



Office for
National Statistics



From process to publication: understanding your census estimates

June/July 2012

Welcome and introductions

- Domestics
- What you can expect from the day

Session overview

- Aims for today
- Outline first release material
- How it all fits together

Aims for today

- Build confidence in the methods
- Improve understanding of methods to produce census population estimates.
- Show how the methods relate to material in first release.

What we won't be covering

- The census field operation
- Processes to capture, code and clean data from questionnaires
- Timetable and plans for more detailed releases and analysis
- any results or outcomes but
 - we will use real examples (anonymised!)

Outline of first release material

Census first release (1)

Statistical bulletin and tables

- Usually resident population (E&W):
 - Single year of age and sex at England and Wales level
 - Five year age and sex at Local Authority levels
- Short-term residents by LA
- Household estimates
- Results rounded to nearest 100

- Tables available to download as Excel tables
- Commentary to highlight key inter-censal and geographic changes

Census first release (2)

Explanatory material:

- Excel based tool to view QA materials for any Local Authority, includes:
 - Comparator data used in the QA process
 - Response rates and confidence intervals
 - Print/PDF friendly
 - Scope is limited by the content of the first release e.g. below Local Authority comparisons
- Series of more detailed papers explaining each of the components of the census estimates

How it all fits together

Why produce census estimates?

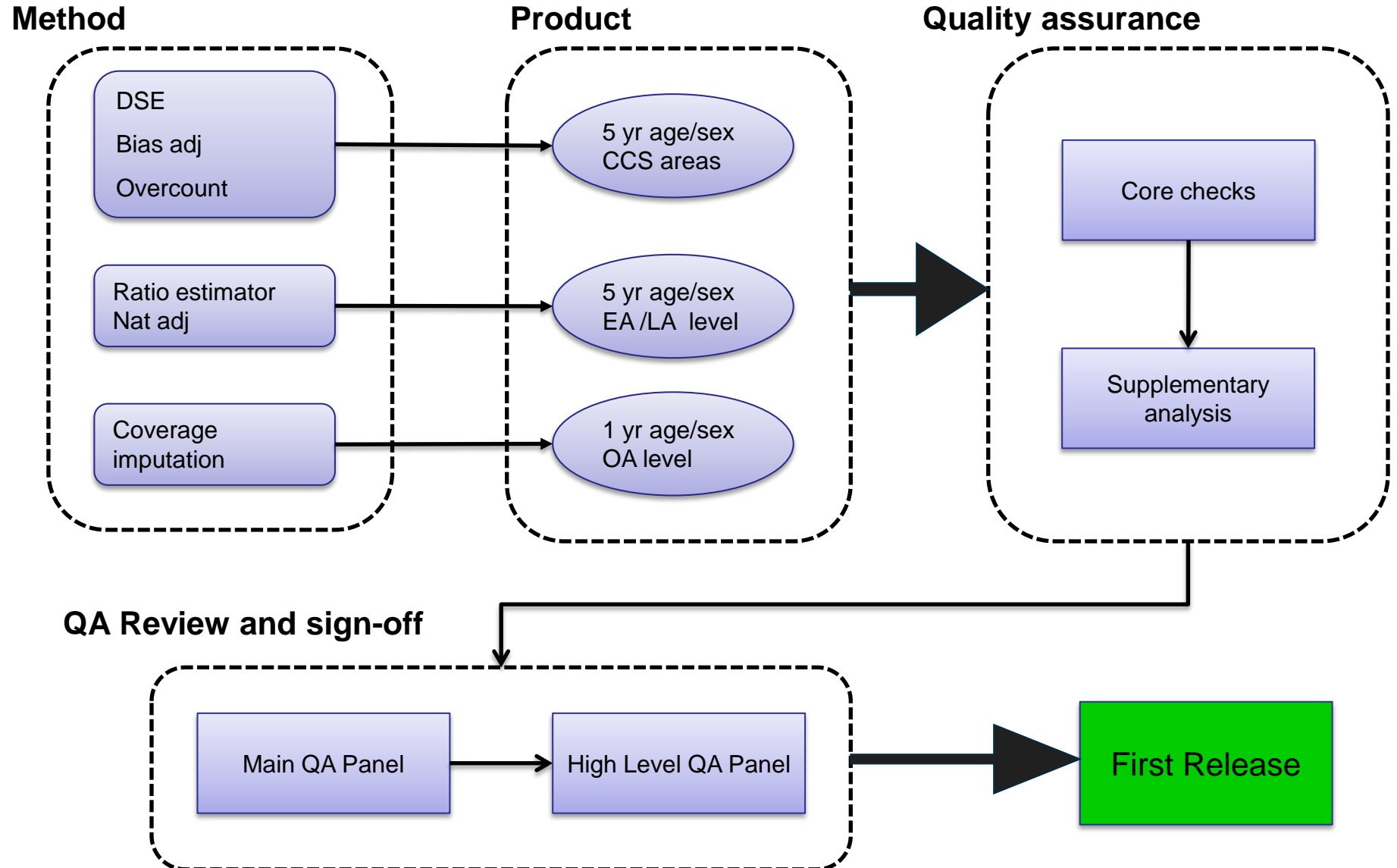
- Successful field operation though censuses never count every household or person
- They also count some people twice
- But, users need robust census estimates - counts not enough
- Estimate and adjust for under (and over) enumeration
- Improved the methodology used in 2001 to measure and adjust for undercount

Quality assuring the estimates

Objectives:

- Ensure 2011 Census estimates are fit for purpose
- Use comparator sources to identify discrepancies with census estimates
- Where required use contingencies to improve census estimates
- Ensure Census population characteristics are accurate
- Build user confidence through transparency in the methods

An overview of the methods



Census estimates - Key components

Component	Action
Raw Census count	Start
Dual system estimation	Add
Bias adjustment	Add
Overcount	Subtract
CE Adjustments	Add
National adjustments*	Add*
Census population estimates	Finish to QA

Agenda for the day

Welcome

Introduction

Estimating under-enumeration

-----Break-----

Creating an Alternative Household Estimate

Estimating for bias

-----Lunch-----

Estimating for overcount

Estimating for under-enumeration in Communal Establishments

Estimating for residual under-enumeration at the national level

-----Break-----

Quality assuring the census estimates

Questions and Answers

Summary

Questions

Estimating under-enumeration

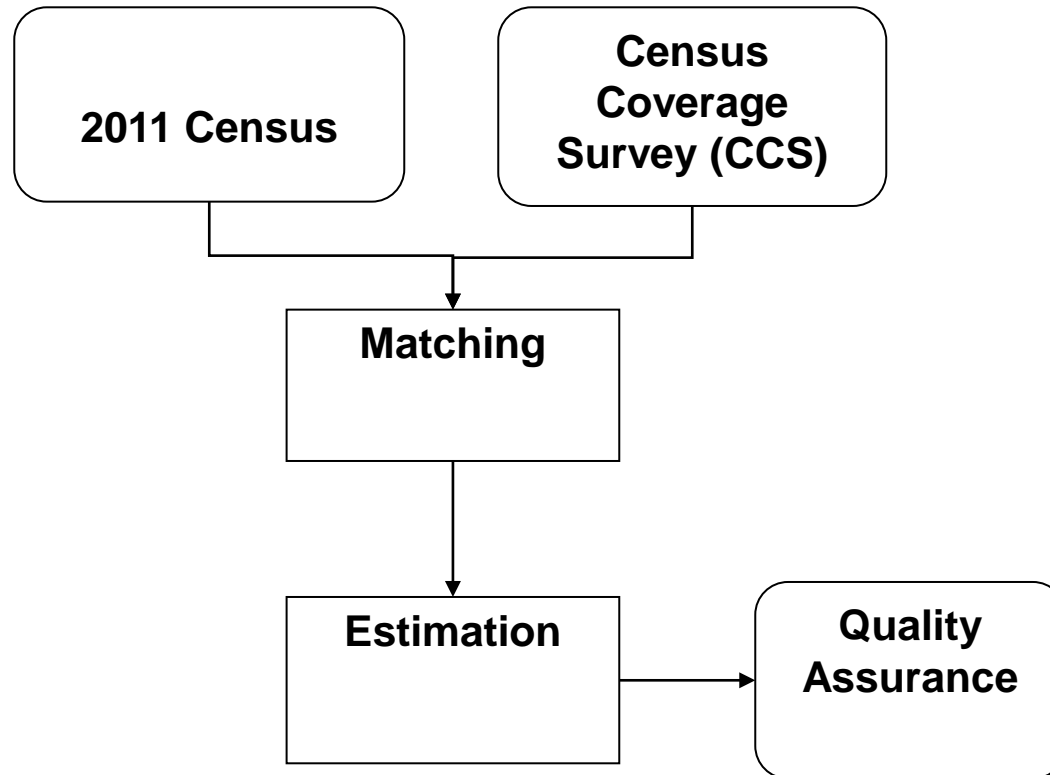
What this session will cover

- Quick overview of coverage process
- Focus on estimation process
- Worked example of estimation process using an anonymous case study
- Adjustments to estimates in later sessions

Overview of coverage process

- Coverage assessment:
 - Method for estimating the missed population
 - Based on a Survey
 - Uses standard statistical techniques
 - Produces estimates of population
 - Output database is adjusted by adding households and persons
- Quality assurance (this afternoon)
 - Checking plausibility of estimates and outputs

Coverage assessment overview



Case study area

- We will use a case study area to follow the estimation process
- This will help with:
 - Understanding the estimation process
 - Showing some of what you will see in the first release material

Estimation Areas and the HtC index

- Estimation Areas
 - Groups of contiguous LAs
 - Have enough sample for estimation
- Hard to Count index
 - Nationally consistent index
 - Built at LSOA level using data associated with non-response
 - Split into 40%, 40%, 10%, 8%, 2% distribution
 - Easiest lowest 40%, hardest top 2%

Case study – HtC index

- Our case study area is an EA with 4 LAs
- Our case study area has 1500 OAs and
- These are classified as follows:
- HtC 1 – 900
- HtC 2 – 540
- HtC 3 – 60

Census Coverage Survey

- **Reminder:**
 - Independent survey of small areas (postcodes)
 - Doesn't use address listing or any census information
 - Doorstep interview, ~13 questions
 - Prompts for population we know are missed (babies etc)
 - Call back lots of times
- Sample of 17,400 postcodes in 5,800 Output Areas = 340,000 households
- Sample of OAs for each LA by HtC
- Sample half postcodes in each OA
 - Called a 'cluster'

Case study – the CCS

- Our case study area has 1500 Output areas
- We sampled 41 of these – 21 in HtC1, 16 in HtC2 and 4 in HtC3
- From these OAs we sampled 158 postcodes in total, about 2500 households
- Sample fractions: 2.7% OAs, 1.3% postcodes, 1.2% households
- The CCS then managed to get valid interviews from 2040 households and 4500 persons (an 82% interview rate)

Processes prior to estimation

- Matching
 - Mixture of automated (65% household match rate, 59% person match rate) and clerical
- Resolving multiple matches (49 hhs)
- Resolving out of scope records (23 records)
 - Some forms of overcount
 - Strikethroughs, Localised duplicates, CCS errors etc
- Collapsing HtC (generally when less than 7 clusters)
 - Collapsed HtC 3 into HtC 2 for case study area
- Drop CCS postcodes where no data (1 postcode)

Estimation

3 parts to the estimation process:

(1) Dual System Estimation (DSE)

- What is the true population in the sampled areas

(2) Ratio Estimation

- Estimates for non-sampled areas
- Estimation Area (EA) level

(3) Local Authority Estimation

- Disaggregate EA level estimates to get LA level estimates

Part 1 – Dual System Estimation

Bang goes the theory

<http://www.bbc.co.uk/programmes/p00qq9c4>

Part 1 – Dual System Estimation

3 parts to the estimation process:

(1) Dual System Estimation (DSE)

- What is the true population in the sampled areas
- Makes adjustment for 'missed in both'
- Applied in each sampled cluster by age-sex

Dual System Estimation

- estimates those missed in both Census and CCS in each cluster by age-sex group

		Counted By CCS	
		Yes	No
Counted By Census	Yes	a	b
	No	c	d

- The DSE is $d = b \times c \div a$
- [Jonny's estimate was $((a+b)/a) \times (a+c)$]
- The total estimate is $a+b+c+d$
 - Initially assumes independence (more later)

Case study – DSE

- Males aged 35-44 in collapsed HtC 2
 - HtC 2 includes the HtC 3 clusters
 - Males 35-39 and Males 40-44 collapsed (more on this later)
- All clusters had some in this group in the Census or CCS

Case study – DSE (M35-44 in HtC 2)

Cluster	Both (a)	Census only (b)	CCS only (c)	Simple DSE(d)	DSE Total (a+b+c+d)
1	5	1	0	0	6
2	5	2	1	0.4	8.4
3	2	0	0	0	2
4	6	0	0	0	6
5	11	0	0	0	11
6	5	1	1	0.2	7.2
7	3	3	0	0	6
8	6	1	1	0.16666667	8.166667
9	9	2	0	0	11
10	1	0	0	0	1
11	9	5	0	0	14
12	13	1	0	0	14
13	7	0	0	0	7
14	5	1	0	0	6
15	13	1	0	0	14
16	4	0	0	0	4
17	5	2	3	1.2	11.2
18	12	0	0	0	12
19	10	3	1	0.3	14.3
20	5	0	0	0	5

Case study – DSE (M35-44 in HtC 2)

Both (a)	Census only (b)	CCS only (c)	DSE Total (a+b+c+d)	Chapman DSE Total
5	1	0	6	6
5	2	1	8.4	8.333333333
2	0	0	2	2
6	0	0	6	6
11	0	0	11	11
5	1	1	7.2	7.166666667
3	3	0	6	6
6	1	1	8.166667	8.142857143
9	2	0	11	11
1	0	0	1	1
9	5	0	14	14
13	1	0	14	14
7	0	0	7	7
5	1	0	6	6
13	1	0	14	14
4	0	0	4	4
5	2	3	11.2	11
12	0	0	12	12
10	3	1	14.3	14.27272727
5	0	0	5	5

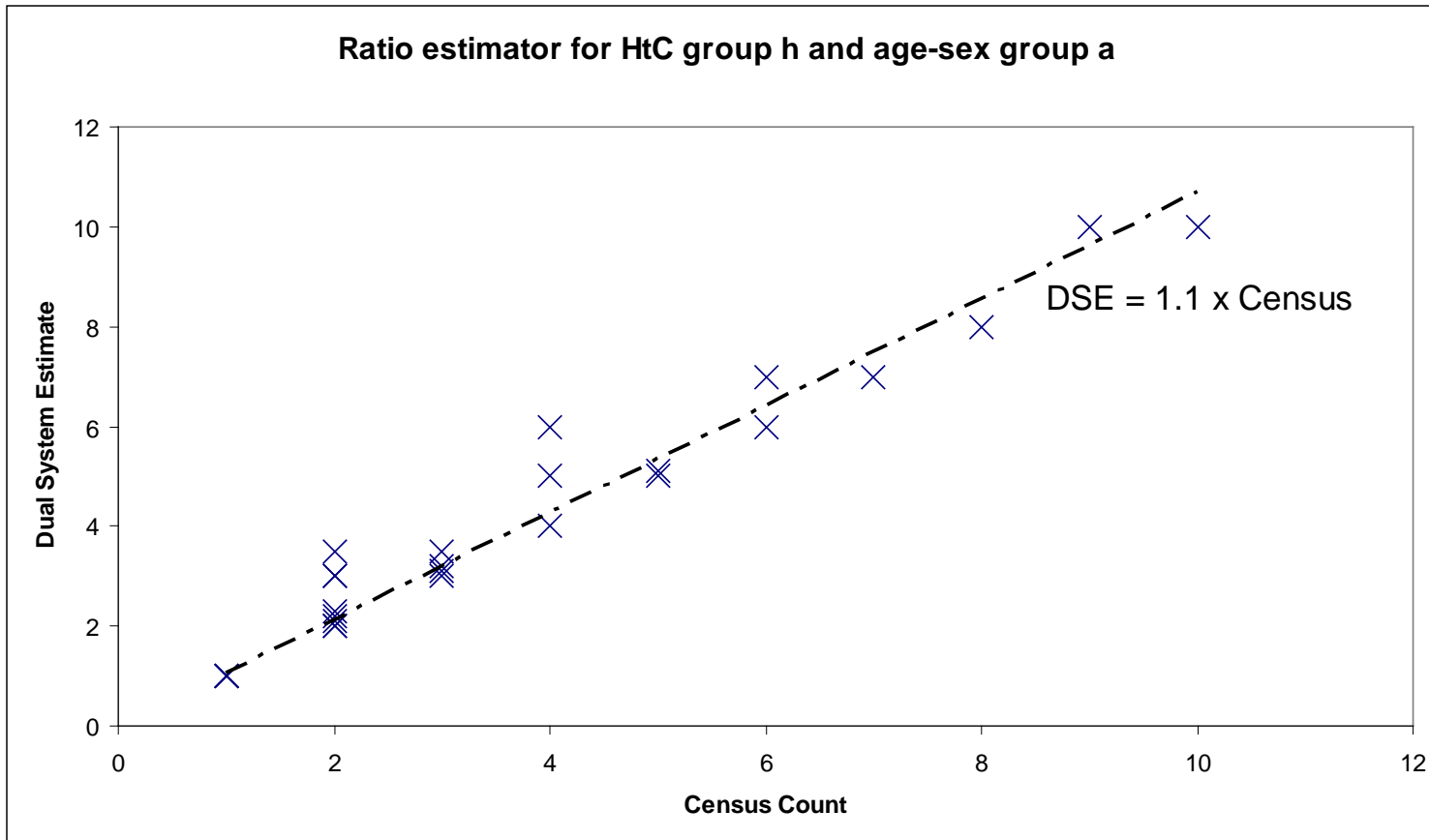
Part 2 – Ratio estimation

(2) Ratio estimation

- Estimates for non-sampled areas
- Estimation Area (EA) level
- Find relationship between DSE and Census count
- Line of best fit

Ratio estimation

- Coverage 'rate' is obtained by ratio between DSE and census count across the clusters (slope of the line of best fit through the origin)



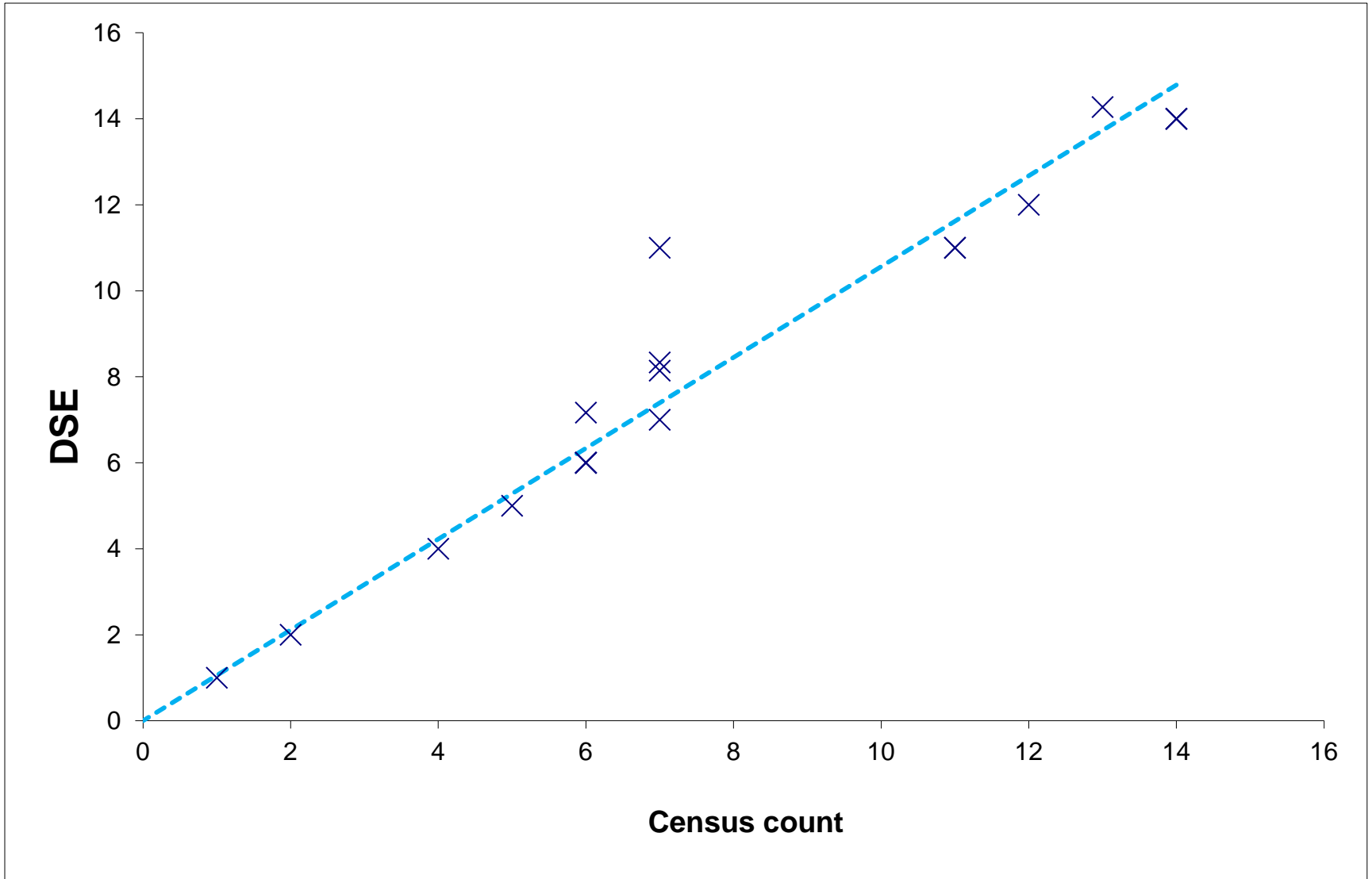
x Each point marks the DSE population and the Census count for an age-sex group in a cluster of postcodes within a hard-to-count stratum for an Estimation area.

