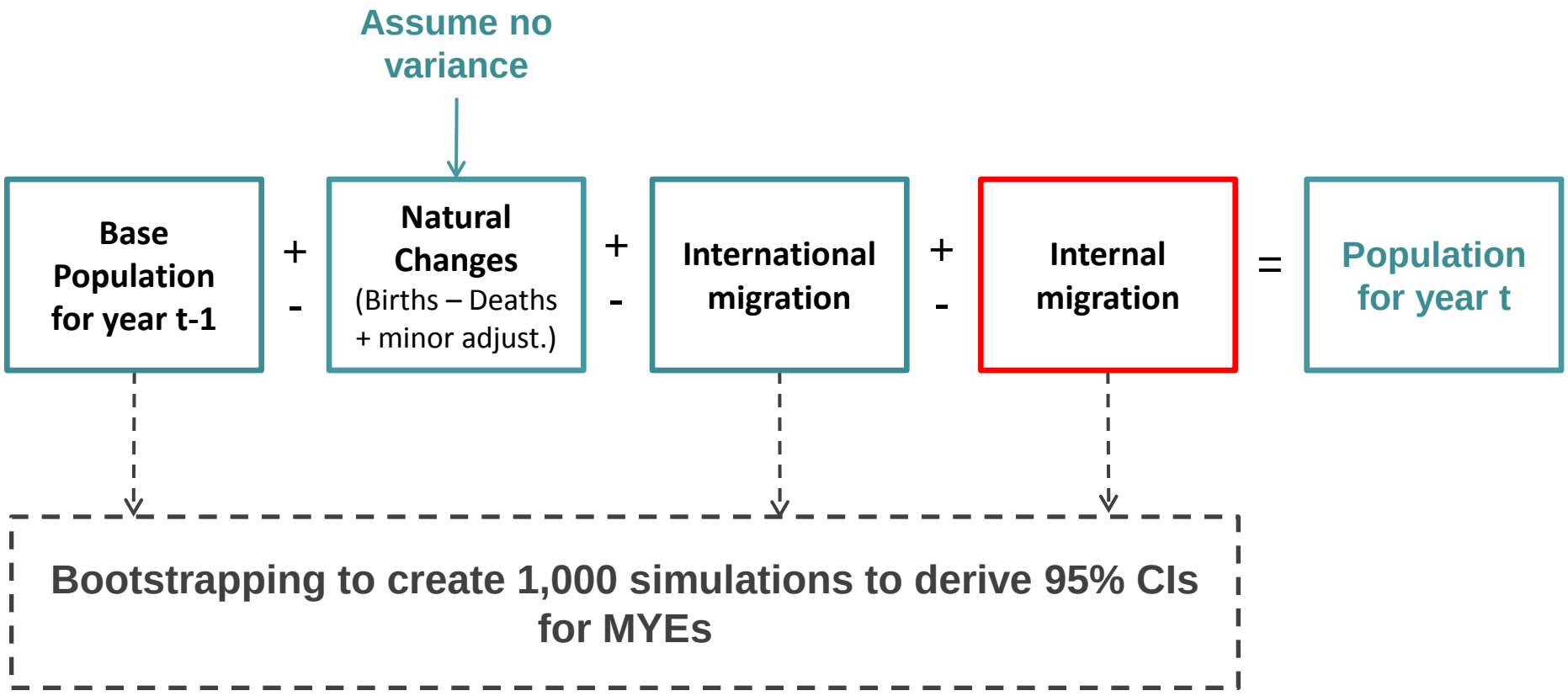
Cohort component method for MYEs

Uncertainty Estimates =



Cohort component method for MYEs

Uncertainty Estimates =



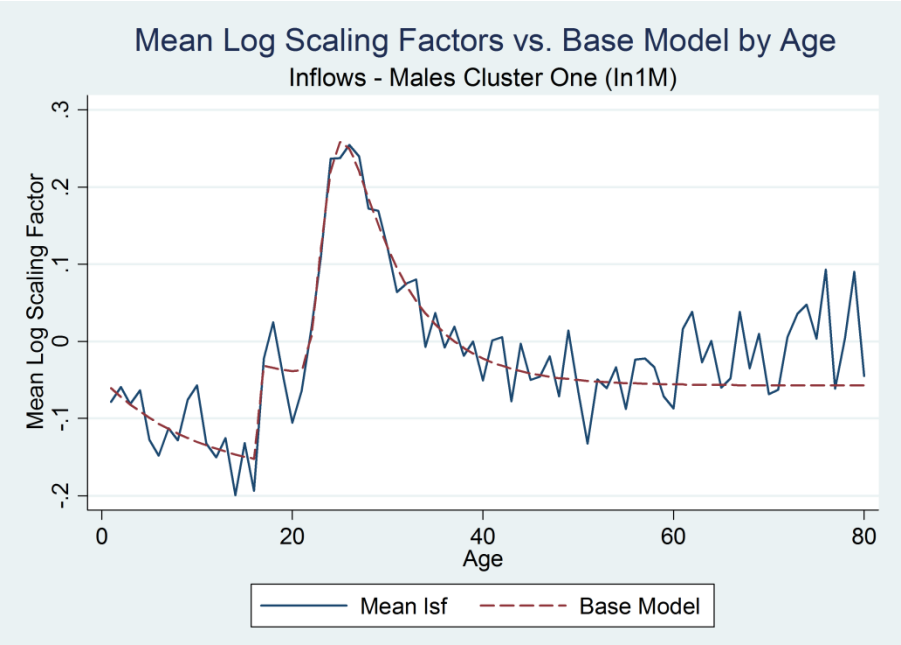
Bootstrapping methods used

Component	Re-sampling method	Type of bootstrap
Census base	Uses published variances Assumes errors are normally distributed.	Parametric
Correction of census base to mid-year	Assumes the same CV as for the 2012 mid-year estimate Assumes errors are normally distributed.	Parametric
International in-migration	IPS: re-sampling with replacement, 1,000 new samples re-run through IPS imputation.	Non-Parametric
	Admin-based allocation to LAs: benchmark approach, comparing each admin source to 2011 Census to derive variances. Assumes errors are log-normally distributed.	Parametric using a benchmark
International out-migration	IPS: re-sampling with replacement, 1,000 new samples re-run through Poisson regression.	Non-parametric
Internal migration	Benchmark approach. Re-samples residuals from the non-linear (Rogers-Castro) regression model of LSFs (Census/ PR moves). Sampled residuals added to predicted values from the model. Model updated with contemporaneous covariates for the inter-censal period	Re-sampling residuals using a benchmark