Part 2 – Ratio estimation

(2) Ratio estimation

- Find Line of best fit between DSE and Census count
- Coverage rate is the sum of the DSEs divided by the sum of the Census in the sampled areas
- i.e. $\text{sum}(a+b+c+d) / \text{sum}(a+b)$
- or $\text{Sum (DSEs)} / \text{Sum (Census count)}$
- Census estimate is the rate applied to the **total** census count in that strata (age-sex by HtC)

Case study – Ratio estimates (M35-44 in HtC 2)



Case study – Ratio estimates (M35-44 in HtC 2)

- This is a plot of the DSE data seen previously
- The ratio is calculated as: $167.915 / 159 = 1.056$
- The Census counted 5057 males aged 35-39 and 5943 males aged 40-44 (in HtC2)
- So the estimates for these two groups for HtC 2 are:
 - $1.056 \times 5057 = 5340.5$
 - $1.056 \times 5943 = 6276.2$

Local Authority estimation

- Use age-sex by HtC patterns at EA level to get LA level estimates

Case study – LA estimation (M35-44 in HtC 2)

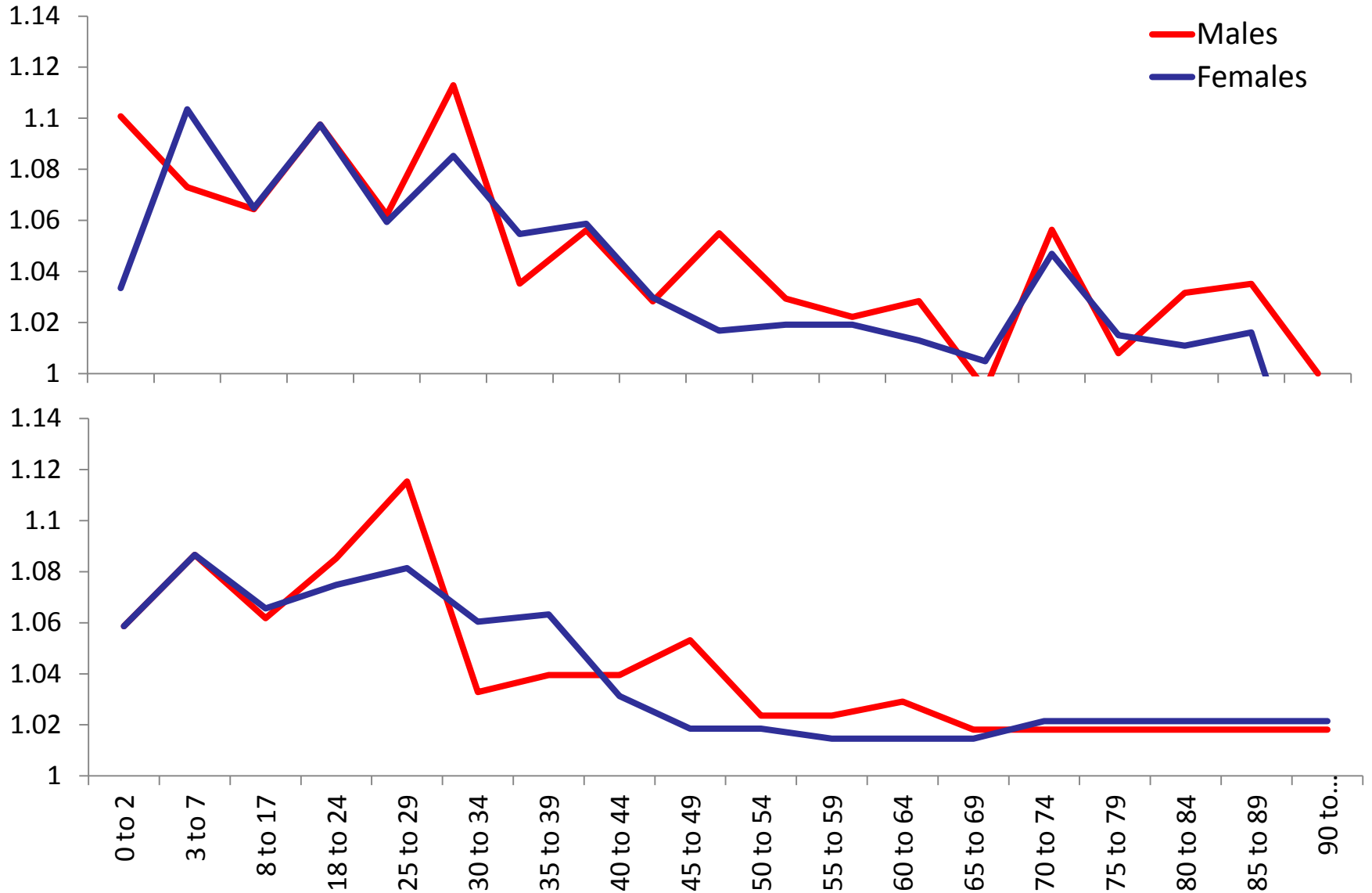
- Apply the 1.056 at LA level for Males 35-39 and Males 40-44 in HtC 2:

LA	Age-sex group	Census count	Estimate
1	M35-39	2200	2323.2
2	M35-39	870	918.7
3	M35-39	452	477.3
4	M35-39	1535	1621.0
1	M40-44	2423	2558.7
2	M40-44	1147	1211.2
3	M40-44	650	686.4
4	M40-44	1723	1819.5

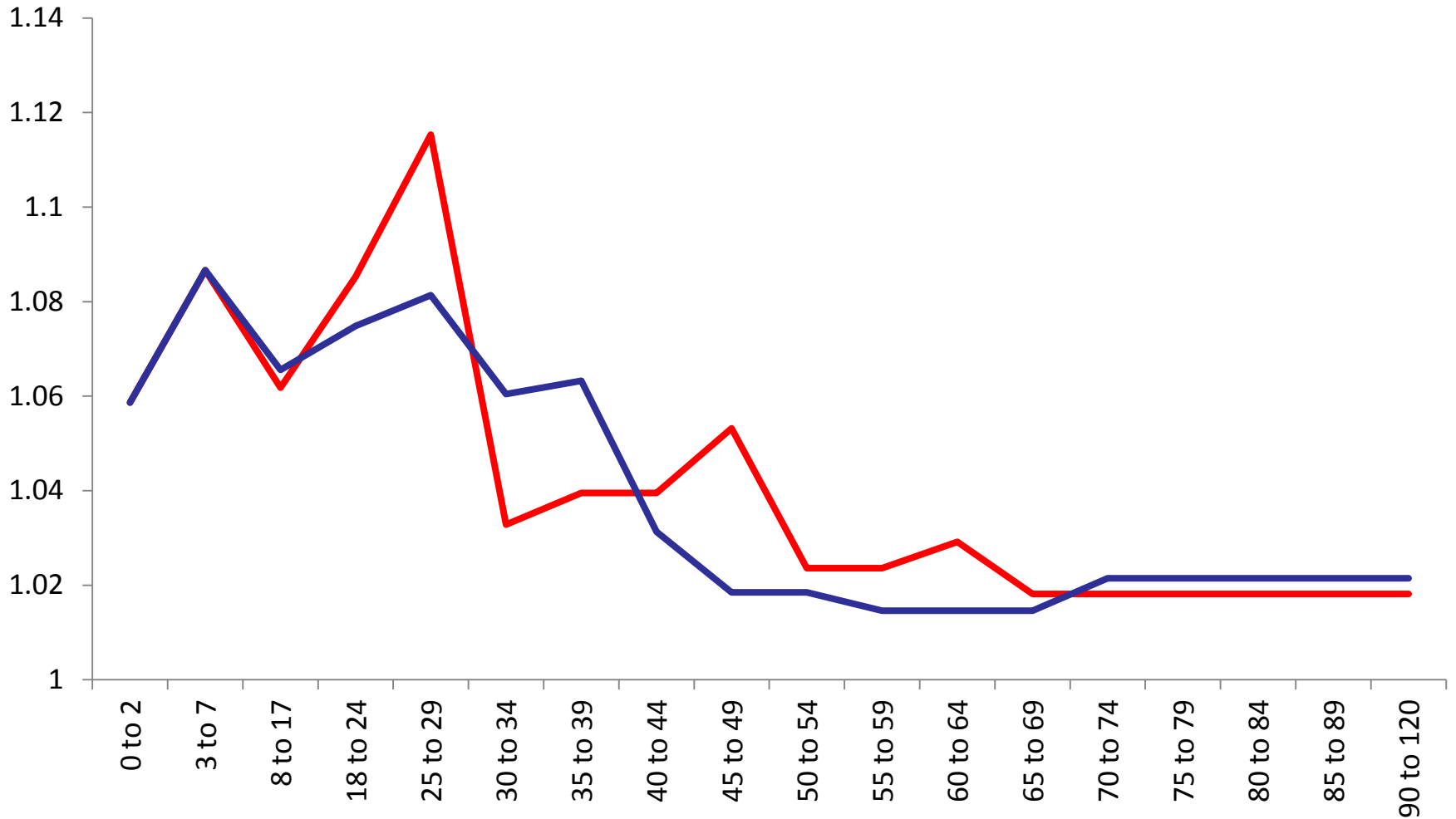
Collapsing in estimation

- We had standard rules for collapsing age-sex groups
- This helped to:
 - stabilise DSEs where sample sizes were small
 - stabilise ratios where sample sizes were small or data was inconsistent
 - reduce variance where there were outliers
- This was an iterative process as estimation and QA progressed

Case study – Impact of collapsing



Case study – Collapsed ratios



Case study – Summary

- All of the estimates can be aggregated to obtain 5 yr age-sex estimates by LA and EA
- And added to get to the total population
- For this EA the total estimate is 469643
- Compared to a census count of 450305
- Implies coverage is 95.9%

Case study - Key components

Component	Action	Number
Raw Census count	Start	450,305
Dual system & Ratio estimation	Add	19,338
Bias adjustment	Add	0
Overcount	Subtract	0
CE Adjustments	Add	0
National adjustments*	Add*	0
Census population Estimates	Finish to QA	469,643
Quality Assurance	Sign-off estimates	

Confidence intervals

- A 95% confidence interval is a measure of sampling variability/reliability/confidence in the estimate
- ‘If we did the CCS 100 times, approximately 95 times the true value would be within the interval’
- Obtained using a bootstrap replication method

Case study – Confidence intervals

The 95% confidence intervals are:

- Males 35-39 in HtC 2 – (4886.1 , 5794.4)
 - Estimate is 5340.5
- Males 40-44 in HtC 2 – (5723.6 , 6828,4)
 - Estimate is 6276.2
- i.e. the estimate plus or minus 8.5%
- Total EA population – (461601 , 477546)
- i.e. plus or minus 1.7%
- (Note CIs are smaller for large populations)

Coverage adjustment

Coverage adjustment

- Estimation produces LA by age-sex estimates
 - With confidence intervals
- Imputation process imputes households and persons
 - Uses CCS data to decide characteristics of the missed, inc Ethnicity, Tenure, ALW, Migrant status
 - Also provides the other characteristics of those missed (for those variables not measured in CCS)
 - Places households into dummy questionnaires (i.e. into a postcode and Output Area)

Summary

- This session has gone through the basic estimation process
- The next sessions look at how improvements can be made when some of the assumptions underpinning the methods are not met
- These can result in bias
 - Bias is when the estimates will always be too low or too high (if the Census/CCS were to be repeated)

Creating an alternative household estimate

Overview

Alternative estimate of occupied households

Estimates produced

- for each Estimation Area
 - for CCS postcode clusters only
 - by Hard to Count Group
-
- Alternative household estimate compared against the DSE: to assess for negative bias

Methodology

Usually resident households

+

A proportion of dummy forms

+

A proportion of blank questionnaires

+

A proportion of unaccounted for addresses

+

A proportion of additional addresses identified from
March 2011 address products (NLPG and PAF)

Usually resident households

- Questionnaire returned with one or more usual residents
- Excludes short term migrant only households, or dwellings with no usual residents (e.g. second homes)

Dummy forms

- Dummy forms completed by field staff if no response at an address
 - Field staff assess occupancy of dwelling
 - Misclassifications can occur if non-contact
- RMR 'remove multiple response' data used to calculate dummy form misclassification rates
- Used to estimate the proportion of dummy forms that were occupied

Blank questionnaires

- 18% of blank form images clerically reviewed to identify:
 - if occupied (e.g. *'I'm not filling this in'*)
 - or unoccupied/invalid (e.g. *'This is a post office'*)
- Sample focussed on CCS areas
- Results from clerical work used to estimate the proportion of blank questionnaires that were occupied

Unaccounted for addresses

- Addresses with no questionnaire return, deactivation or dummy form
- Field exercise checked 15% of UFAs
 - Focussed in CCS areas and those with greatest proportion of UFAs
 - Dummy forms completed for genuine households; or address deactivated
- For the remainder of UFAs: The proportion occupied was estimated based on field check results

Additional addresses

- Source products used to create Census address register were “cut-off” in December 2010
- Additional addresses in March 2011 version of PAF and NLPG identified
- Numbers adjusted to determine likely occupied

Case study

	Number of addresses	Proportion occupied	Alternative household estimate
Occupied Households	1,164	100%	1,164
Dummy questionnaires (reason code = 'occupied')	4	74%	3
Dummy questionnaires (reason code = 'non contact')	54	86%	47
Dummy questionnaires (reason code = 'unoccupied')	48	39%	19
Blank questionnaires	3	5%	0
Unaccounted for addresses	20	41%	8
Additional addresses	0	100%	0
			1,241

Validation of process

- Alternative Household Estimates by LA also produced, for validation
 - Less accurate than estimates for CCS postcode clusters
- Census estimates of occupied households quality assured against other sources e.g.
 - Council Tax
 - Patient Register
 - Household estimates from CLG

Estimating for bias

Estimating for bias

- DSE can be biased when its assumptions are not well met
- Two types:
 - Between household bias – e.g. when households that are not likely to be counted in the census are also not likely to be counted in the CCS
 - Within household bias – e.g. when persons that are not likely to be counted in the census in a counted household are also not likely to be counted in the CCS

Estimating for bias

- Example of between household bias
 - a household that will always refuse in Census **and** CCS
 - or a household that changes its behaviour in the CCS dependent on its Census outcome (i.e. I filled in your questionnaire, I don't want to do another)

Estimating for bias

- Example of within household bias
 - a person within a **counted** household that will always be excluded in Census **and** CCS (i.e. partner of single parent mother due to benefit fraud)

Estimating for bias

- We assess between household bias using the AHE
- We assess within household bias using social survey data
- Note: This is the equivalent the 2001 'dependence' adjustment

Estimating for between hh bias

- Within each HtC stratum
- If the AHE $>$ Household level DSEs for the sample, then there is between household bias

Estimating for within hh bias

- Social survey data matched to Census data
- Analysed within household coverage by Region, HtC and broad age-sex (where sample sizes were sufficient)
- If the Social Survey found significantly lower coverage within households than the CCS then there is within household bias

Adjusting for DSE bias

- Based on the AHE and Survey information
- A model is used to work out the adjustments to apply to the DSEs by age – sex
- This takes the adjustment needed at household level and works out what adjustment is needed at person level
- The adjustments are multiplying factors to apply to the person level estimates

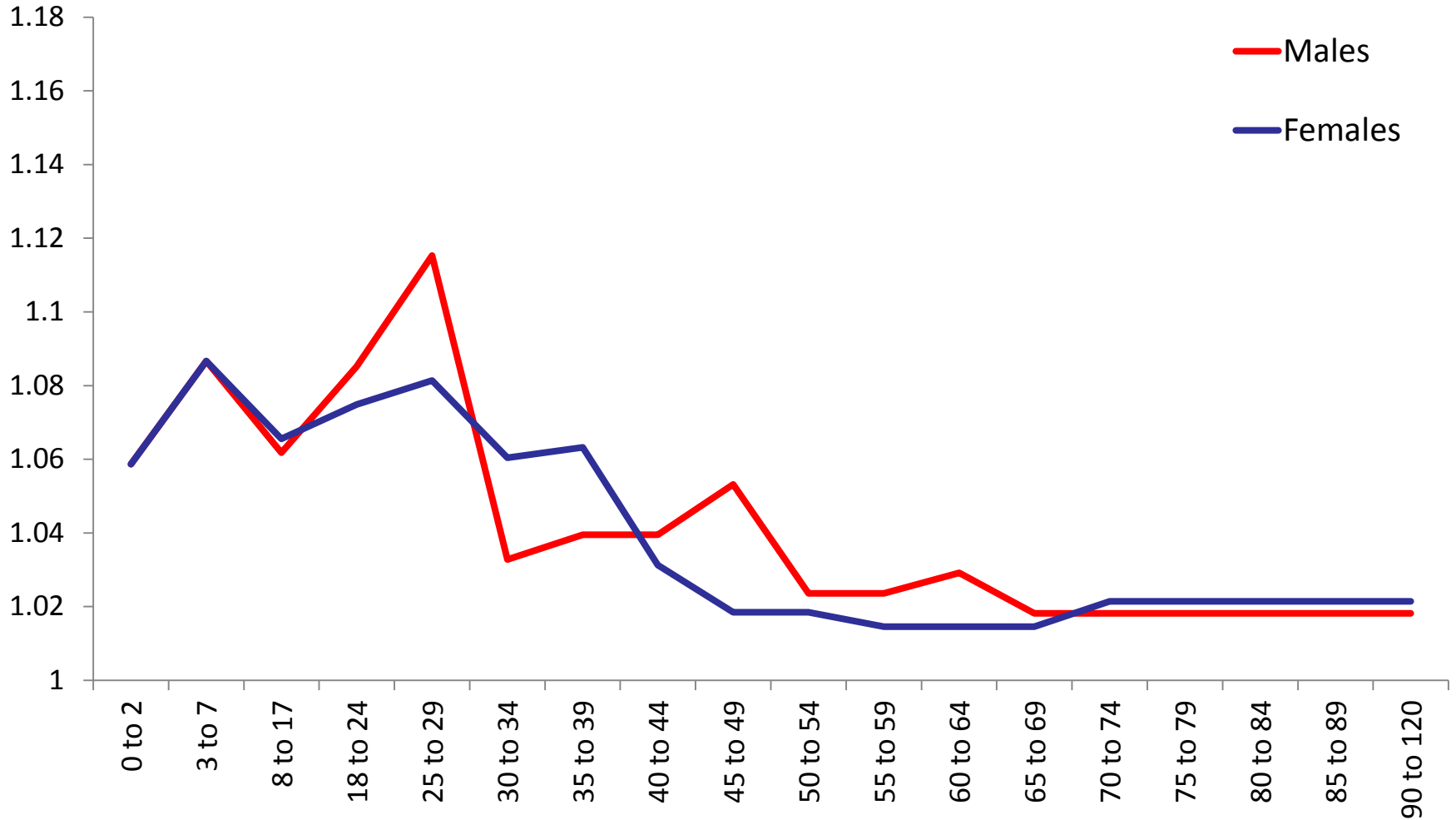
Case study – Bias adjustment

- The AHE for HtC 2 was 1241
- The DSE by tenure for households in HtC 2 was 1198.6
- No evidence of within household bias in this area
- So a bias adjustment made on the basis of the AHE so that the household DSE by tenure will be 1241
- For Males 35-39 in HtC 2 the model for adjustment calculates a bias adjustment factor for this group at person level of 1.051

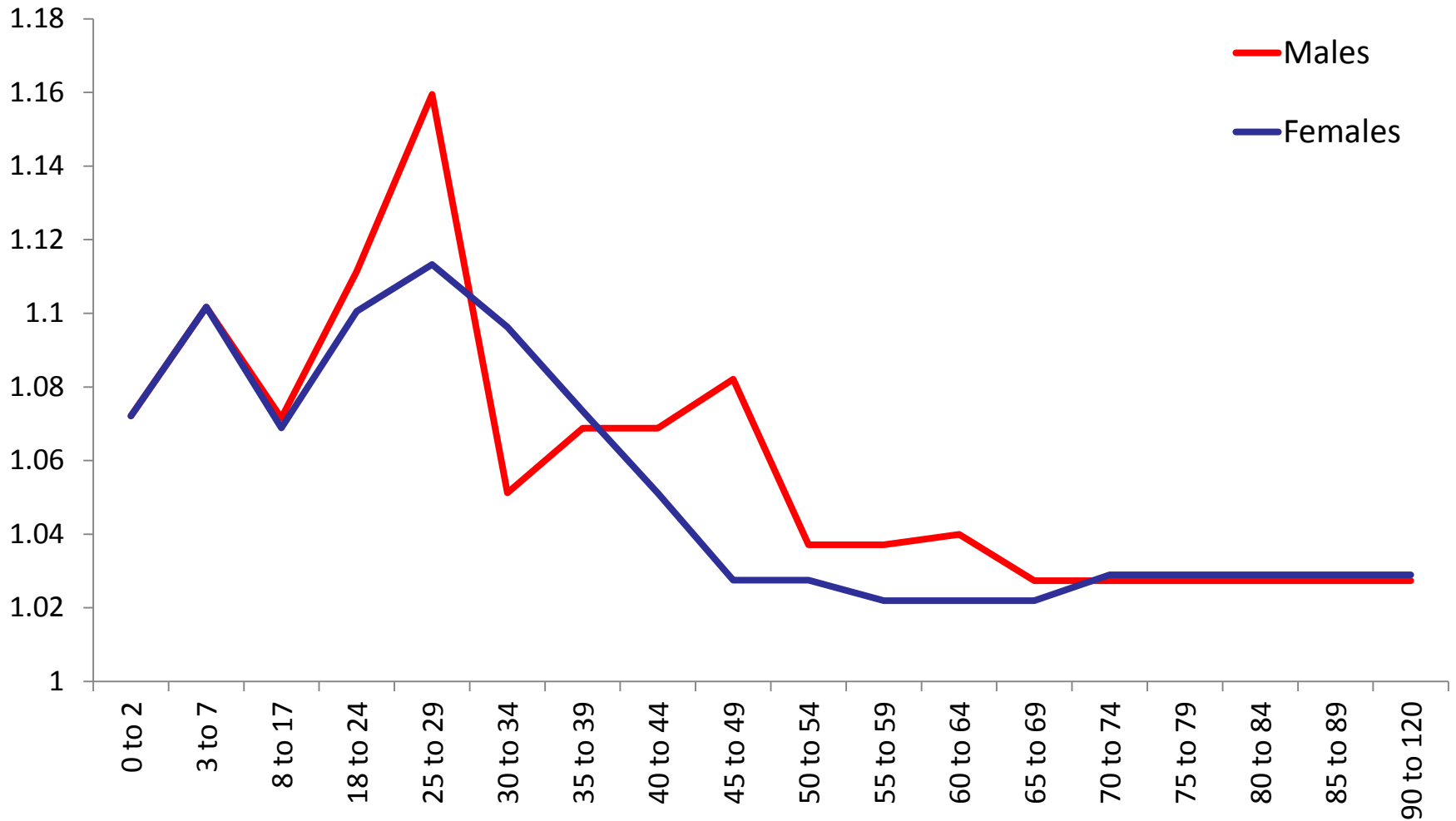
Case study – Bias adjustment

- For Males 35-39 in HtC 2 the adjustment factor of 1.051 is applied to the estimate
- So the new estimate is $1.051 \times 5340.5 = 5612.9$
- The adjustment factor varies according to:
 - Coverage levels in CCS
 - Split between missed in counted/wholly missed households
- Not always high (for example in this area the adjustment factor for older persons is <1.01)

Case study – Before bias adjustment



Case study – After bias adjustment



Case study – Bias adjustment

- The adjusted census estimate is 475779
- (The unadjusted estimate was 469643)
- Compared to a census count of 450305
- Implies coverage is now 94.6%
- The adjustment is also made at LA level

Case study - Key components

Component	Action	Number
Raw Census count	Start	450,305
Dual system & Ratio estimation	Add	19,338
Bias adjustment	Add	6,136
Overcount	Subtract	0
CE Adjustments	Add	0
National adjustments*	Add*	0
Census population estimates	Finish to QA	475,779
Quality Assurance	Sign-off estimates	

Estimating for overcount

Estimating for overcount

- Two types of person level overcount:
- Duplication
 - e.g. Child of separated parents
 - Student at term time address and with parents
- Counted in the wrong location
 - e.g. Student counted at parents address and NOT at term time address
 - Person who moved prior to census day but sent back questionnaire early

Estimating for overcount

- Note we don't remove duplicates from the database, we make a net adjustment
- Estimated regionally
- Combination of:
 - Searching for duplicates in a large sample of census persons (measures duplication)
 - Wider searching for all persons in the CCS sample (measures duplication and in wrong place)

Estimating for overcount

- Outcome is a set of regional overcount propensities by:
 - Hard to Count and
 - Broad age (3-17, 18-24, 85+, the rest) and
 - Student or not (18-24 only)
- These are used to weight each census individual in the DSE
 - Each person counts for 0.99 instead of 1

Case study –overcount

- For the region that contains this EA:
- Sampled 400,000 records (about 5%) and found 6100 duplicates
- When combined with CCS information, estimated overcount propensity for Persons aged 0-2 or 26-84 (i.e. the 'rest' group) in HtC 2 was 1.00393
- This means overcount for this group in this region is about 0.4%

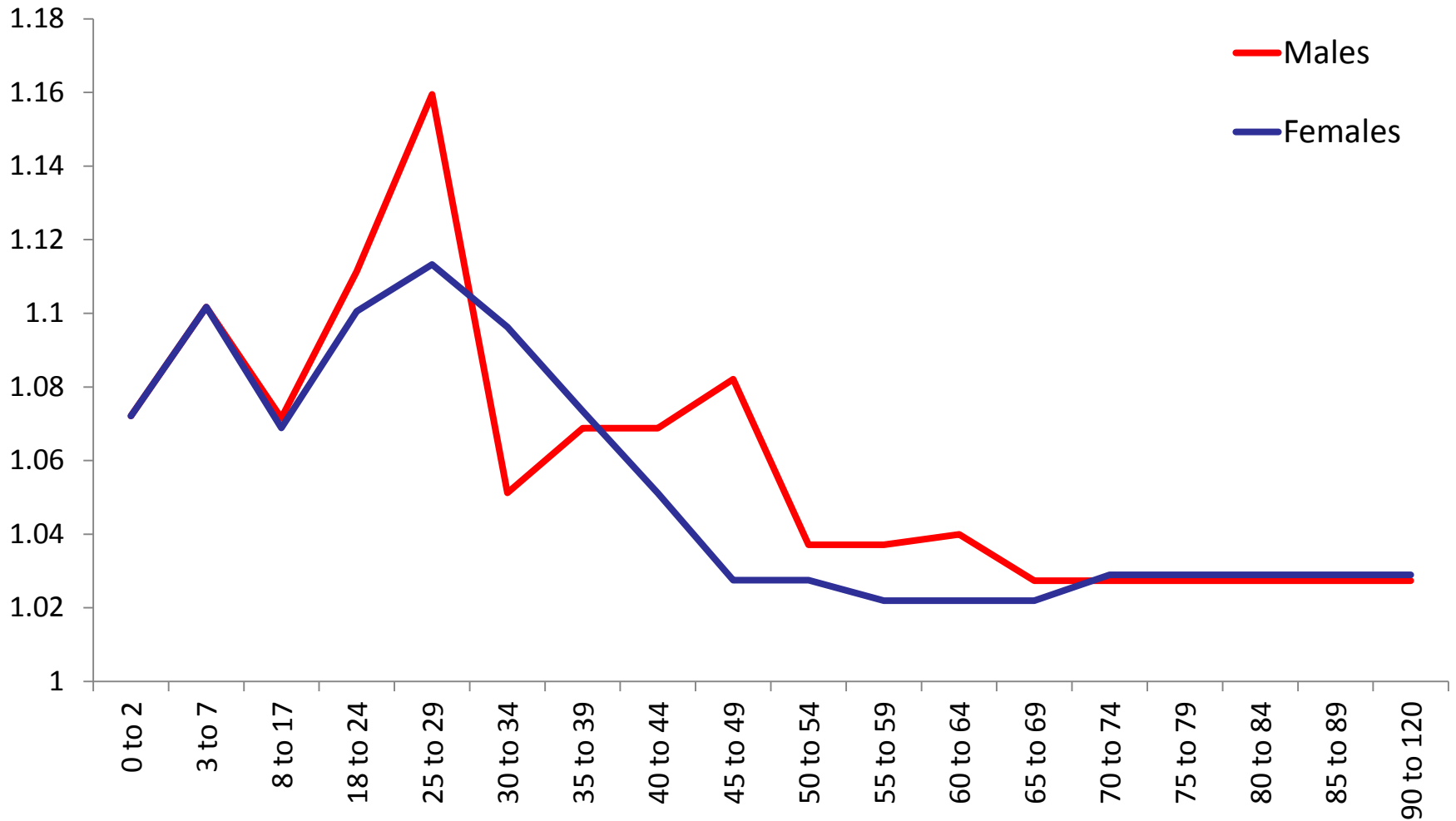
Case study – Overcount revised DSEs

Both (a)	Census only (b)	CCS only (c)	Chapman DSE Total	Chapman DSE Total with overcount
5	1	0	6	5.977
5	2	1	8.333	8.301
2	0	0	2	1.992
6	0	0	6	5.977
11	0	0	11	10.957
5	1	1	7.167	7.139
3	3	0	6	5.977
6	1	1	8.143	8.112
9	2	0	11	10.957
1	0	0	1	0.996
9	5	0	14	13.945
13	1	0	14	13.945
7	0	0	7	6.973
5	1	0	6	5.977
13	1	0	14	13.945
4	0	0	4	3.984
5	2	3	11	10.959
12	0	0	12	11.953
10	3	1	14.273	14.217
5	0	0	5	4.980

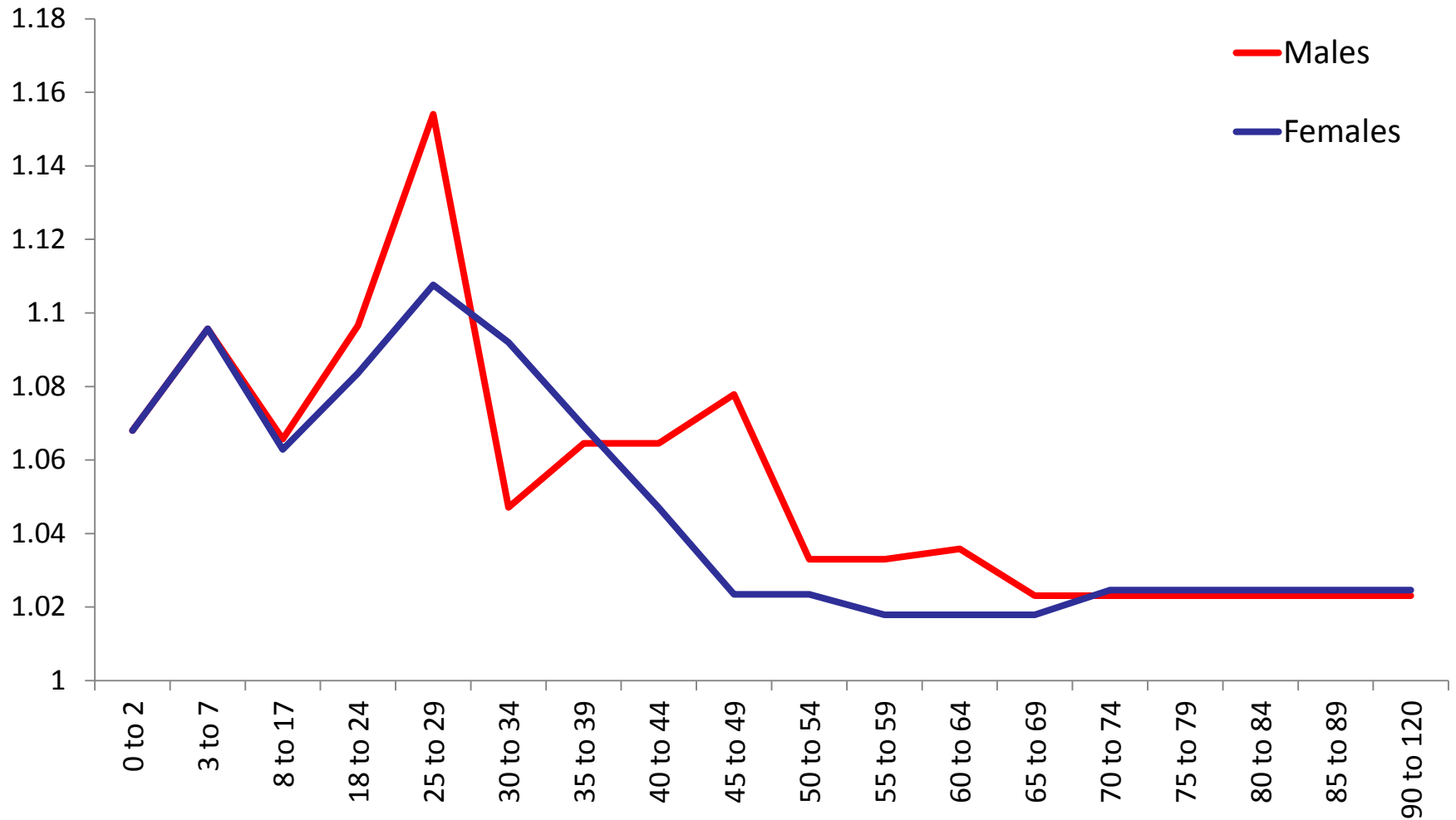
Case study – overcount

- The DSEs are a bit smaller, and sum to 167.263 (it was 167.915 before)
- So the new ratio estimate is $167.263 / 159 = 1.052$
- And the so revised estimate for Males 35-39 in HtC 2 is $1.052 \times 5057 \times 1.051 = 5591.1$
- Note the bias adjustment still applies
- The previous estimate (inc bias adjustment) was 5612.9

Case study – After bias adjustment



Case study – Overcount revised ratios



Case study – Overcount

- The adjusted census estimate is 473387
- (The previous estimate was 475779)
- Compared to a census count of 450305
- Implies coverage is now 95.0%
- So overcount in this EA is about 0.3%
- Note we don't remove duplicates from the database, we make a net adjustment

Case study - Key components

Component	Action	Number
Raw Census count	Start	450,305
Dual system & Ratio estimation	Add	19,338
Bias adjustment	Add	6,136
Overcount	Subtract	-2,392
CE Adjustments	Add	0
National adjustments*	Add*	0
Census population estimates	Finish to QA	473,387
Quality Assurance	Sign-off estimates	

Estimating for under- enumeration in Communal Establishments

Communal Establishments

Component	Action	Number
Raw Census count	Start	450,305
Dual system & Ratio estimation	Add	19,338
Bias adjustment	Add	6,136
Overcount	Subtract	-2,392
CE Adjustments	Add	0
National adjustments*	Add*	0
Census population estimates	Finish to QA	473,387
Quality Assurance	Sign-off estimates	Yes

Communal Establishments

- Communal Establishments (CEs) are managed residential accommodation
- CE address register – based on third party sources supplemented with field checks and Local Authority engagement (twice)
- Each CE sent a CE questionnaire plus questionnaires for each individual
- Enumerated by 1,744 special enumerators
- This section looks at how estimates were made for under-enumeration in communal establishments- large and small
- Examples include halls of residence, armed forces bases and prisons

Small Communal Establishments

- A small CE has up to 99 bed spaces
- Covered by Census Coverage Survey
- Dual System Estimation approach used as for households
- Estimates made by region, broad CE type and broad age-sex
- Estimating for under-coverage within a CE
- **For our exercise – assume small CE adjustment = 598**

Large Communal Establishments

- A CE with 100 or more bed spaces
- Not covered by Census Coverage Survey
- Dual System Estimation not used to estimate under-coverage

- Quality assurance and adjustment based on case by case assessment of:
 - Returns for each CE
 - Administrative data for each CE

Assessment of returns

- Further investigation carried out where:

The number of individuals who didn't return a form was 50 or more
or

Where the return rate was less than 75%

- Large CE Return rate =
$$\frac{\text{Individual Questionnaires Returned}}{\text{Individual Questionnaires Issued}^*}$$

*Questionnaires issued minus any deactivations in the field

Assessment Against Administrative Data (1)

Large CE Type	Administrative Source
Student Hall of Residence	Higher Education Statistics Agency (HESA)
Boarding Schools	Department for Education (DfE)
Prisons	Ministry of Justice
Immigration Removal Centres	UK Borders Agency (UKBA)
Residential/Nursing Homes	NHS Patient Register
Armed Forces Bases	Defence Analytical Services Agency (DASA)

Assessment Against Administrative Data (2)

- CEs matched between Census and Administrative Source
- Work carried out to ensure consistency between administrative data and census. For example:
 - School Boarder data originally referred to age at 1 January 2011. This was aged on to approximately relate to census day
 - Higher Education data filtered to only include individuals with a communal establishment flag
- Further work carried out when the administrative data was 50 or more greater than the census count for the CE

Adjustments made

- Adjustments made by calibrating to administrative data
- Direct contact made with large CEs where there was inconsistency between administrative data and the number of forms issued
- Approximately 100 cases where direct contact was made (mainly halls of residence)
- Further discussions held with suppliers of administrative data (Department for Education (DfE), Ministry of Justice (MoJ))
- Census field intelligence was also used – e.g. Record books completed by special enumerators

Case study 1

University Hall of Residence

- Questionnaires issued = 237
- Completed questionnaires = 136
- CE Return rate = 57.4%
- Forms not returned = 101

- Census CE count of individuals = 136
- HESA CE count = 241

- This was adjusted to without contacting the establishment.
- **Large CE adjustment made of 105**

Case study 2

Boarding School

- Questionnaires issued = 424
- Completed questionnaires = 402
- CE Return rate = 94.9%
- Forms not returned = 22

- Census CE count of individuals = 402
- DfE CE count = 675

- The school was contacted. They provided a count of 422 students in their accommodation.
- No adjustment was made

Back to Case study

Component	Action	Number
Raw Census count	Start	450,305
Dual system & Ratio estimation	Add	19,338
Bias adjustment	Add	6,136
Overcount	Subtract	-2,392
CE Adjustments	Add	703
National adjustments*	Add*	0
Census population estimates	Finish to QA	474,090
Quality Assurance	Sign-off estimates	Yes

Estimating for under- enumeration at the national level

What are we assessing?