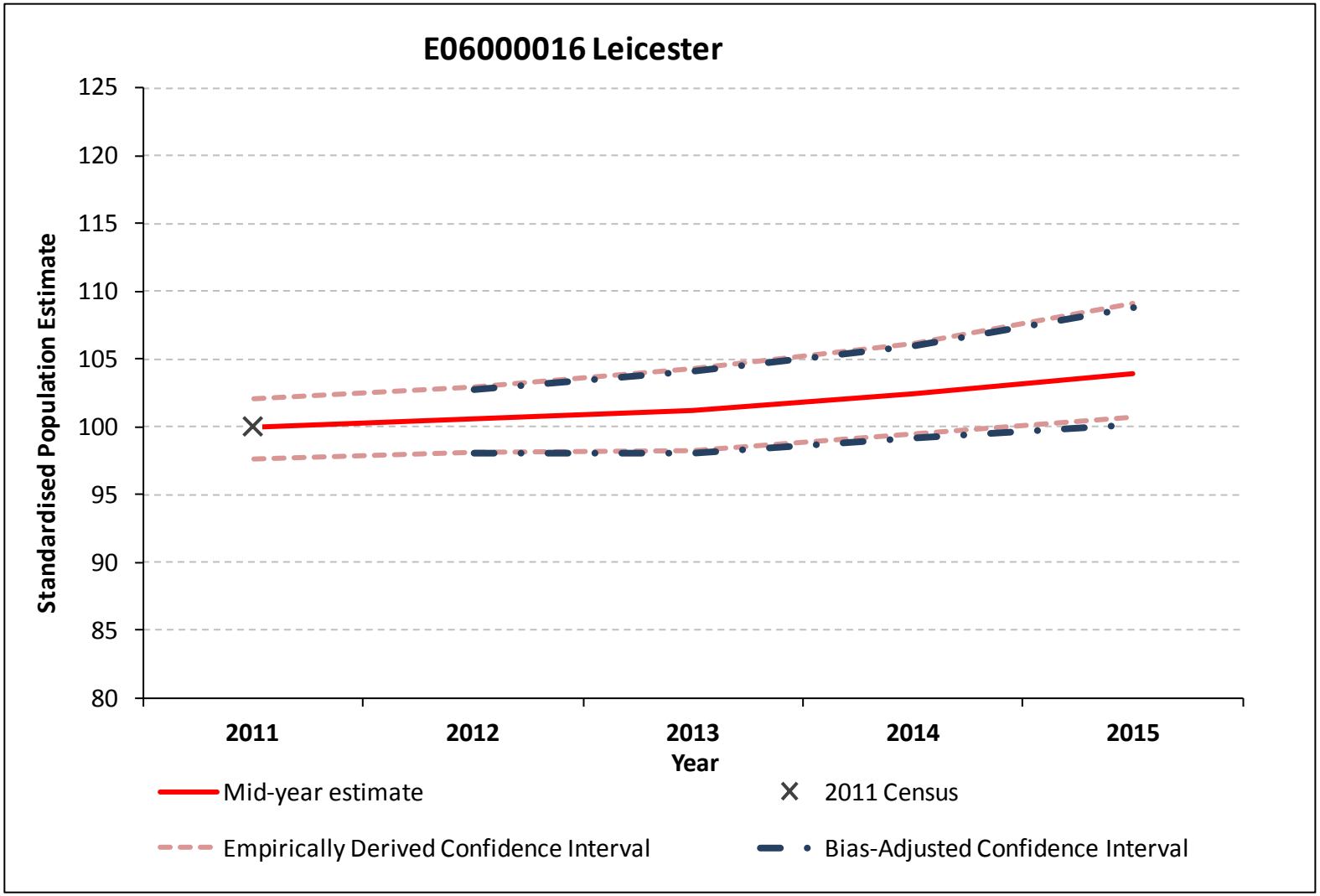
Rogers Castro Model

7 parameter Rogers-Castro Migration Curve

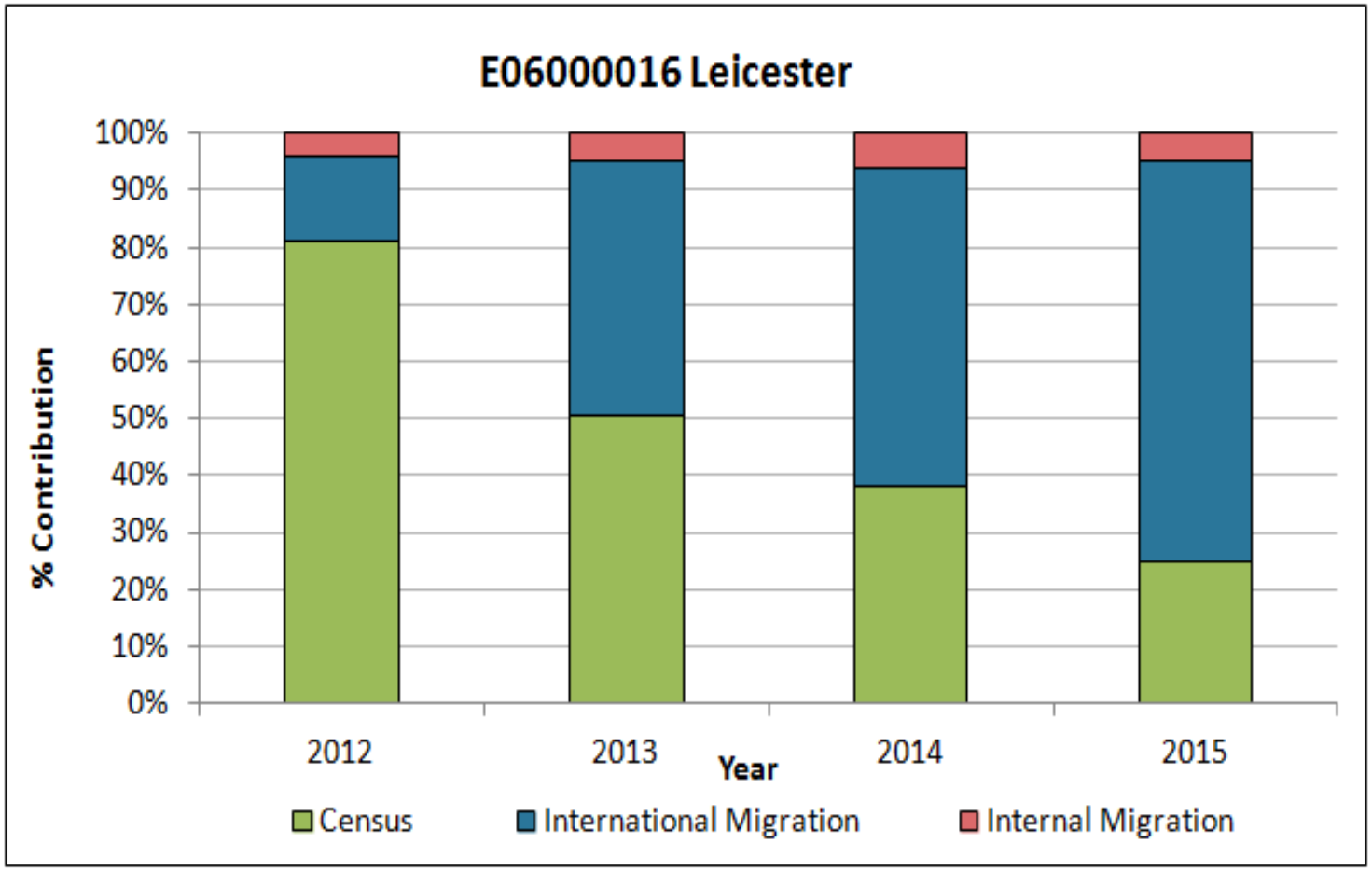
$$\underbrace{a_6}_{\text{constant}} - \underbrace{a_7 \sum_{j=0}^{17} I_j(x)}_{\substack{\text{step change} \\ \text{lowers curve for } <18}} + \overbrace{\underbrace{a_0}_{\substack{\text{height of} \\ \text{childhood} \\ \text{curve}}} \exp[-\underbrace{a_1}_{\substack{\text{rate of} \\ \text{descent}}} x]}^{\text{childhood}} + \overbrace{\underbrace{a_2}_{\substack{\text{height of} \\ \text{curve}}} \exp\left\{-\underbrace{a_3}_{\substack{\text{rate of} \\ \text{descent}}} (x - \overbrace{a_4}^{\substack{\text{position of} \\ \text{curve}}}) - \exp[-\underbrace{a_5}_{\substack{\text{rate of} \\ \text{ascent}}} (x - a_4)]\right\}}^{\text{labour force}}$$