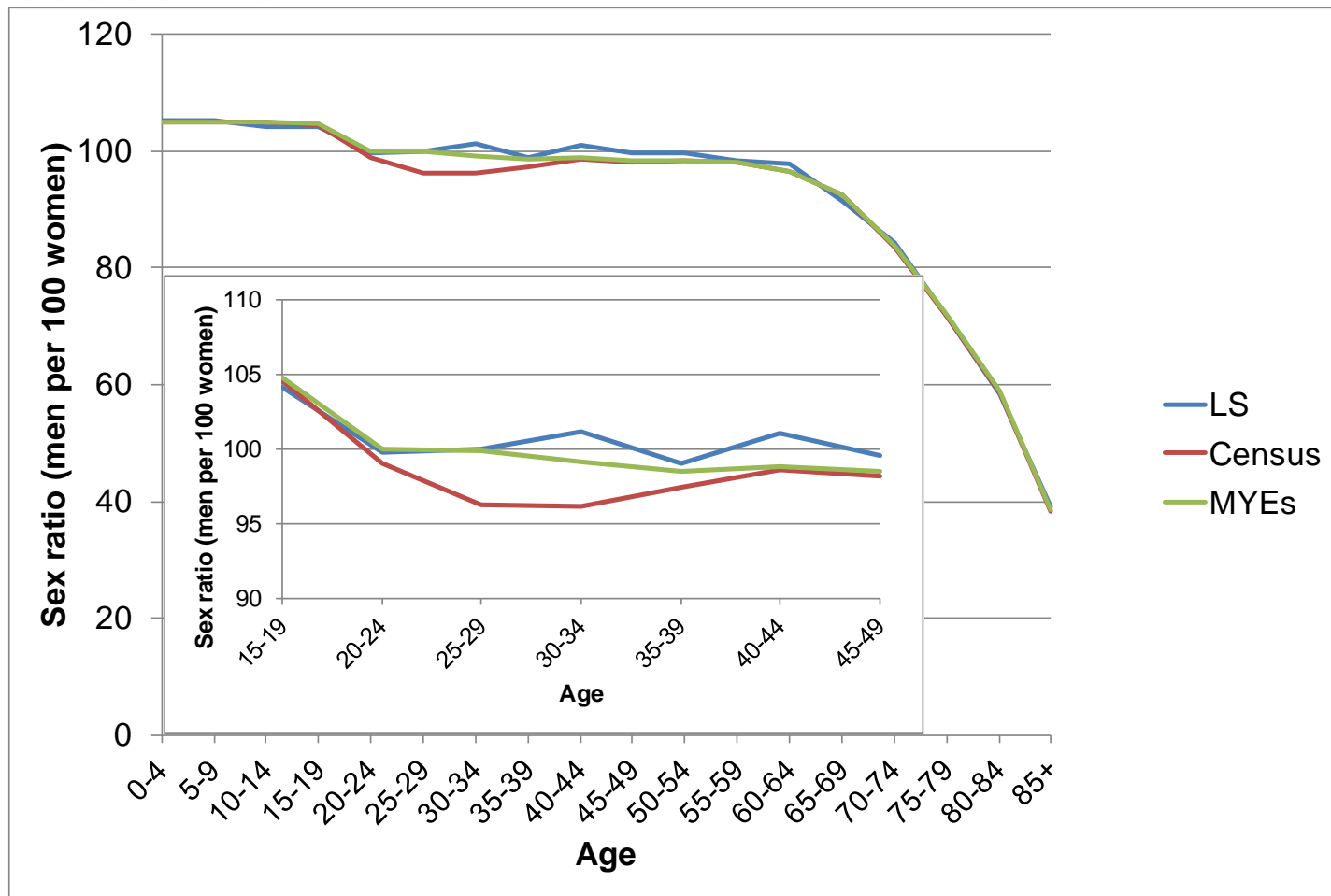
- Most adjustments in Census bottom up:
 - Estimation
 - Bias
 - Communal Establishments
 - Overcount
- Assessing national estimates for any residual under (or over) enumeration
- Note much of adjustments to MYEs following 2001 was to address residual under-enumeration

Method (1)

- Compare alternative sex ratio patterns from other sources with census estimates
 - ONS Longitudinal Study 2011 link,
 - implied ratios from demographic analysis,
 - Lifetime Labour Market database
- Does the evidence suggest an adjustment is required?

Example 2001 – post Census adjustment

Used ONS LS to derive potential number of men missing and added them in.



Method (2)

- Methods developed (and published) to adjust if evidence suggests necessary
 - derive a sex ratio target
 - decide whether one or both sexes to be adjusted
- Decide on method to geographically distribute
 - Proportional to population size
 - Proportional to coverage adjustment
 - Proportion missed by both (correlated with census coverage and CCS coverage)

Key components – case study

Component	Action	Number
Raw Census count	Start	450,305
Dual system & Ratio estimation	Add	19,338
Bias adjustment	Add	6,136
Overcount	Subtract	-2,392
CE Adjustments	Add	703
National adjustments*	Add*	0
Census population estimates	Finish to QA	474,090
Quality Assurance	Sign-off estimates	No

Quality assuring the census estimates

Session overview

Component	Action	Number
Raw Census count	Start	450,305
Dual system & Ratio estimation	Add	19,338
Bias adjustment	Add	6,136
Overcount	Subtract	-2,392
CE Adjustments	Add	105
National adjustments*	Add*	0
Census population estimates	Finish to QA	474,090
Quality Assurance	Sign-off estimates	No

Session overview

- Quality Assurance (QA) overview
- What was considered - the QA evidence
- How it was considered – the QA panels
- Demonstrating QA through practical examples
 - QA of Estimation
 - QA of Final Estimates

Quality Assurance overview

Achieving Quality Estimates

- Quality built in throughout process
 - Design (census form and field work)
 - Operational management (up-to-date questionnaire tracking)
 - Data processing (checking consistency of scanned information)
 - Coverage estimation (census estimates rather than simply counts)
 - Quality assurance process (validation of what was collected and estimated)
 - Quality measurement (response/return rates and confidence intervals)

What did we say we'd do

- Evidence routinely considered
 - Checks against other estimates and administrative sources
 - Demographic analysis
 - Profiles of each local authority area
 - Operational intelligence
 - Cumulative checking data
- 'Supplementary' analysis
 - Low level aggregate comparisons
 - Local authority supplied evidence
 - Cross checking estimates at different processing stages

What did we do differently?

- Supplementary analysis routinely carried out e.g. below LA level
- Greater emphasis on diagnostics from processing – particularly coverage estimation
- Prioritised some of the checks which proved most useful (age-sex, households)
- Local Authority provided intelligence – specifically locally provided Council Tax data routinely used
- More detailed investigations into Mid-Year Estimates than originally proposed

What evidence was considered?

Evidence assessed for all Local Authority estimates

1. Checks against other estimates and administrative sources
2. Demographic analysis
3. Profiles of each Local Authority area
4. Operational intelligence
5. Diagnostics from estimation & adjustment processes

Evidence assessed for all Local Authority estimates

- Comparator sources will not match exactly due to:

- Definition
- Coverage
- Accuracy/timeliness

(Paper published May 2012 – Administrative sources used in census QA)

- Tolerance bounds derived for each Local Authority estimate
- Checks include:
 - Age-sex
 - Household number/size
 - Ethnicity
 - Students
 - Armed Forces
 - International Migration

Comparator checks and data sources

QA Check	Comparator dataset
Age and sex	<ul style="list-style-type: none">• Patient Register• Mid-year Population Estimates*• School Census• Child benefit/pensions data
Household Number and Average Size	<ul style="list-style-type: none">• Council Tax• Address Register• Patient Register• Communities and Local Government household projections
Ethnicity	<ul style="list-style-type: none">• Population Estimates by Ethnic Group• Integrated Household Survey• School Census

- Mid-2011 Population Estimates rolled forward (extrapolated) from published mid-2010 estimates including recent improvements to migration statistics

Comparator checks and data sources

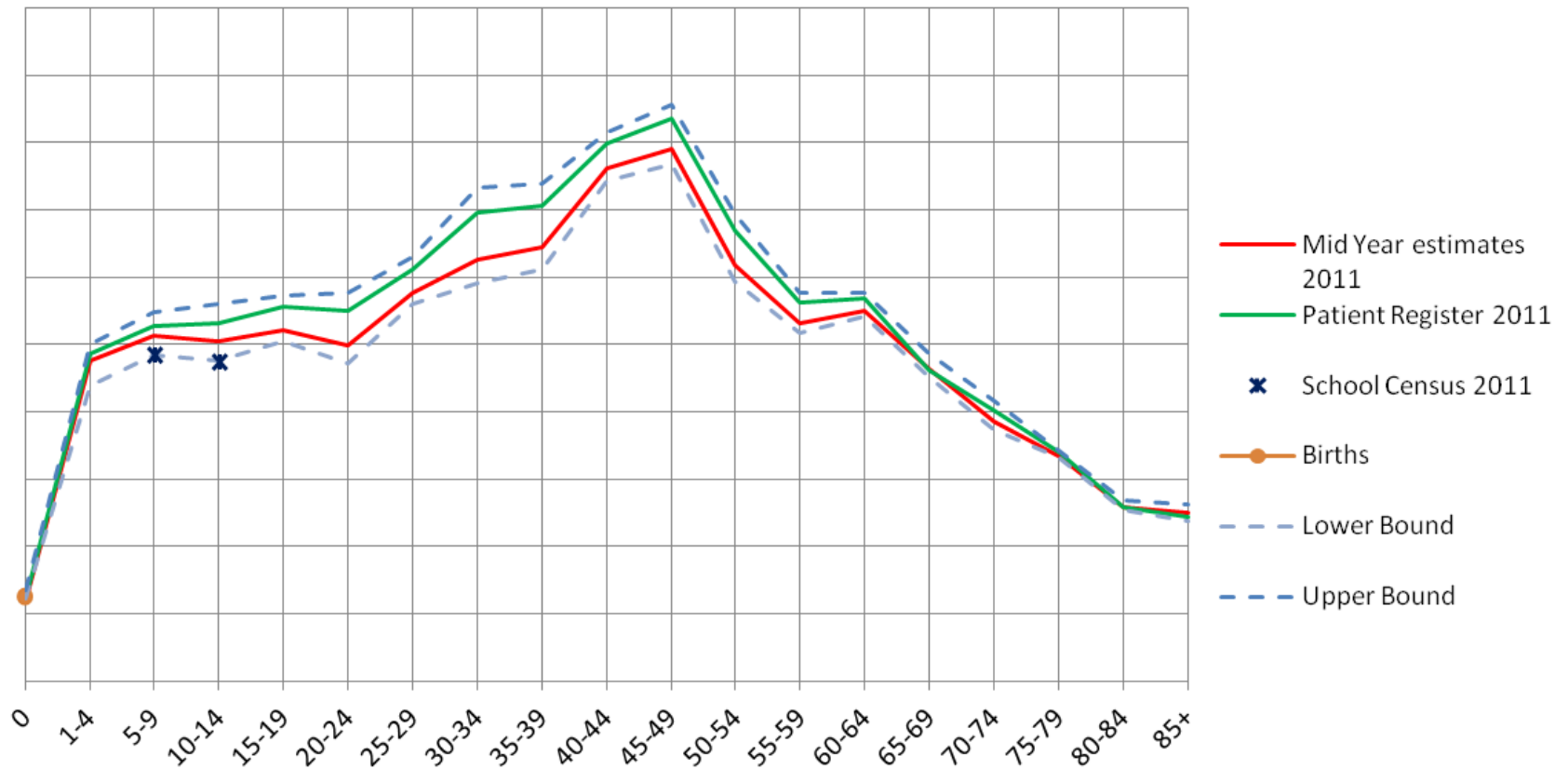
QA Check	Comparator dataset
Students (residential/communal)	Higher Education Statistics Agency (HESA) Further Education Student Numbers from Business, Innovation and Skills
Armed Forces (Home/Foreign)	Defence Analysis Statistics Agency US Armed Forces
Migration (international)	Patient Register ONS International Migration Estimates Migrant Workers Scan

Evidence assessed for all Local Authority estimates

- Tolerance bounds act as a guide for quality assuring estimates
- Two main approaches:
 1. Diagnostic range approach
 - Used when there are two or more comparators
 - Bounds calculated based on variation between sources
 2. Quality assessment approach
 - Used when there is only one comparator source
 - Based on quantifying known quality issues with the comparator

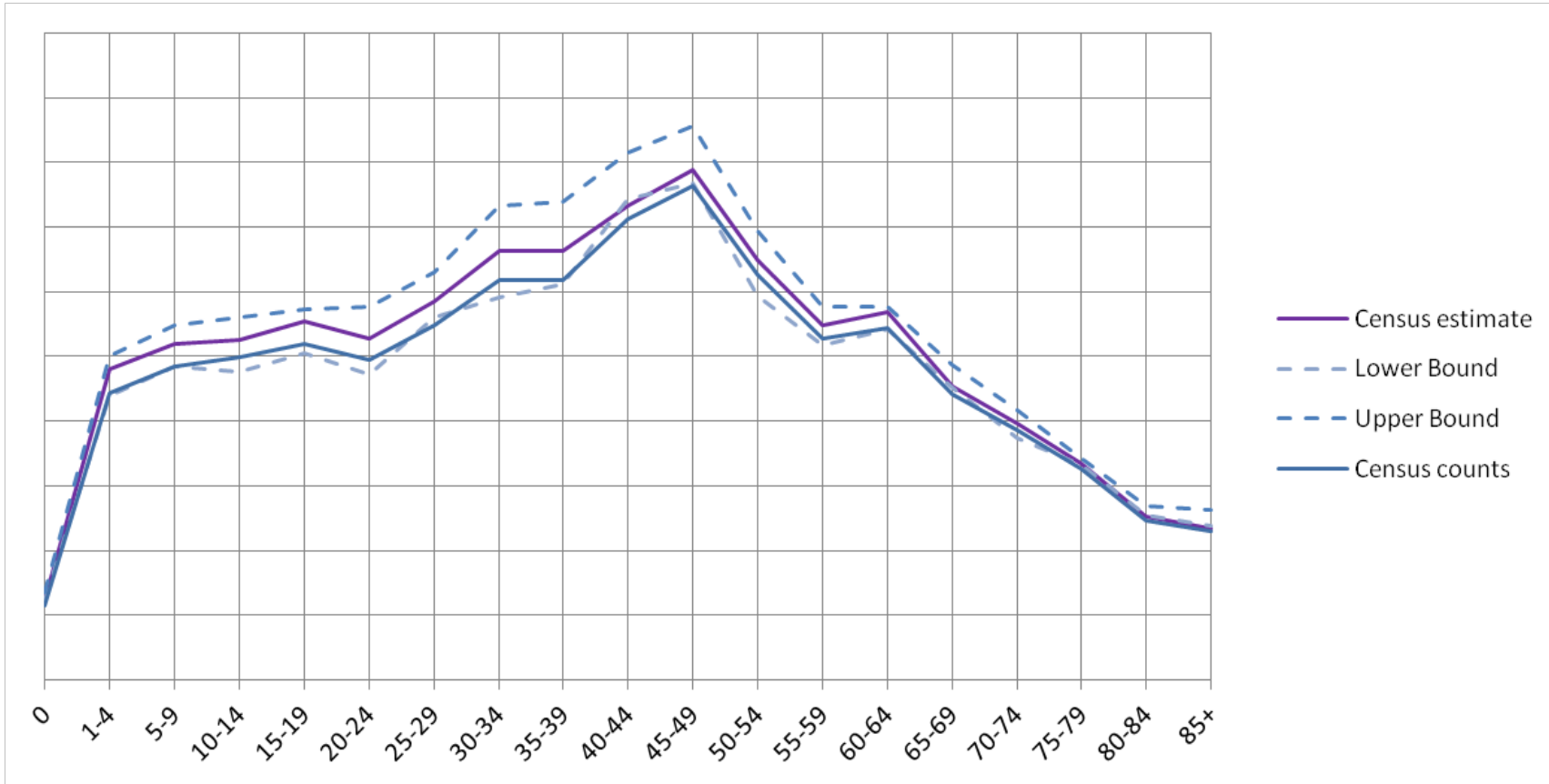
Example of age-sex check

Comparators and Bounds



Example of age-sex check

Census Estimate and Bounds



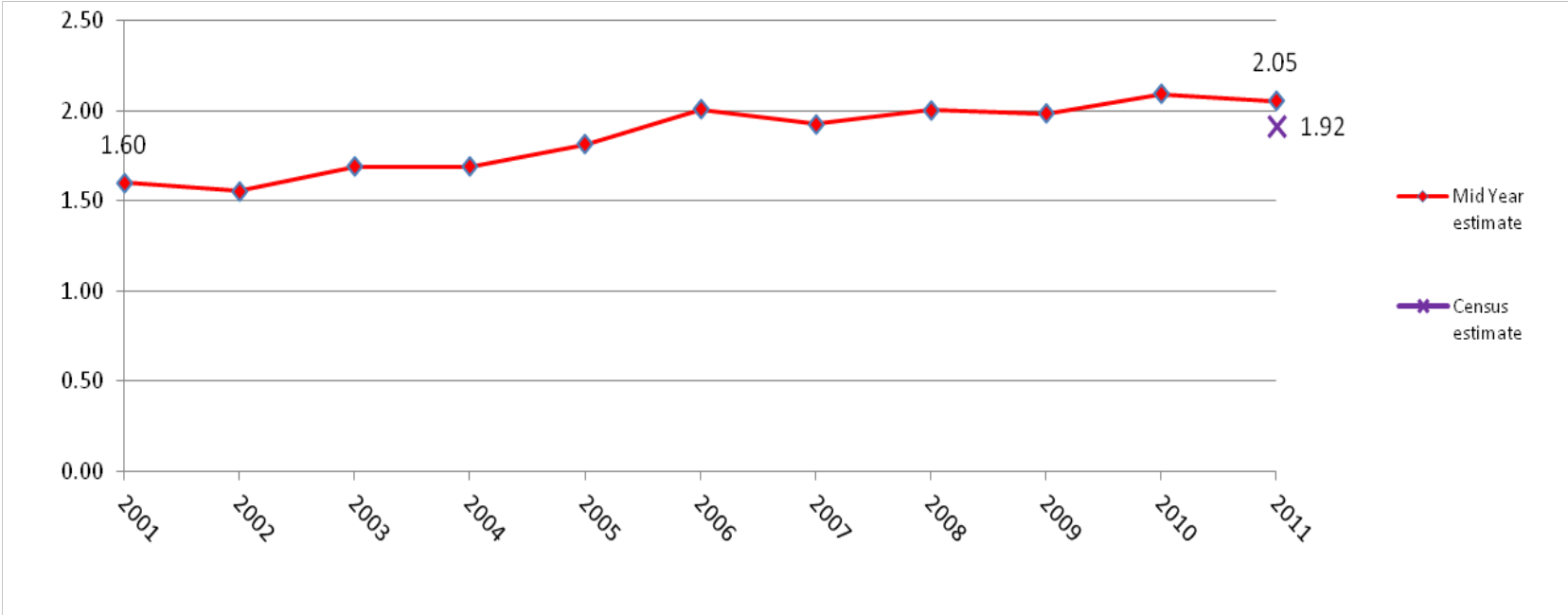
Demographic analysis

Demographic analysis is a key part of the Quality Assurance process:

- Is based on accurate and timely registration data
- Expertise and understanding of fertility and mortality rates
- Assess change over time (based on mid-year estimates) as well as comparison at census day

Examples of Demographic Analysis (Fertility)

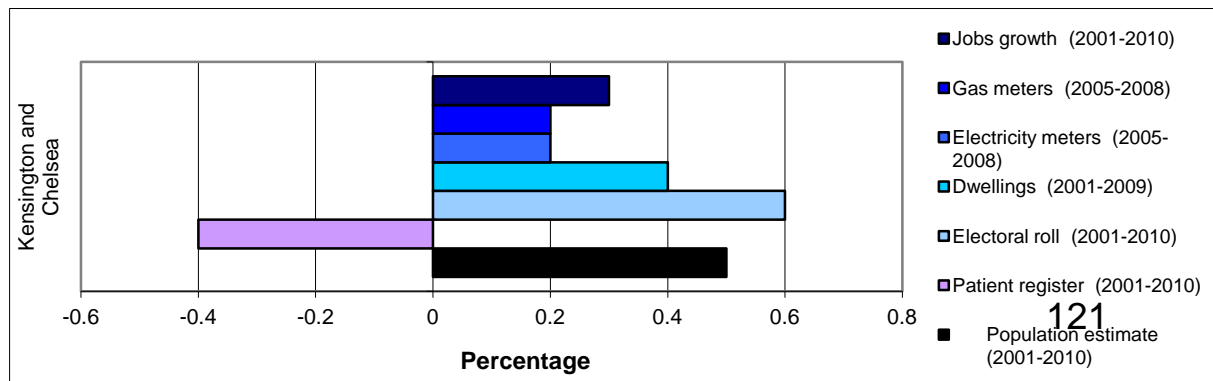
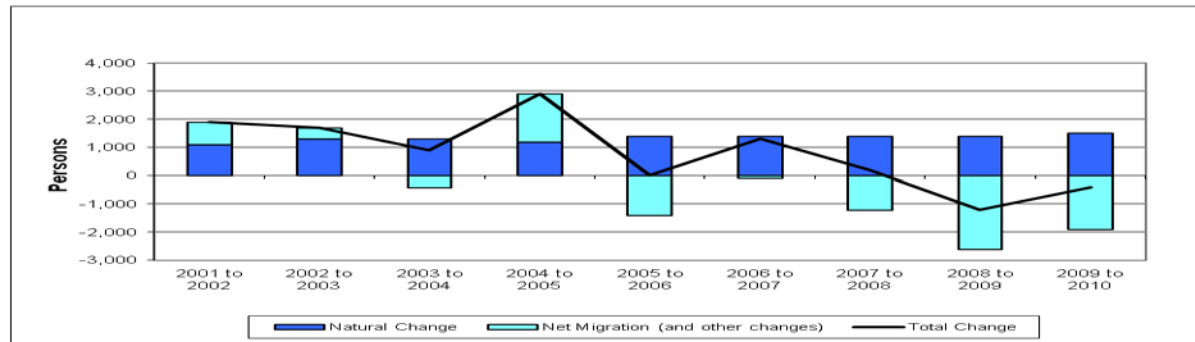
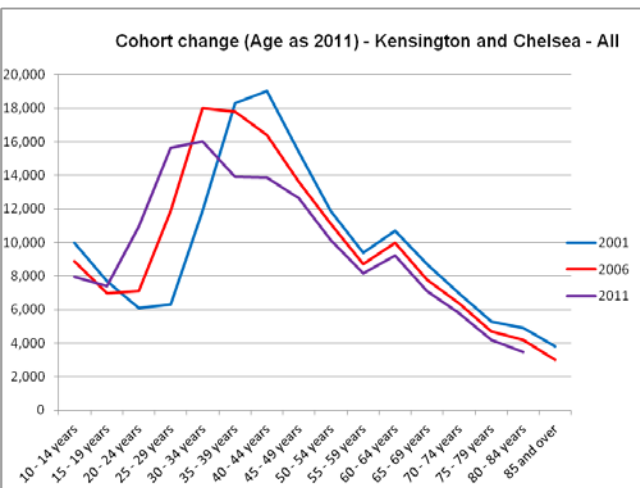
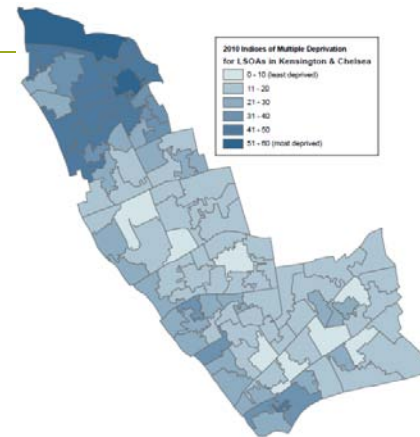
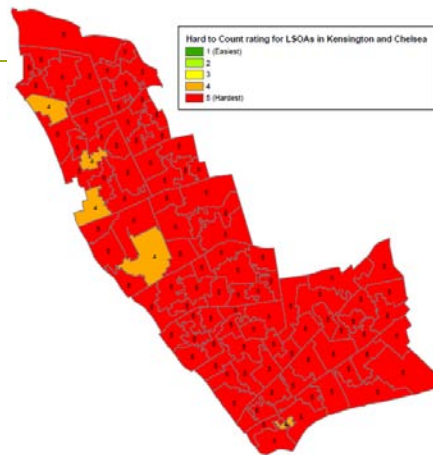
- Fertility rates over past ten years



Profiles of each area

- Mapping of areas (Hard to Count areas, Index of Multiple Deprivation)
- Enumeration challenges (from Census Local Partnership Plans)
- Statistical information on the LA, change over time in:
 - Mid-year estimates (by age-sex)
 - Patient register
 - Gas/electricity meters
 - Dwellings (Council Tax)
 - Electoral Roll
- Information on Communal Establishments – prisons, halls of residence

Example of Area Profile information



Operational Intelligence

- Return rates at Local Authority level
- Return rates within Local Authority
- Information on number of dummy forms (by type)
- Internet / paper responses
- New addresses identified and addresses deactivated
- Census Coverage Survey (CCS) intelligence – interviews completed, addresses listed, refusals

How was evidence considered?

Reviewing evidence and signing-off estimates

- Quality Assurance panels reviewed evidence compiled and analysis carried out
- Approach similar to 2001 but with three panels rather than one
- Important that all Local Authority population estimates pass through the same QA process
- All QA meetings were paperless with all evidence on 20 laptops
 - Security
 - Efficiency
 - Independence
 - Comparability
 - Completeness

Quality Assurance Panels

QA Panel	Membership	Function
QA Steering Group	<ul style="list-style-type: none"> • ONS <i>Experts working on census</i>	<ul style="list-style-type: none"> • Review estimation • Steer on analysis carried out
Main QA Panel	<ul style="list-style-type: none"> • ONS • Welsh Government <i>Census experts and from across ONS</i>	<ul style="list-style-type: none"> • Review all 348 Local Authority estimates • Sign off or request further work
High Level QA Panel	<ul style="list-style-type: none"> • ONS • Academic experts • Expert user • Devolved Administrations <i>Census experts and from across ONS</i> <i>Prof Ludi Simpson (Manchester University)</i> <i>Prof David Martin (Southampton University)</i> <i>Prof Ian Plewis (Manchester University)</i> <i>John Hollis – formerly of GLA</i> <i>Scotland and Northern Ireland</i>	<ul style="list-style-type: none"> • Review emerging regional and national estimates • Sign off or request further work • Review Local Authority estimates as required • Review methodological change • Quality assure process

QA Steering Group

- Aim was to assess the 5 year age-sex estimates after the coverage estimation stage
- Provided a steer on additional analysis to carry out
- Met approximately 50 times
- Focus on age-sex estimates
- Requested further work be carried out and adjustments considered:
 - Estimation processing e.g. collapsing
 - Earlier processing stages
 - Mid-year estimates and comparator data

Main QA Panel

- Aim was to assess all 348 Local Authority estimates
- Met a total of 31 times
- Routinely considered all checks and evidence for all areas
- Requested further work be carried out and adjustments considered:
 - Further investigations into mid-year estimates, comparator data as well as census estimates
 - Local Authority estimates reviewed multiple times in some cases
- Recommendation made to National Statistician to sign off Local Authority estimates

High Level QA Panel

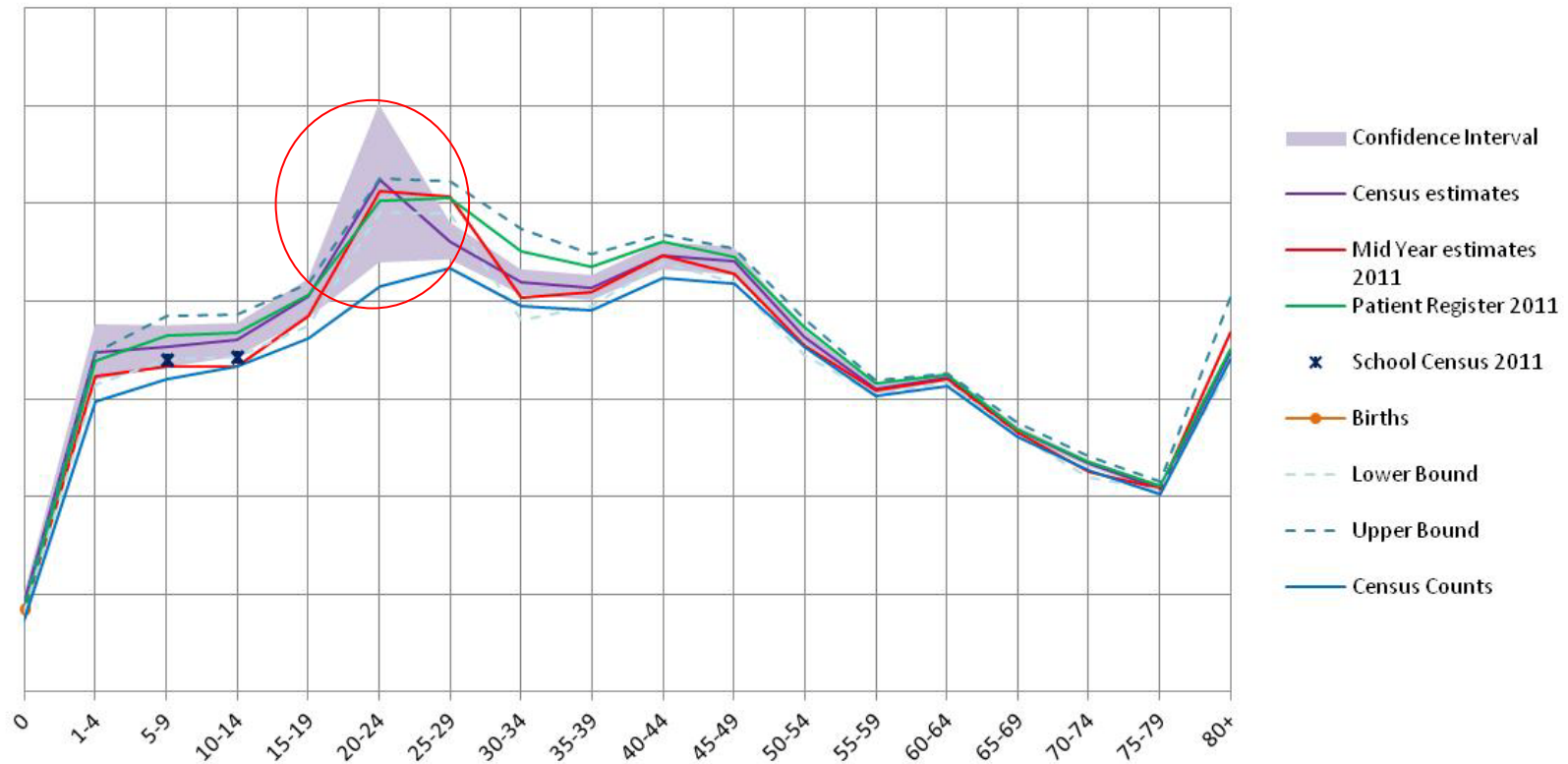
- Aim was to assess regional/national estimates and the QA process as a whole
- Met a total of 12 times
- Also considered the need for and suitability of adjustments
- Recommendation made to the National Statistician to sign off census estimates

Demonstrating QA through practical examples

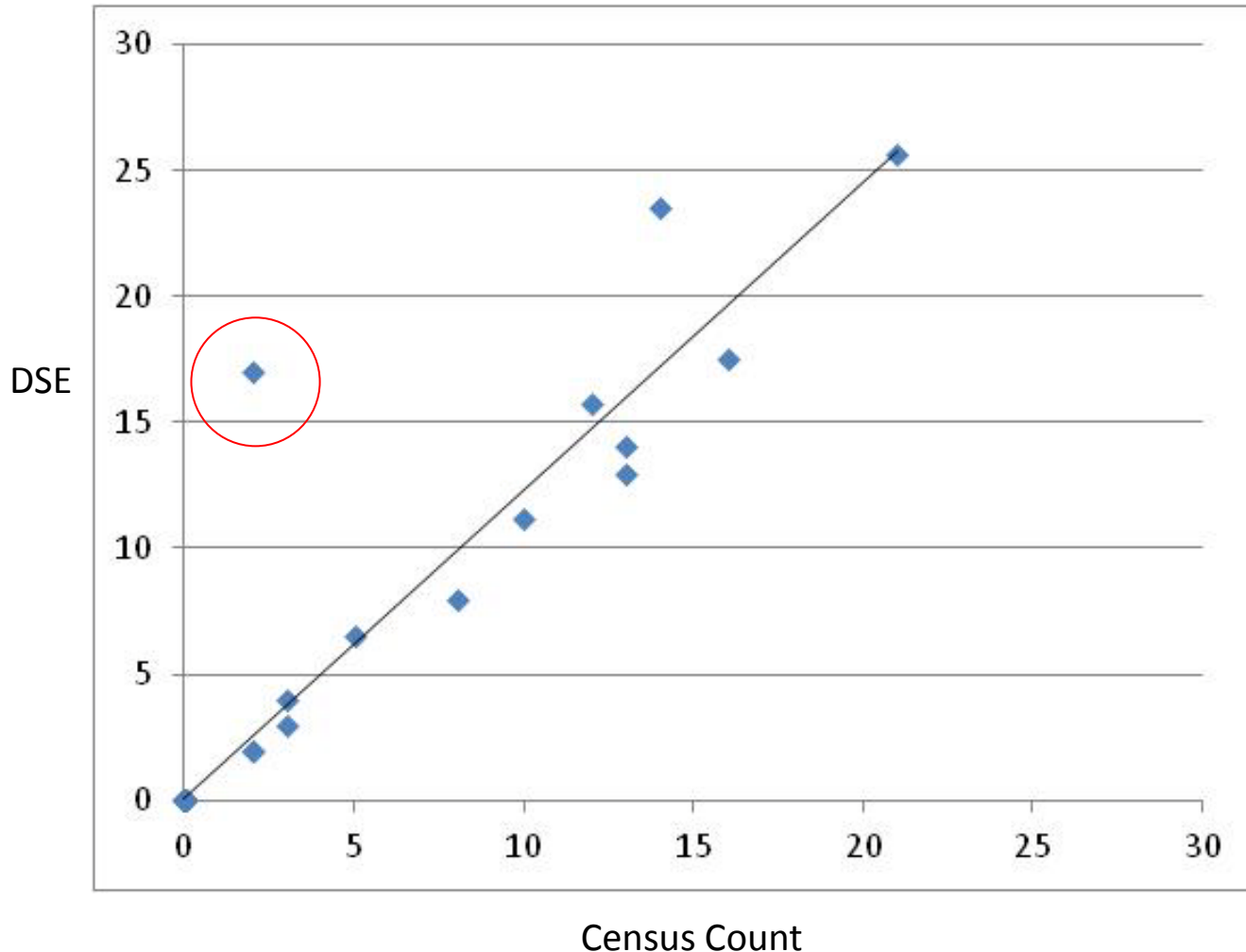
QA of estimation

- Five-year age-sex estimates initially checked for all 348 Local Authorities
- Assessed using comparator data and pre-defined tolerances
- Two typical examples presented:
 1. Width of confidence intervals
 2. Inconsistencies
- In both cases adjustments made to the initial estimates seen

Example 1 – Width of Confidence Intervals



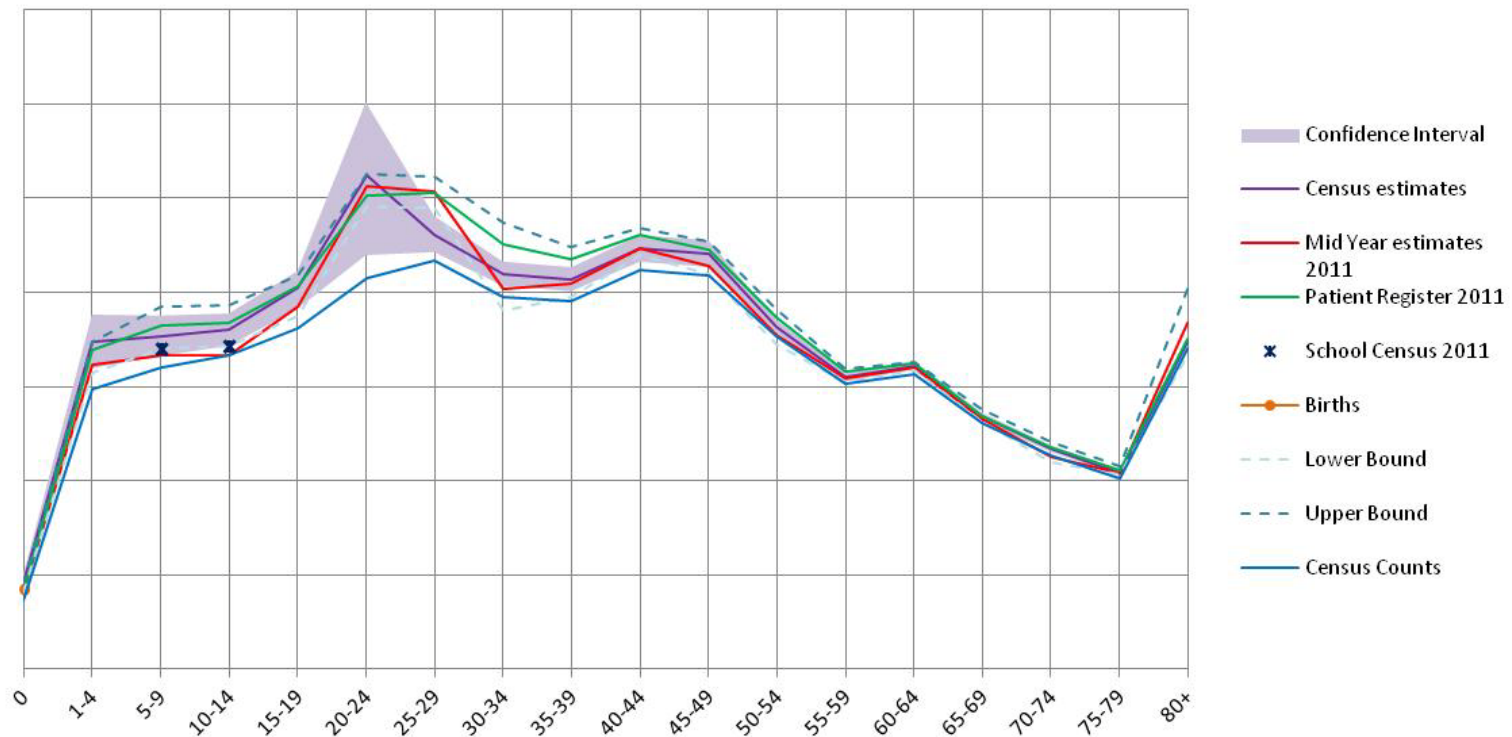
Example 1 – Width of Confidence Intervals