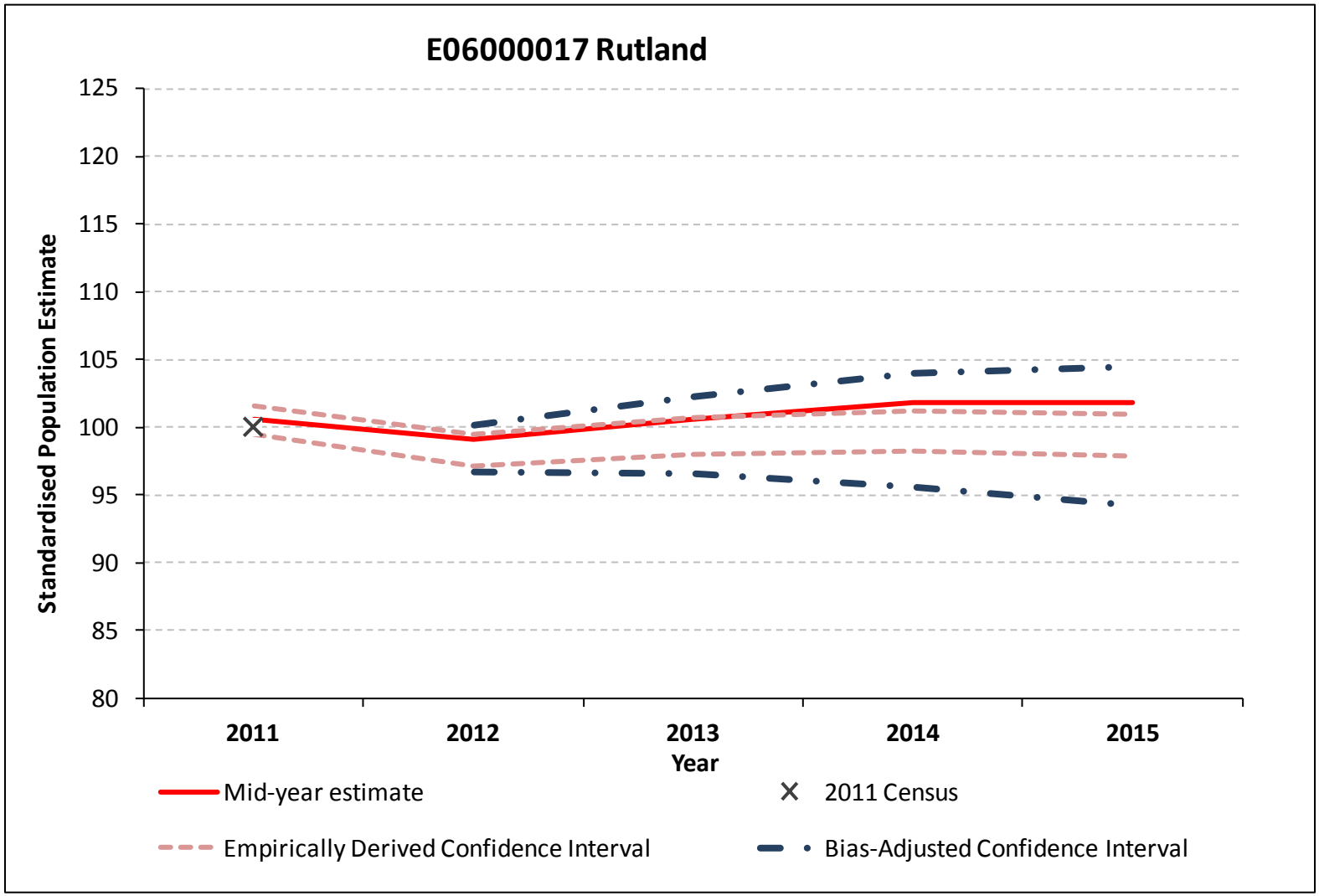
2012-15 CIs for Leicester



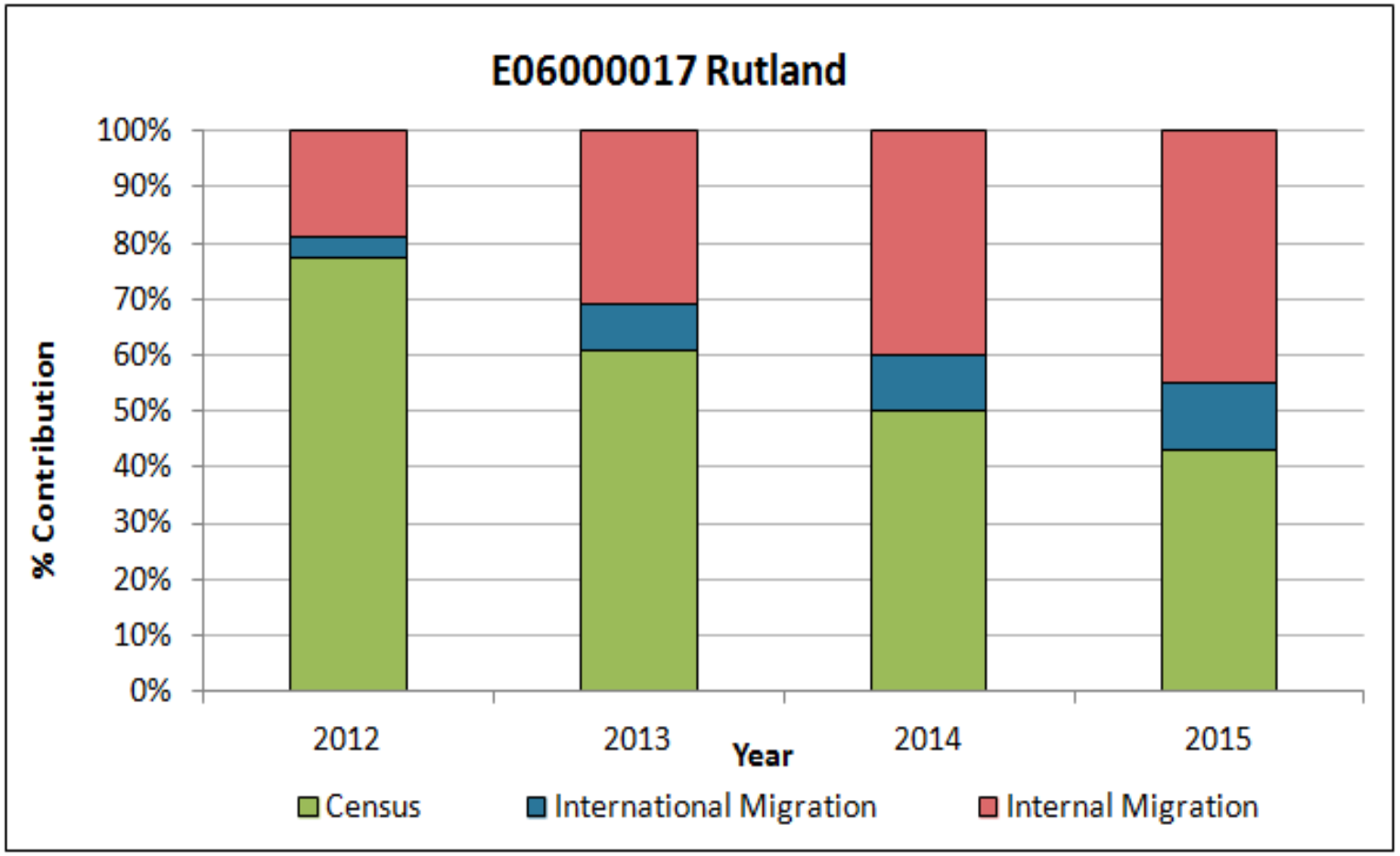
Components driving uncertainty in Leicester



2012-15 CIs for Rutland



Components driving uncertainty in Rutland



What is an admin data census?