- No evidence of error found in matching
- Instead collapse age groups 18-19 with 20-24
- Reduced the influence of the outlier

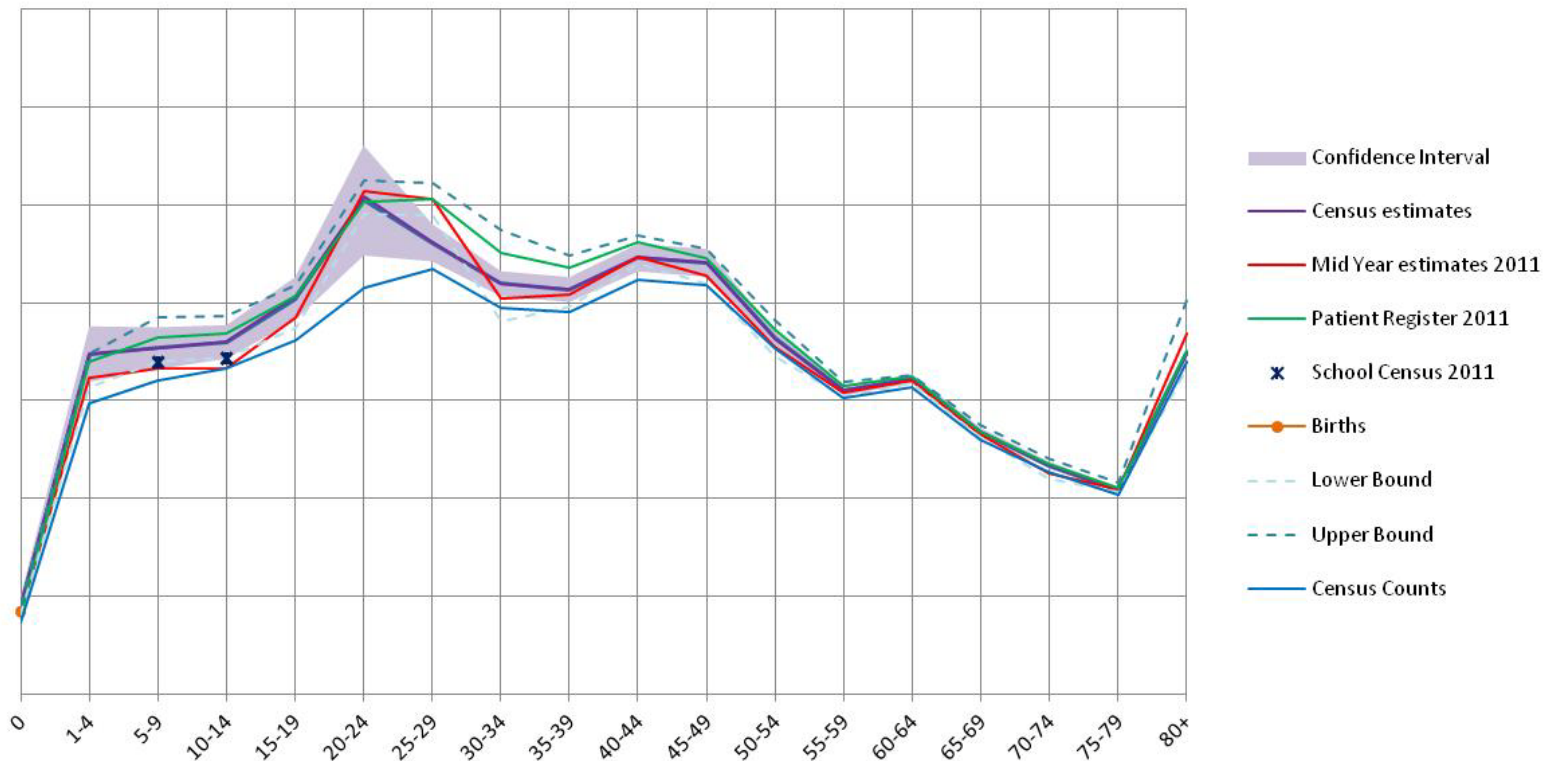
Example 1 – Width of Confidence Intervals

- Before collapsing

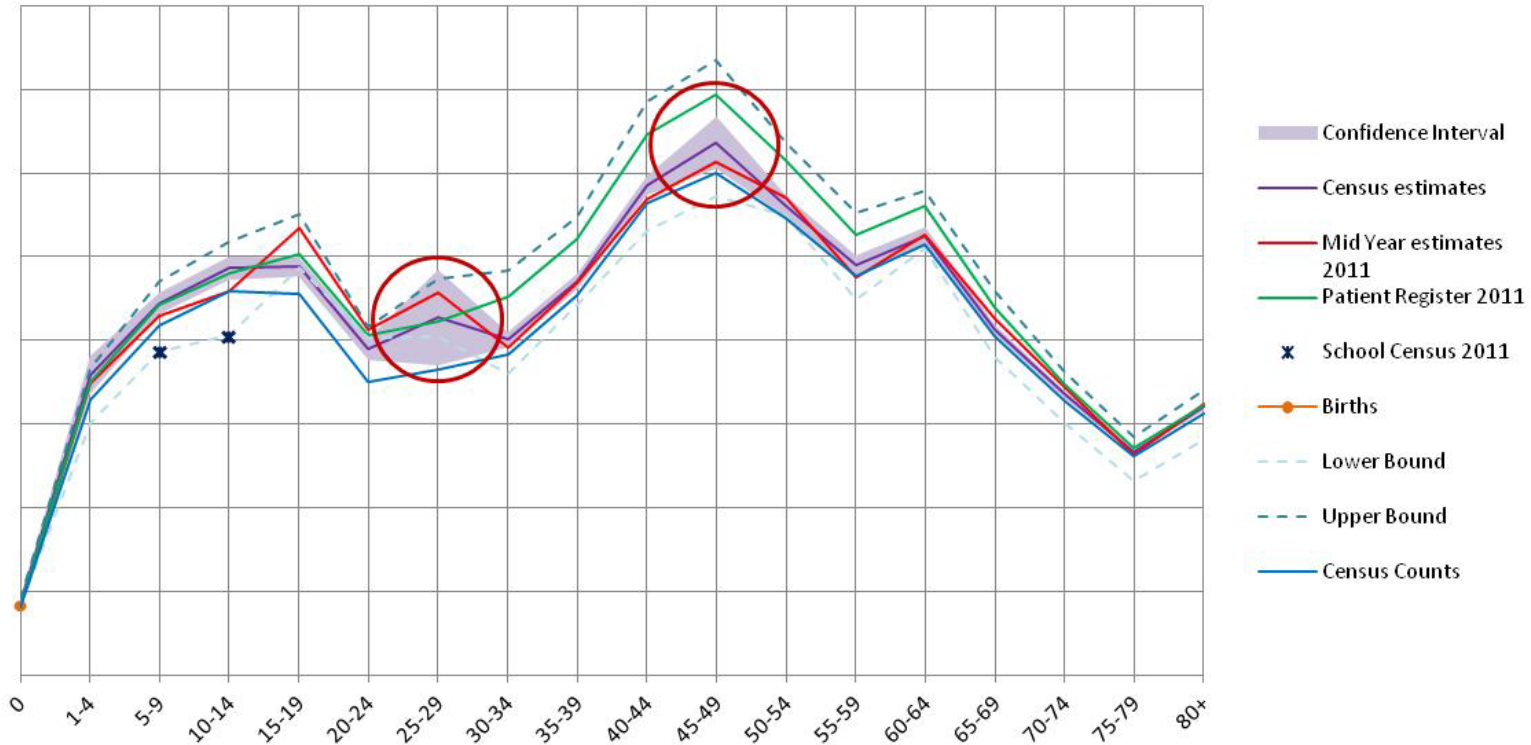


Example 1 – Width of Confidence Intervals

- After collapsing



Example 2 – Inconsistencies



Example 2 – Inconsistencies

Ages 25-29 and 45-49 assessed further:

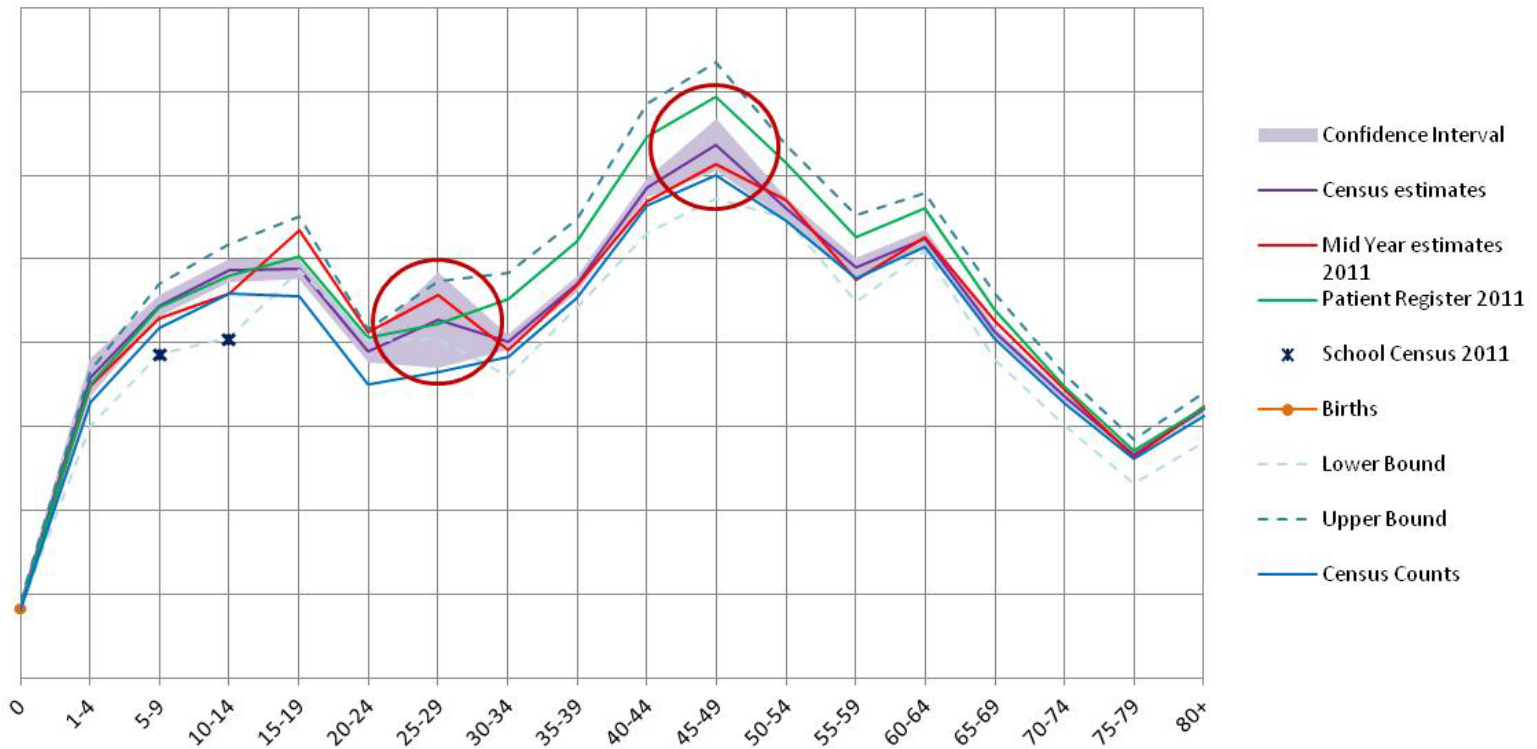
- At age 25-29 and 45-49 estimation is greater than in neighbouring age groups
- Different shape to comparator data
- Confidence intervals also wider at these ages
- Investigate potential outliers – not found

Adjustment: Collapse ages 40-49

 Collapse ages 19-29

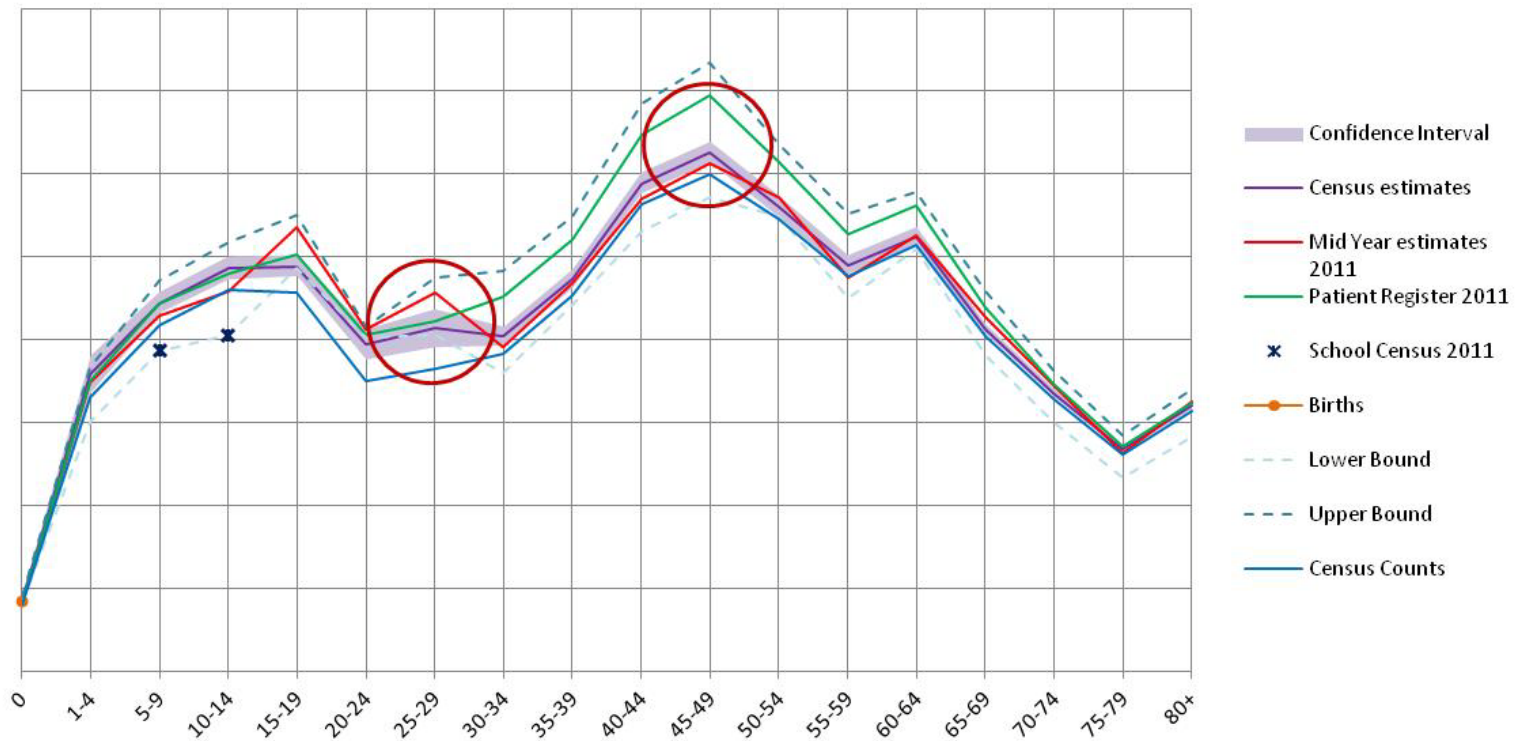
Example 2 – Inconsistencies

- Before collapsing



Example 2 – Inconsistencies

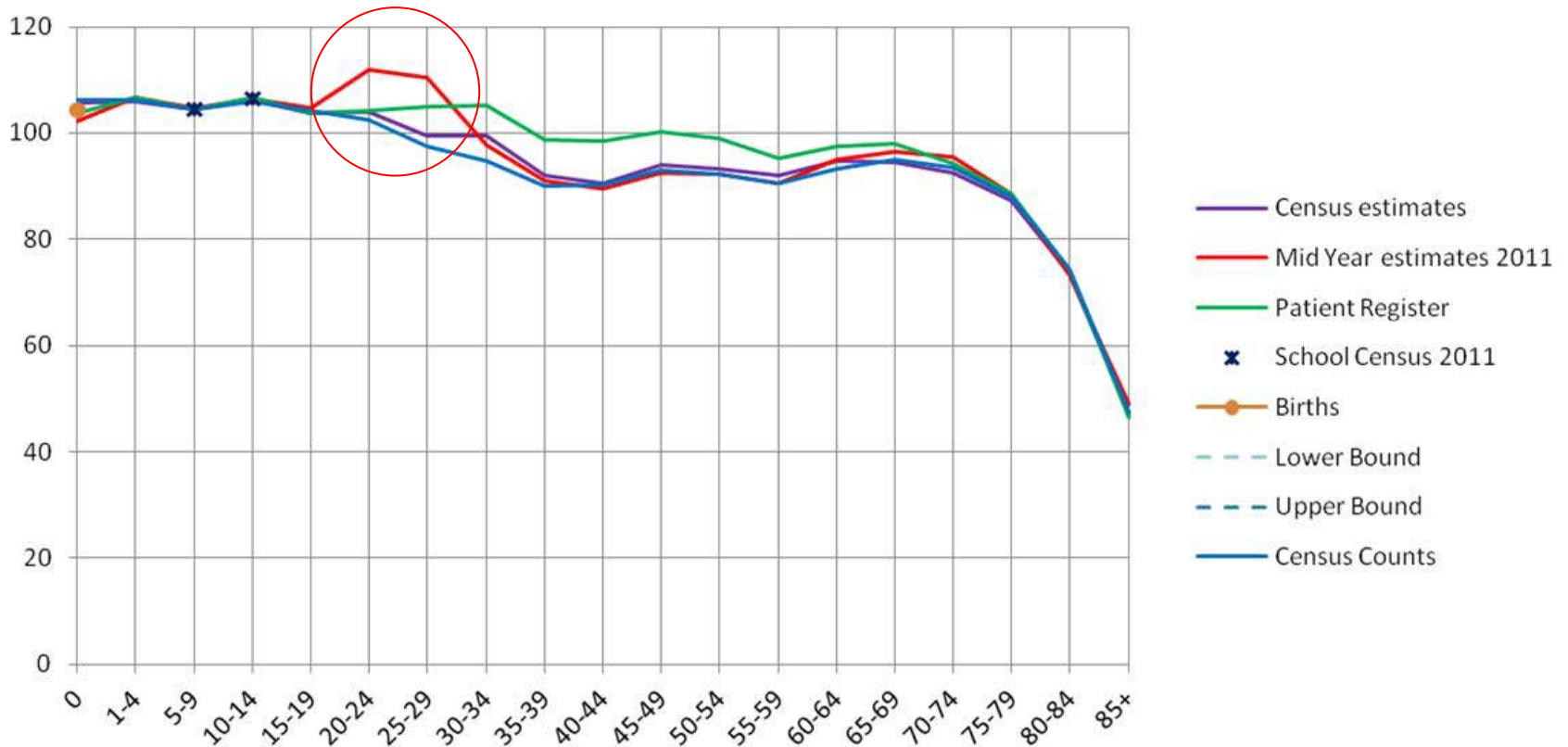
- After collapsing



QA of Final Estimates

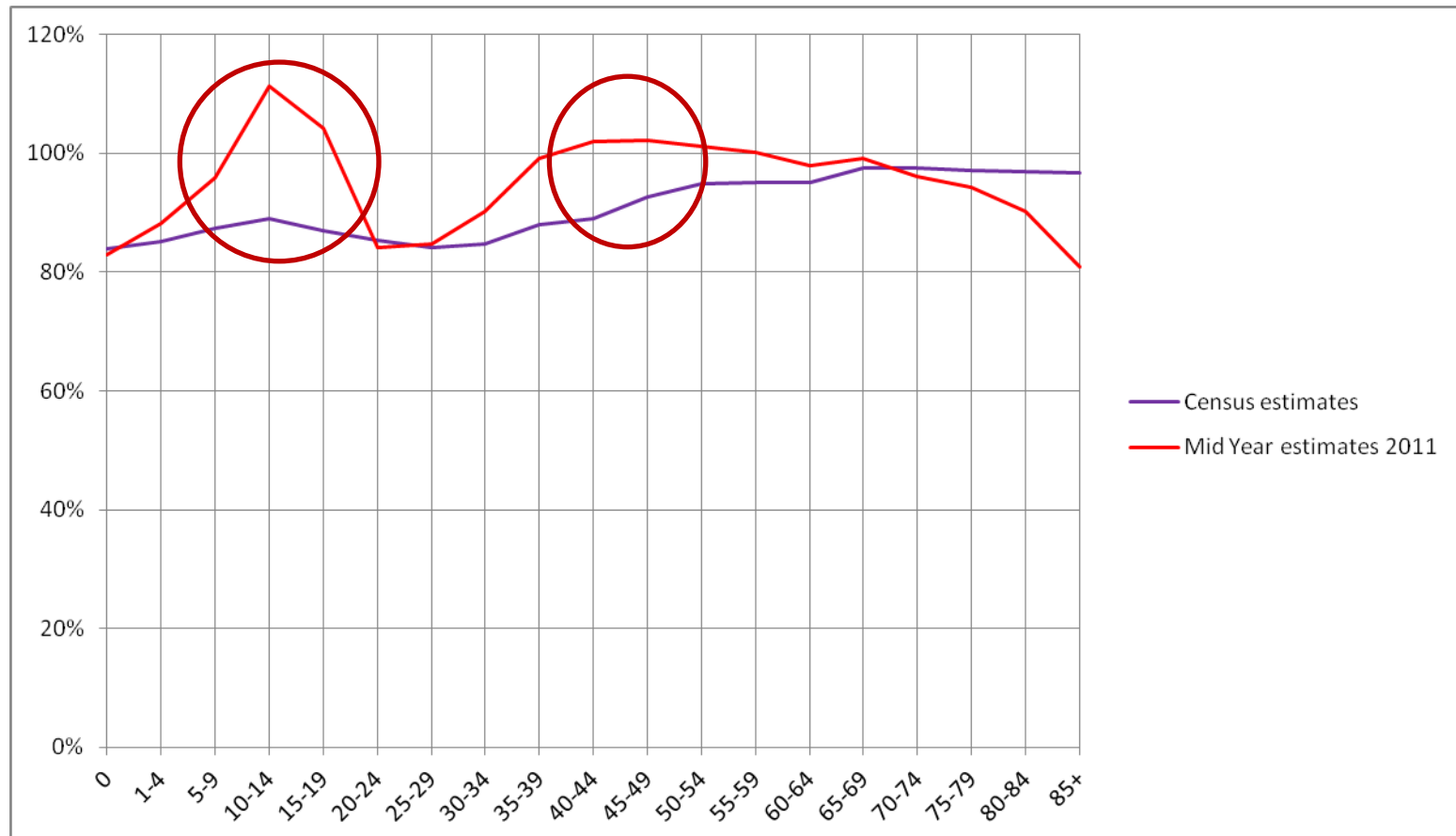
- Full range of QA checks assessed for all 348 Local Authorities
- Four typical examples presented:
 1. Inconsistency with population comparator data (by age)
 2. Inconsistency within a Local Authority (population)
 3. Inconsistency within a Local Authority (households)
 4. Consistency with ethnicity comparator data
- Examples are based on actual census data but are anonymised given pre-release access

Sex ratio analysis



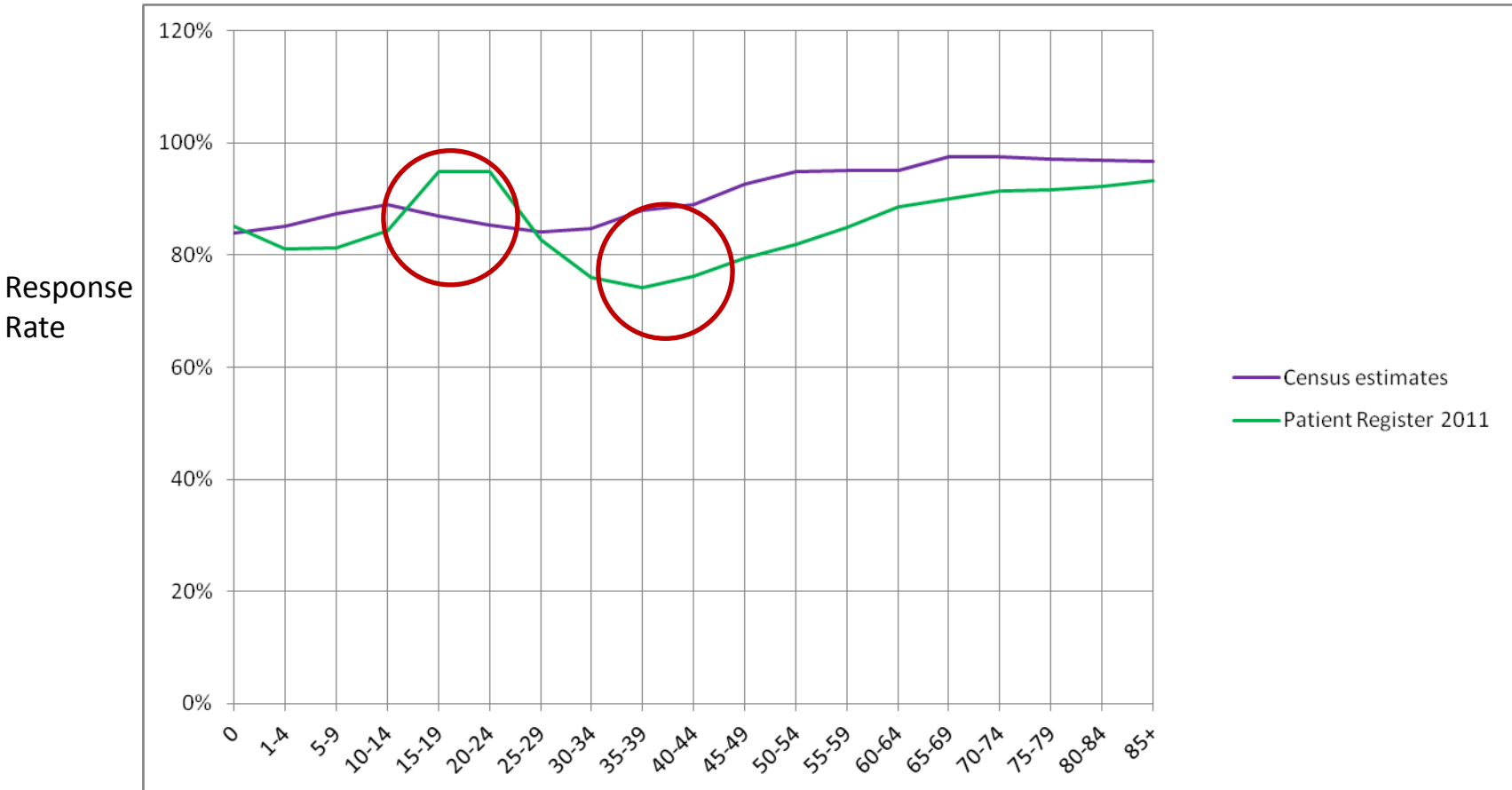
1. Implied Response Rate (Mid-Year Estimates)

- Implied Response Rates = $\text{Census Count} / \text{Comparator Source}$



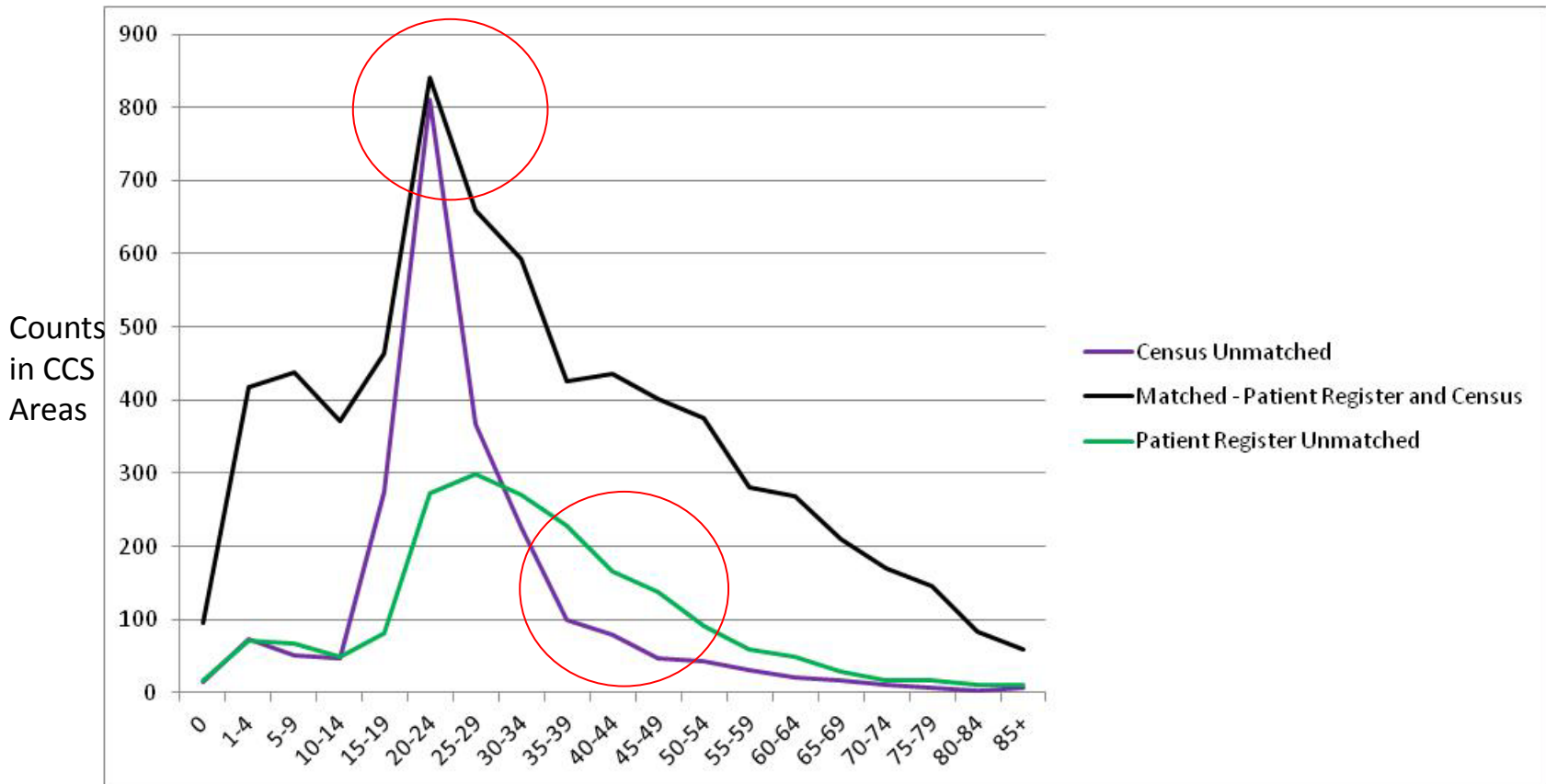
1. Implied Response Rate (Patient Register)