Aiming to replicate as many census outputs as possible using admin data (and surveys) by 2021 to compare with 2021 Census

- **Recommendation in 2023**

Three key types of Census outputs:

- Size of population
- Number and structure of households
- Characteristics of housing and the population

Lot of potential with admin data alone but it will not provide the complete solution.

Need access to range of admin data and combine with surveys.

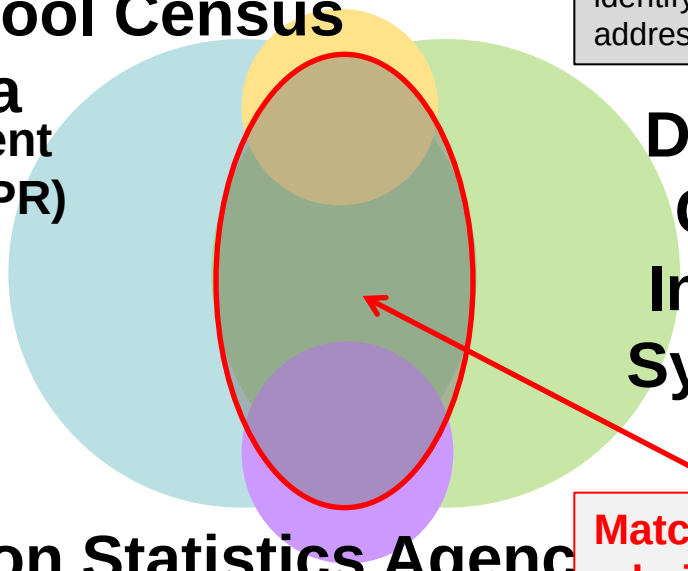
Likely to need two new surveys:

- Annual 1% coverage survey to help measure size of population and households
- Annual characteristics survey – size tbc

Design of the SPD to produce the population estimates

Statistical Population Dataset – SPD

The SPD estimates have been produced by matching individual records across the administrative data sources. To protect privacy of individuals the process involves replacing identifying fields (names, dates of birth and addresses) by one or more artificial identifiers.



Higher Education Statistics Agenc
Data (HESA) - students

Matched records from the various administrative data sources are included in the Statistical Population Dataset (SPD)

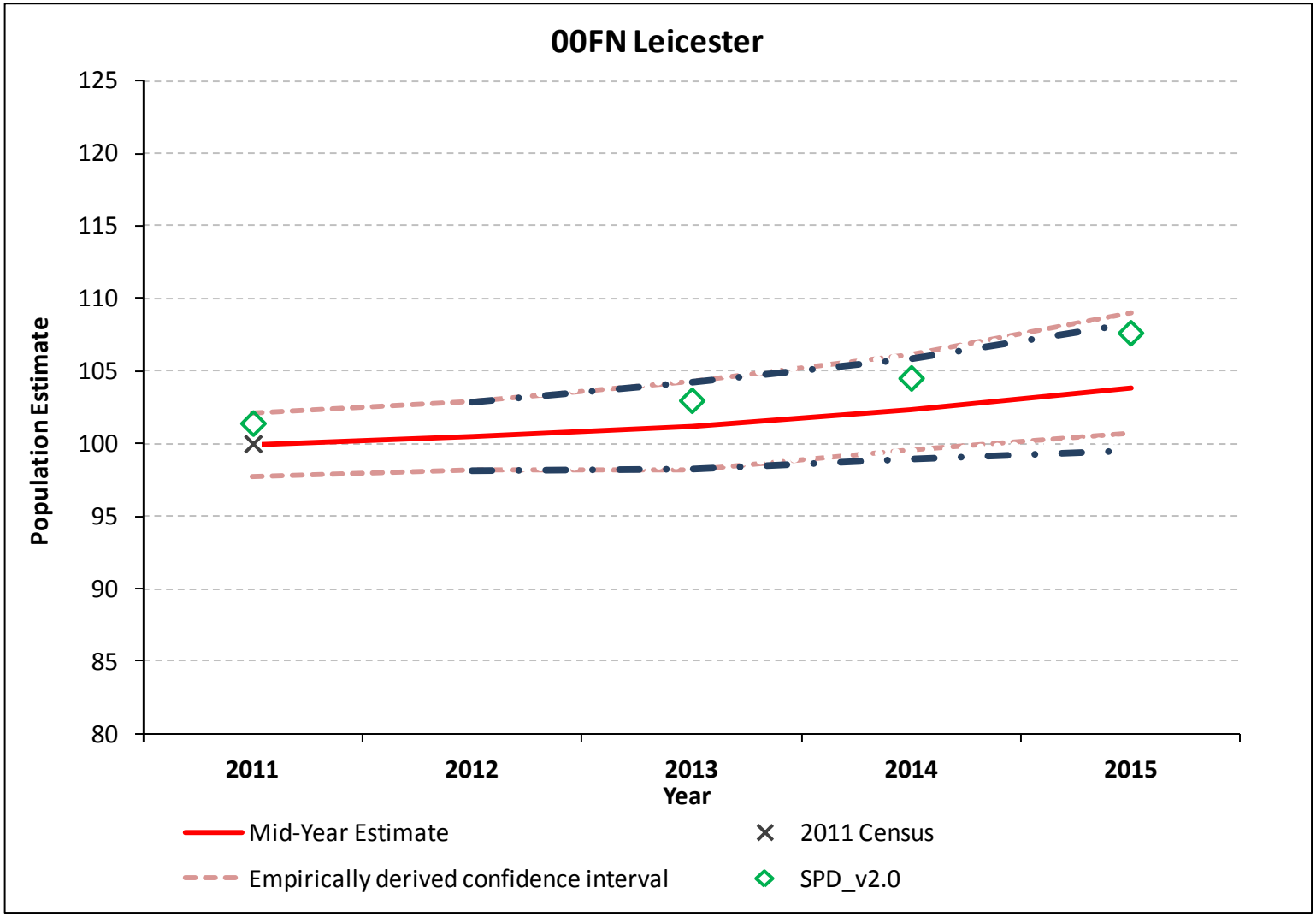
Aggregate totals of Armed Forces personnel are added in to the estimates

=

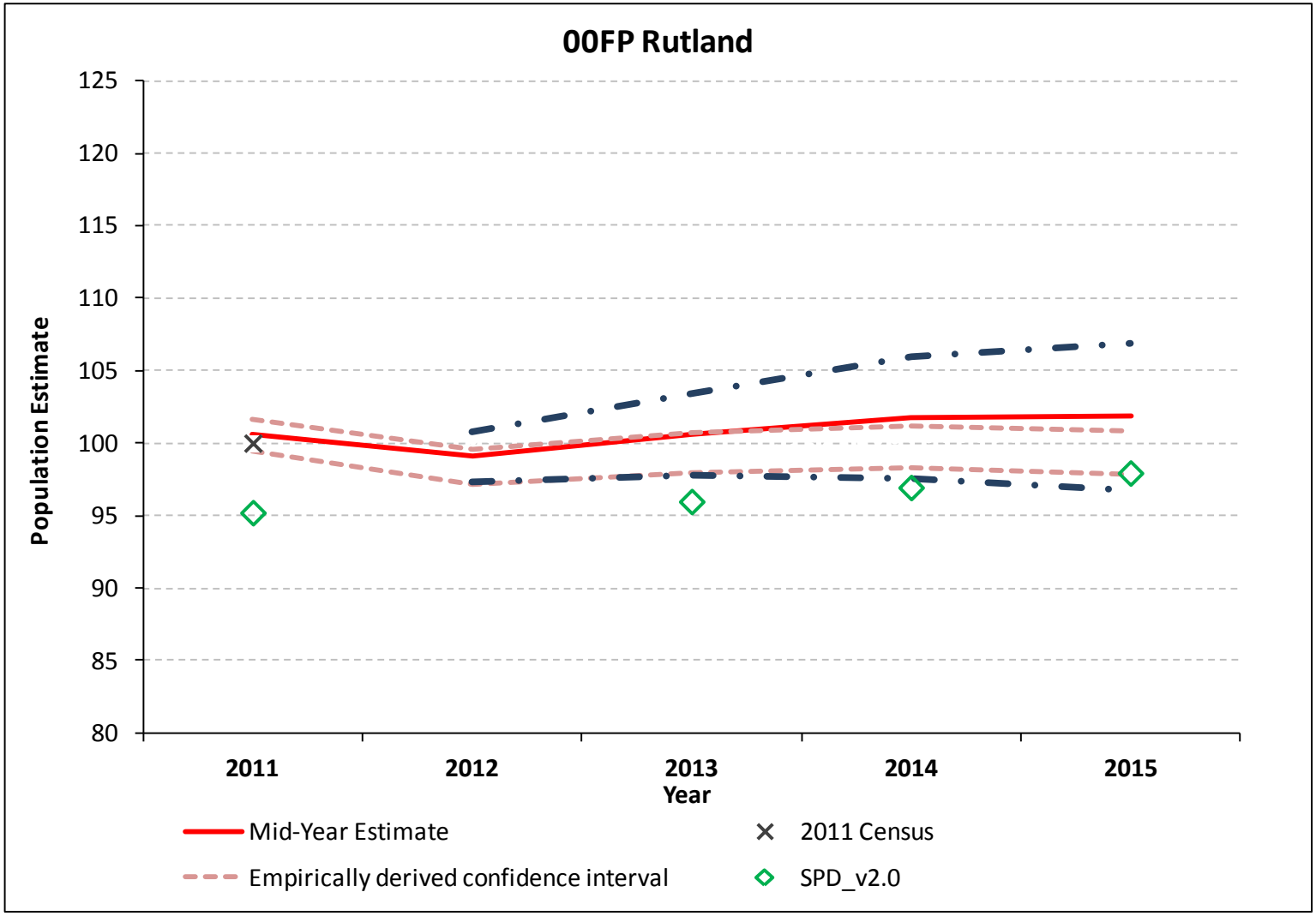
SPD population estimates

A future plan is to use a coverage survey to adjust for biases on the SPD

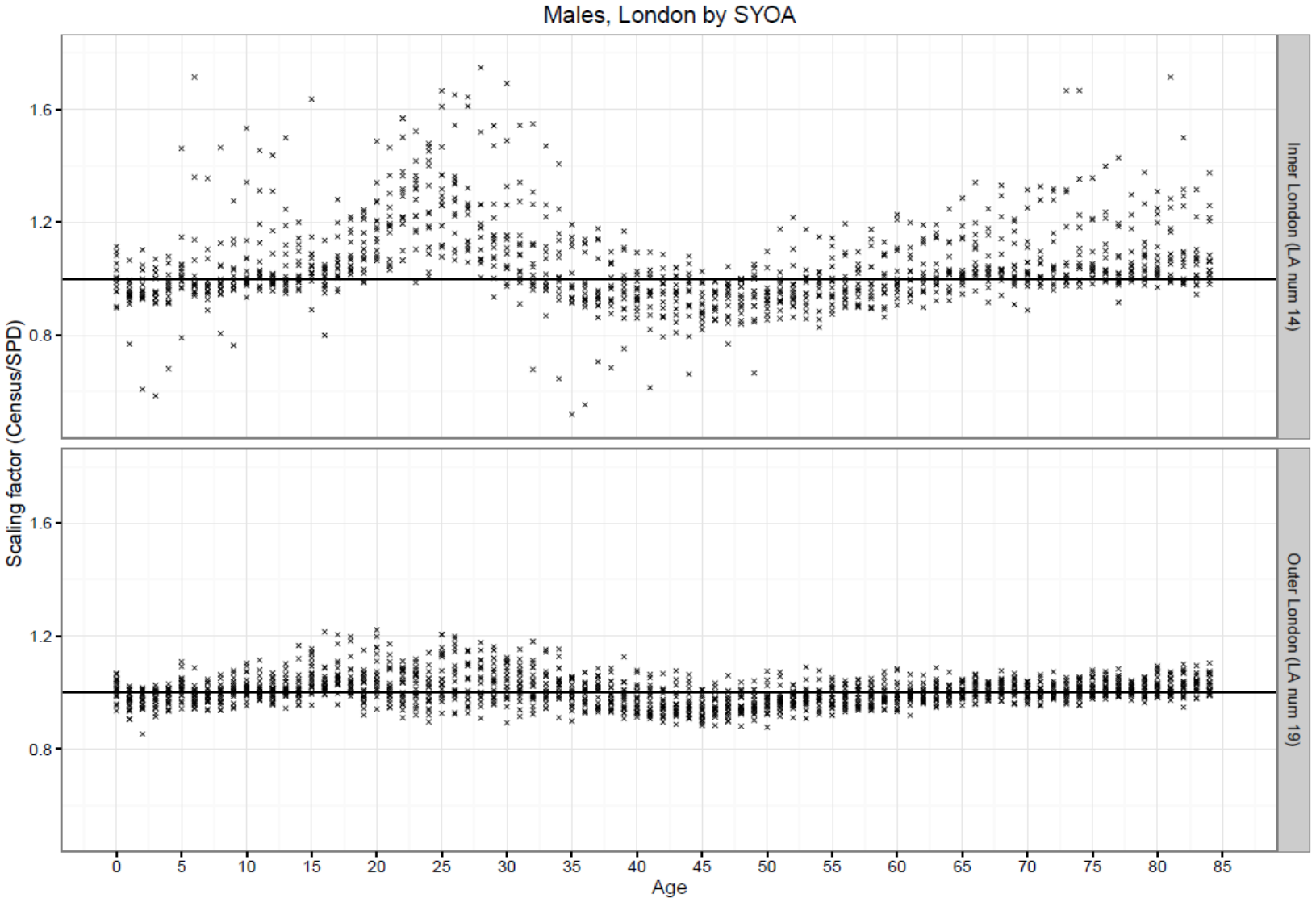
Evaluating SPD2



Evaluating SPD2(2)

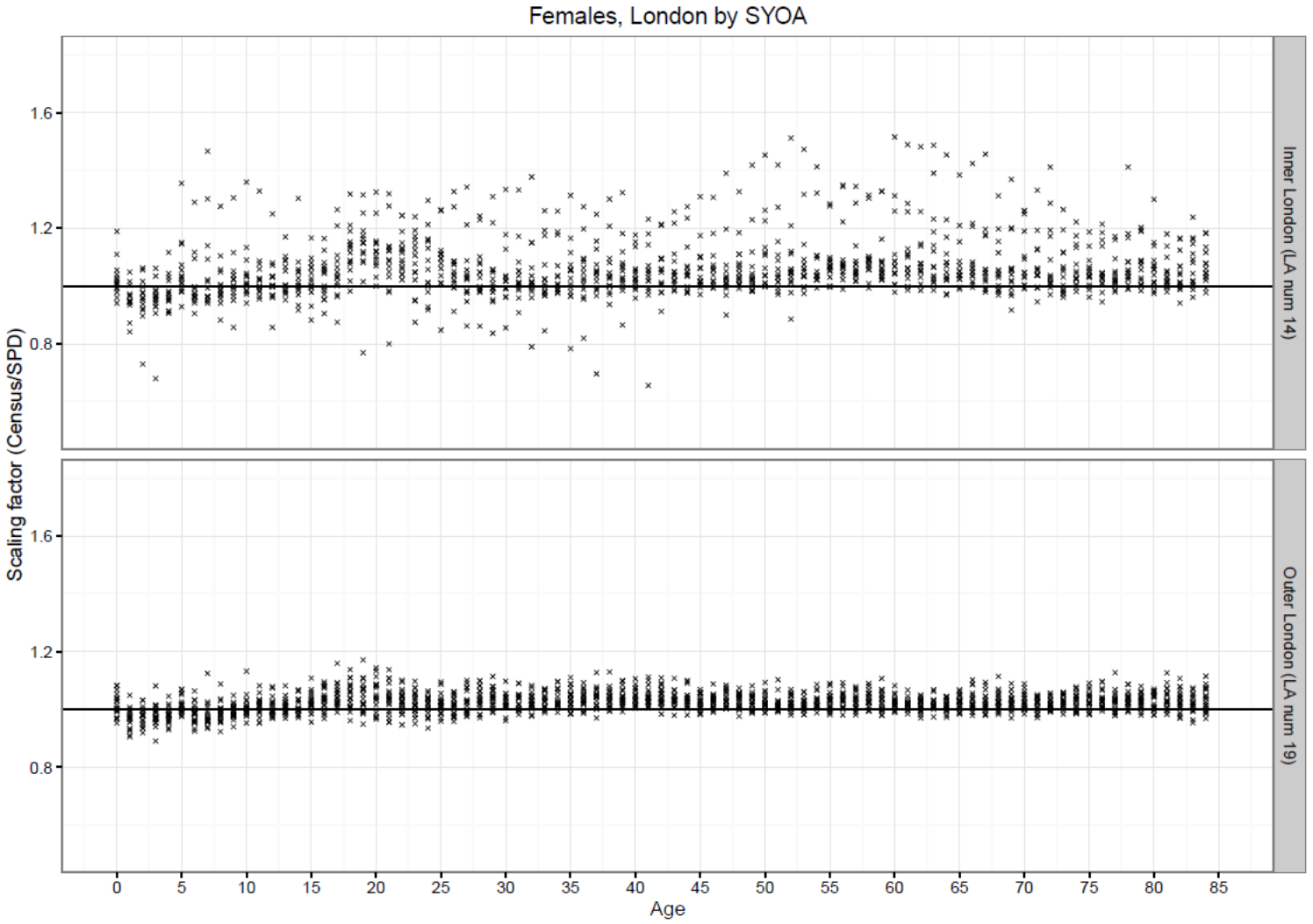


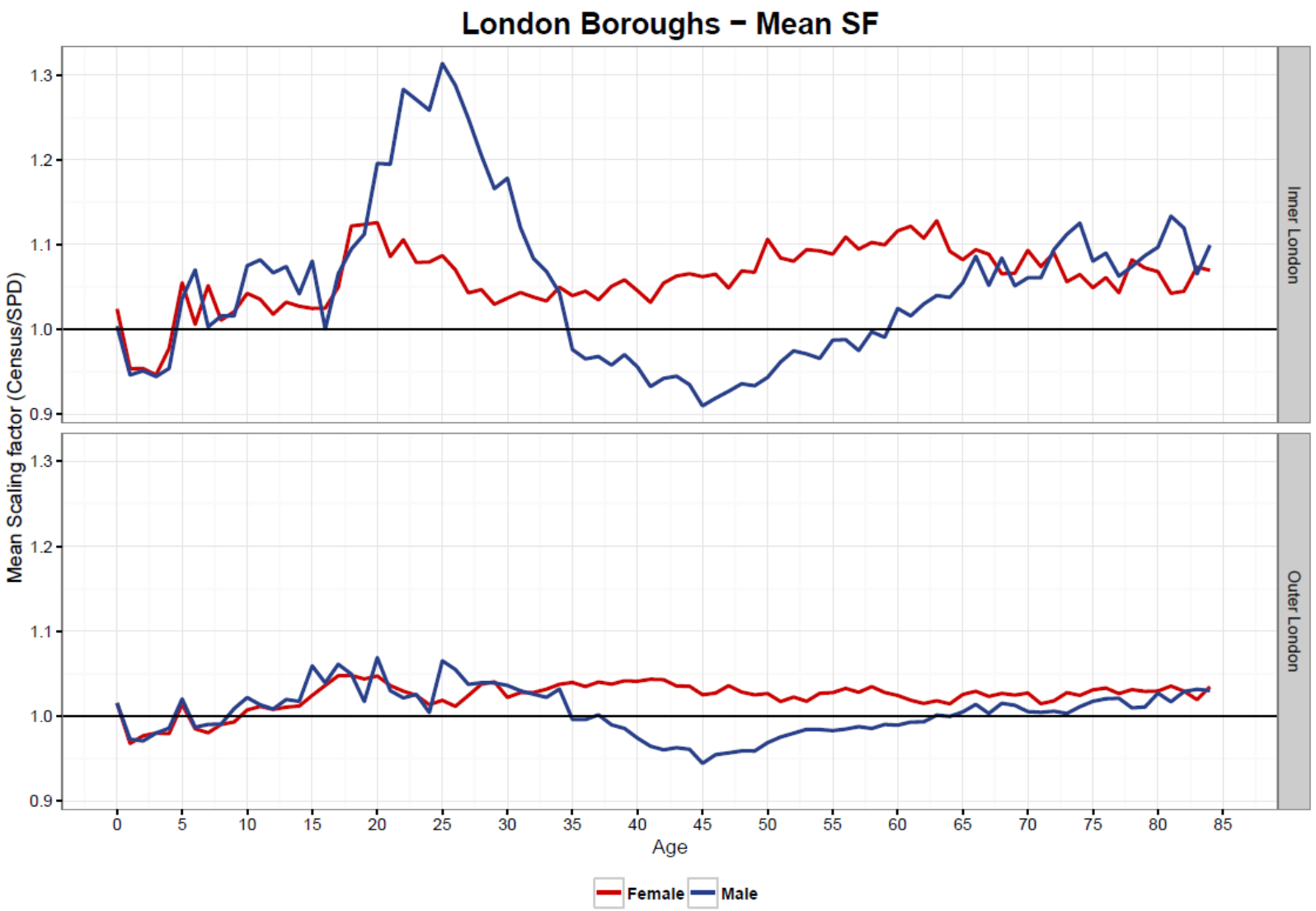
2011 Census/SPD scaling factors



Current research

- Documenting existing SPD methods to identify/ quantify errors from each stage/ process/ source
- Documenting intended changes in methods
- Identifying alternative methods for measuring uncertainty: direct vs benchmark approach
- Cluster analysis to identify LA typology
- SPD Uncertainty by LA, sex, SYOA
- Disaggregating MYE uncertainty to SYOA
- Produce/ publish MYE uncertainty measures for 2016, 2017-20


Females, London by SYOA
Scaling factor (Census/SPD)
1.6
1.2
0.8
Inner London (LA num 14)
Outer London (LA num 19)
0
5
10
15
20
25
30
35
40
45
50
55
60
65
70
75
80
85
Age


London Boroughs − Mean SF
Mean Scaling factor (Census/SPD)
1.3
1.2
1.1
1.0
0.9
Inner London
Outer London
0
5
10
15
20
25
30
35
40
45
50
55
60
65
70
75
80
85
Age
Female
Male