- Implied Response Rates = $\text{Census Count} / \text{Comparator Source}$



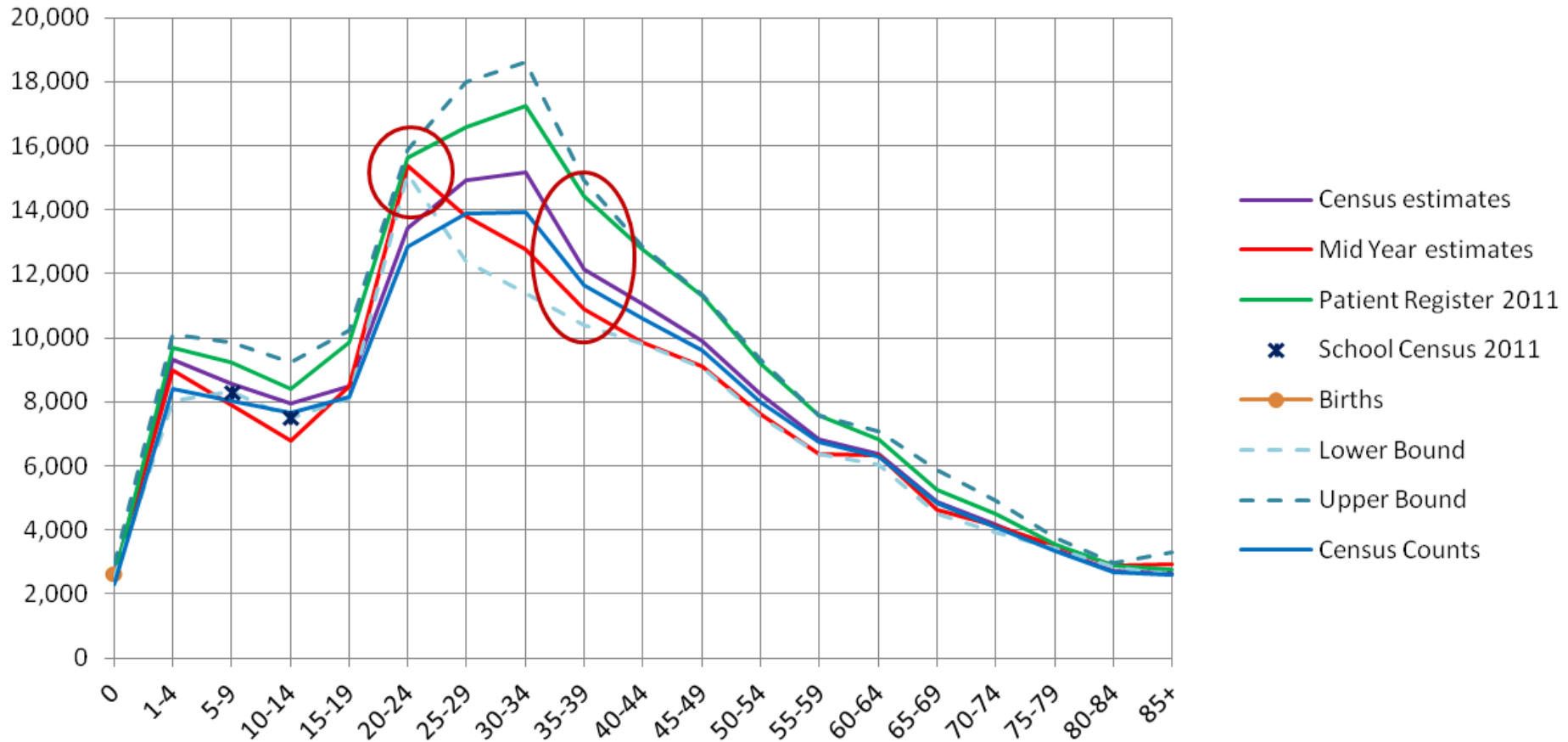
Findings from Data Matching

- Matching Census/CCS to Patient Register

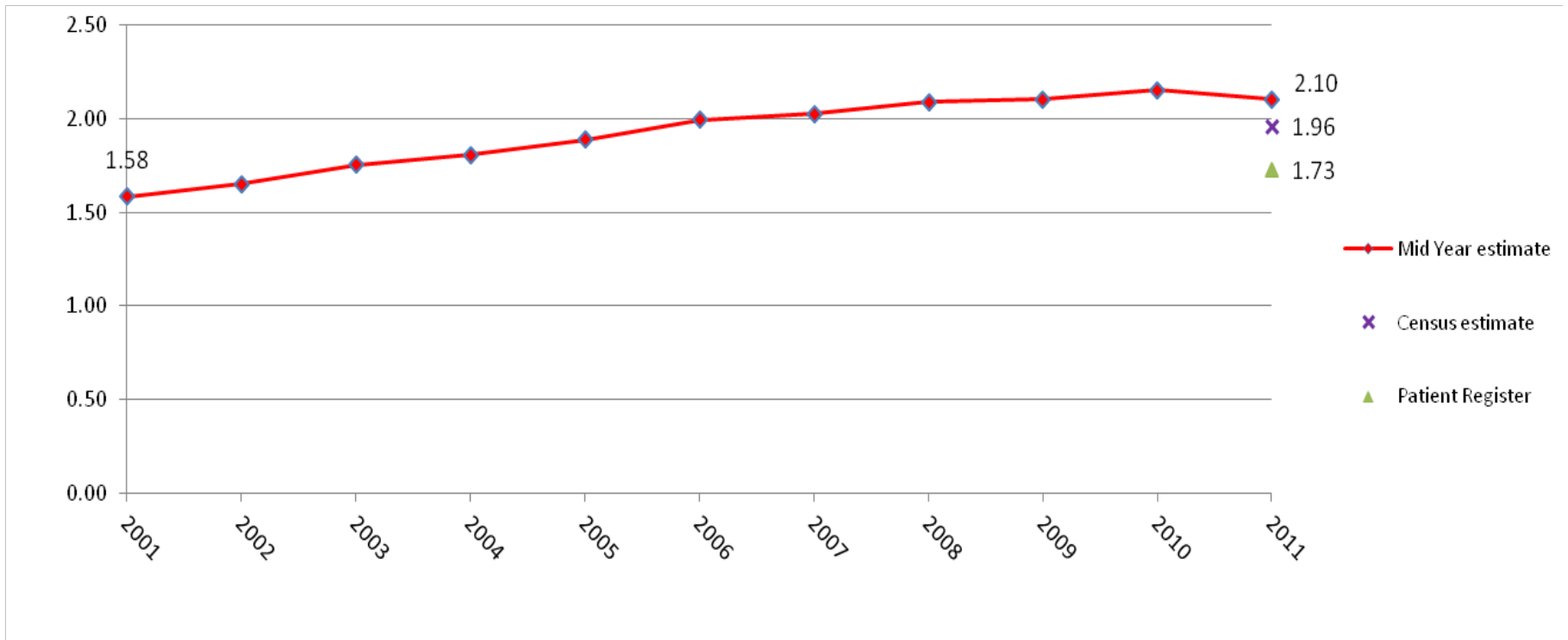


1. Shape of Bounds Across Ages

- Shape of bounds and consistency of comparators across ages

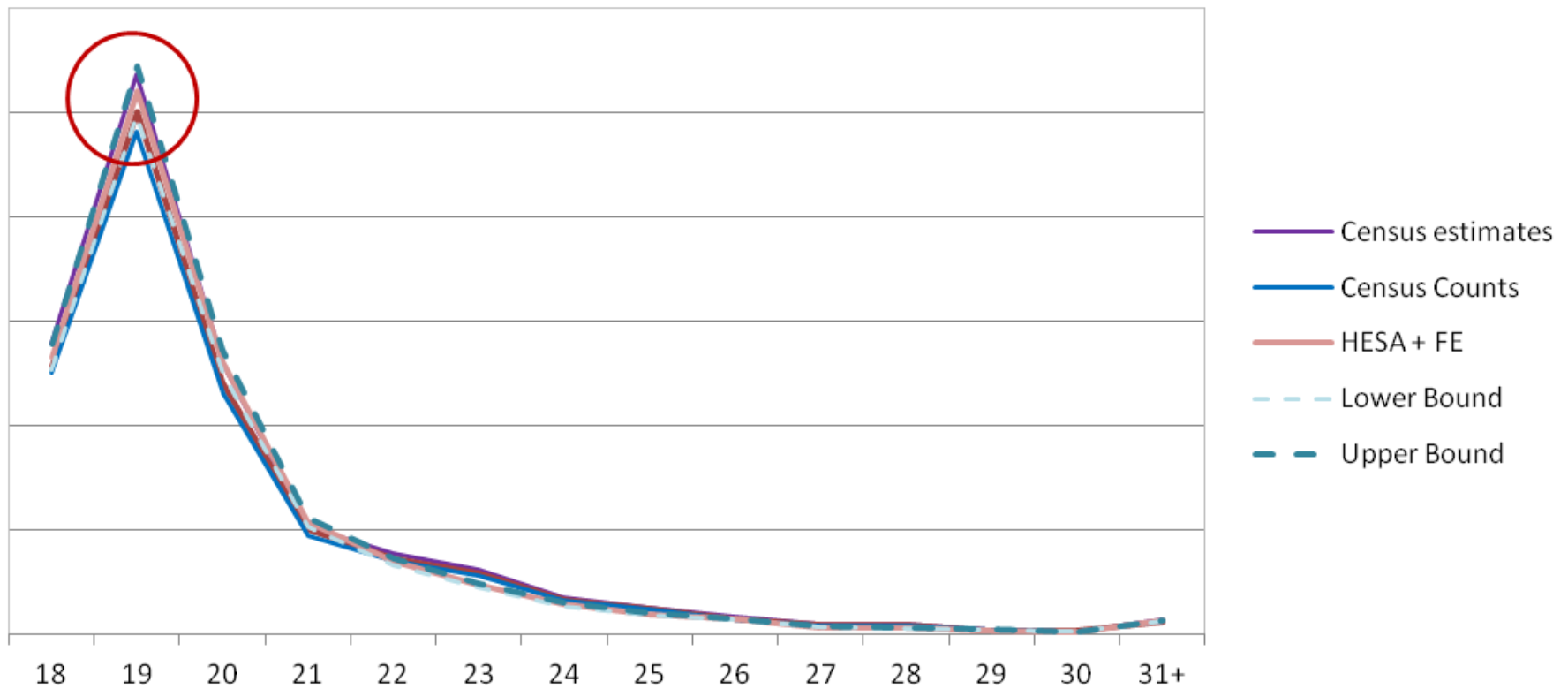


Fertility analysis over time



1. Students in communals

- Students in communals establishments against Higher Education Statistics Agency and Further Education data

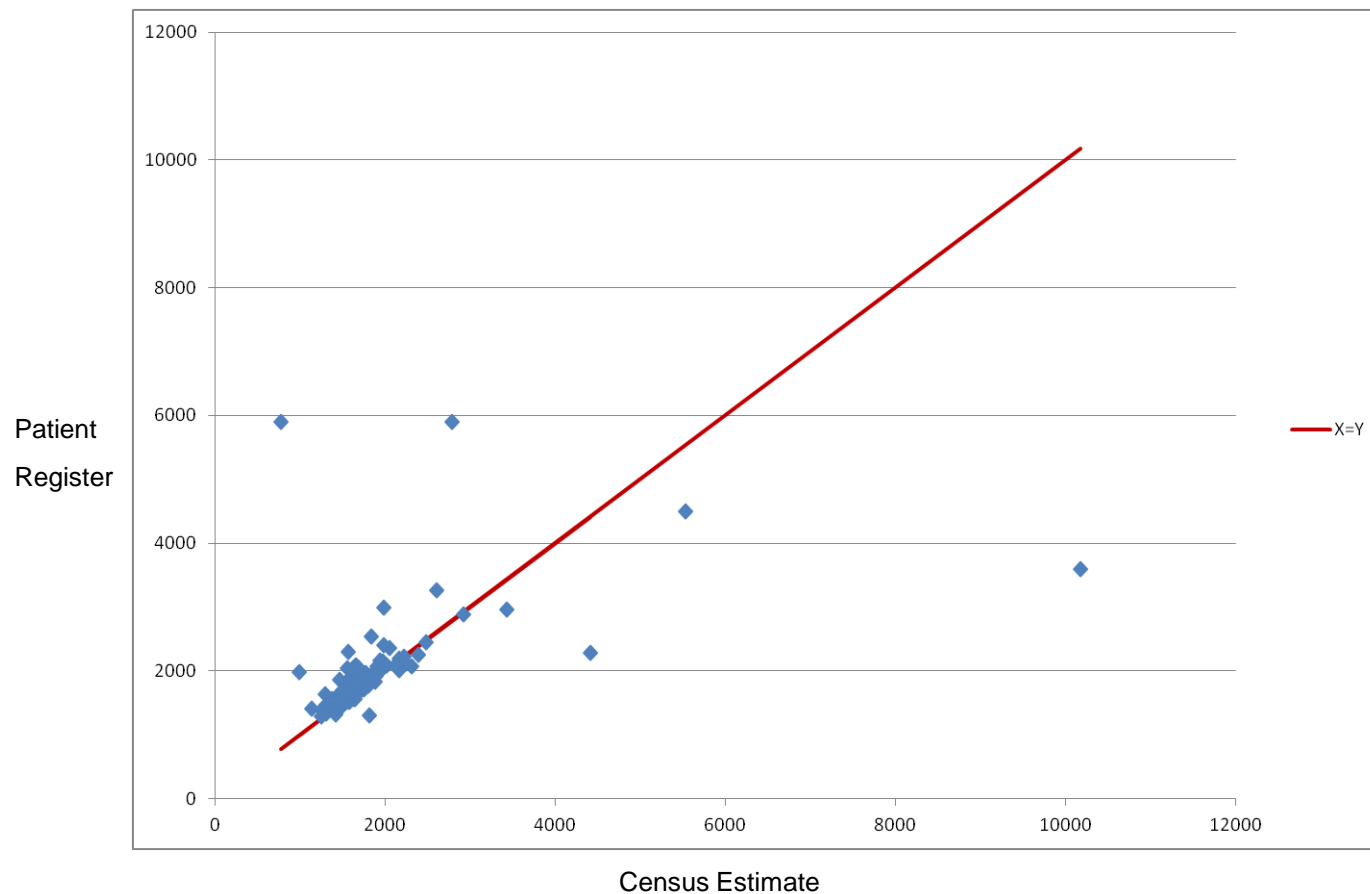


2. Inconsistency within a Local Authority (persons)

- Carried out to identify potential pockets of problems
- Interpreted with caution as coverage adjustment is aimed at producing LA level estimates
- Comparisons made against Patient Register at LSOA level
- Inconsistencies found attributable to:
 - Large Communal Establishments in the wrong LSOA or LA in Census data
 - Issue with Patient Register

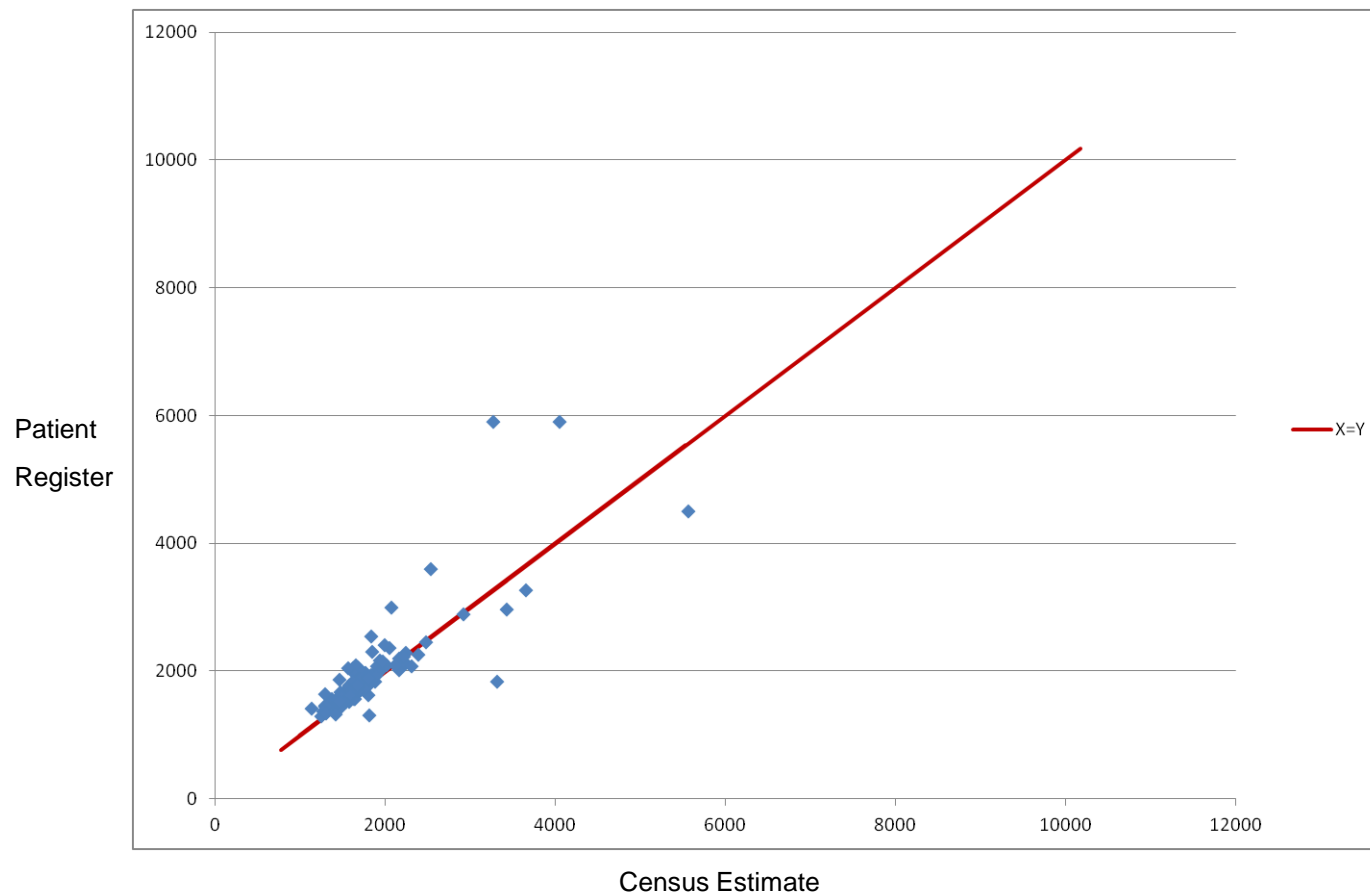
2. Inconsistency within a Local Authority

- Patient Register against Census Estimate
- Identification of Communal Establishments in wrong area (**before**)



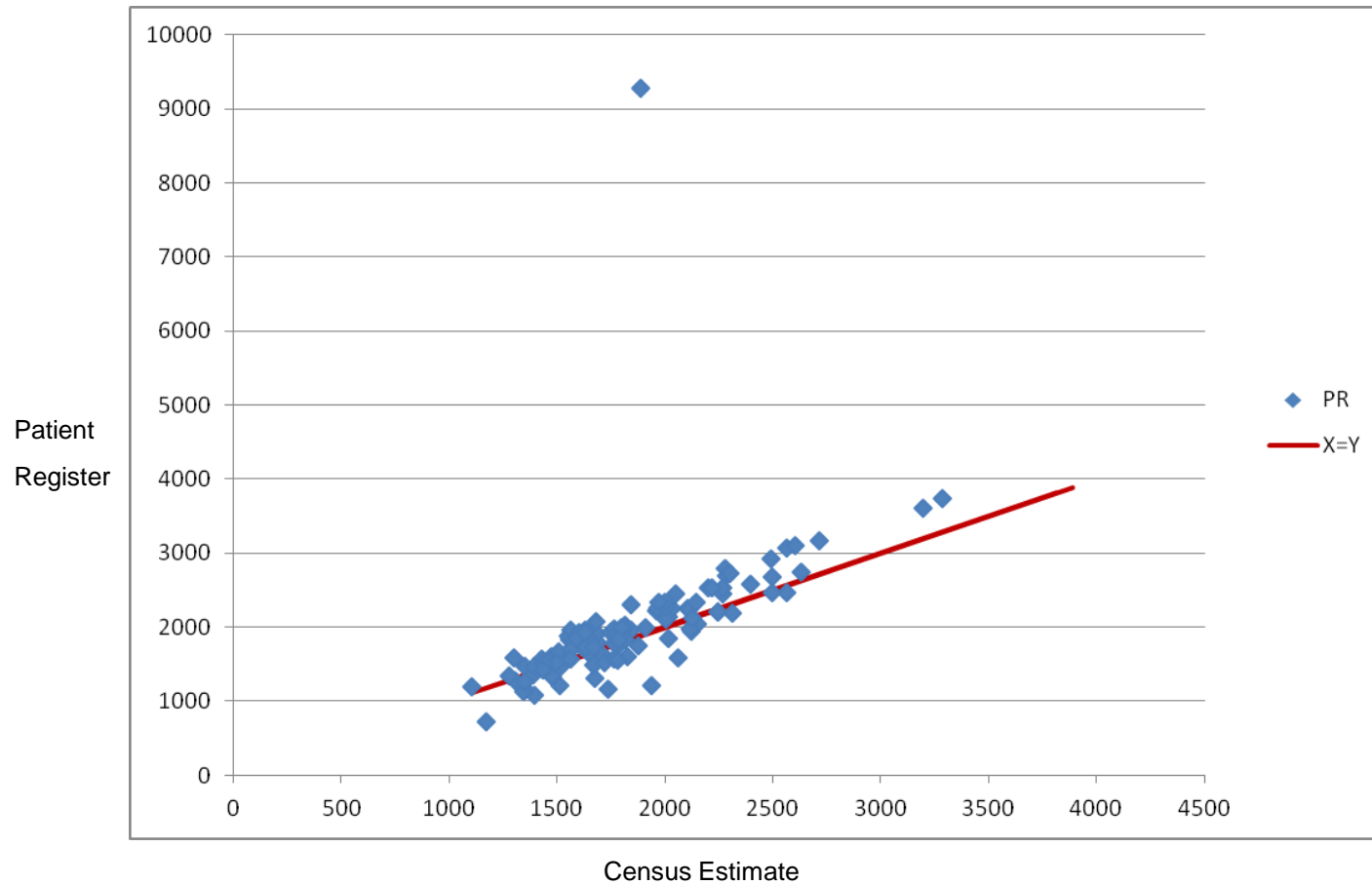
2. Inconsistency within a Local Authority

- Patient Register against Census Estimate
- Identification of Communal Establishments in wrong area (**after**)



2. Inconsistency within a Local Authority

- Patient Register against Census Estimate
- Patient Register outlier – University health centre

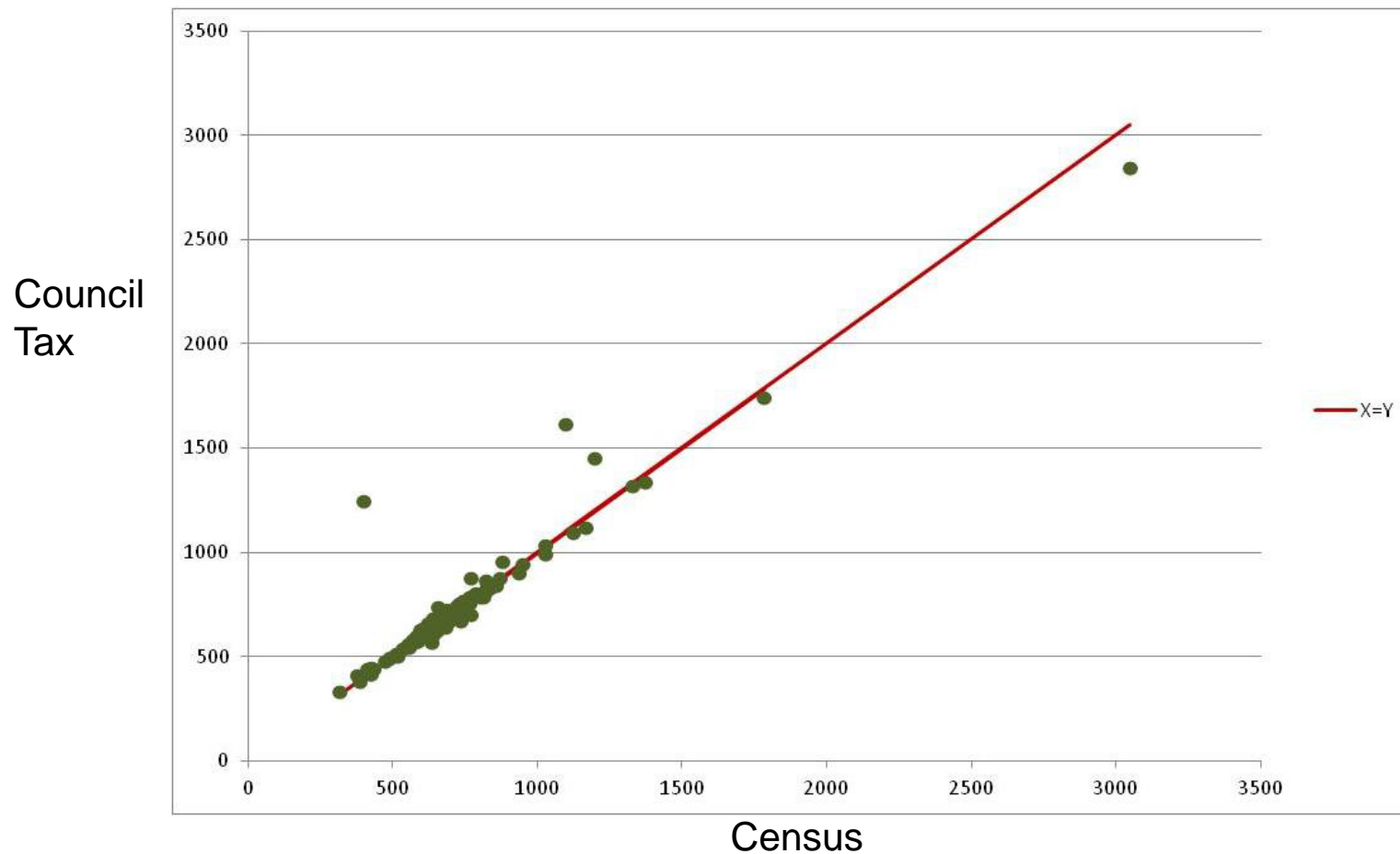


3. Inconsistency within a Local Authority (households)

- Carried out to identify potential pockets of problems
- Interpreted with caution as coverage adjustment is aimed at producing LA level estimates
- Comparisons made against:
 - Patient Register data (grouped into households)
 - Local Authority Council Tax (occupied dwellings using discounts/exemptions)
- Inconsistencies attributed to:
 - Council Tax (quality, student halls, unbanded addresses)
 - Short-term residents

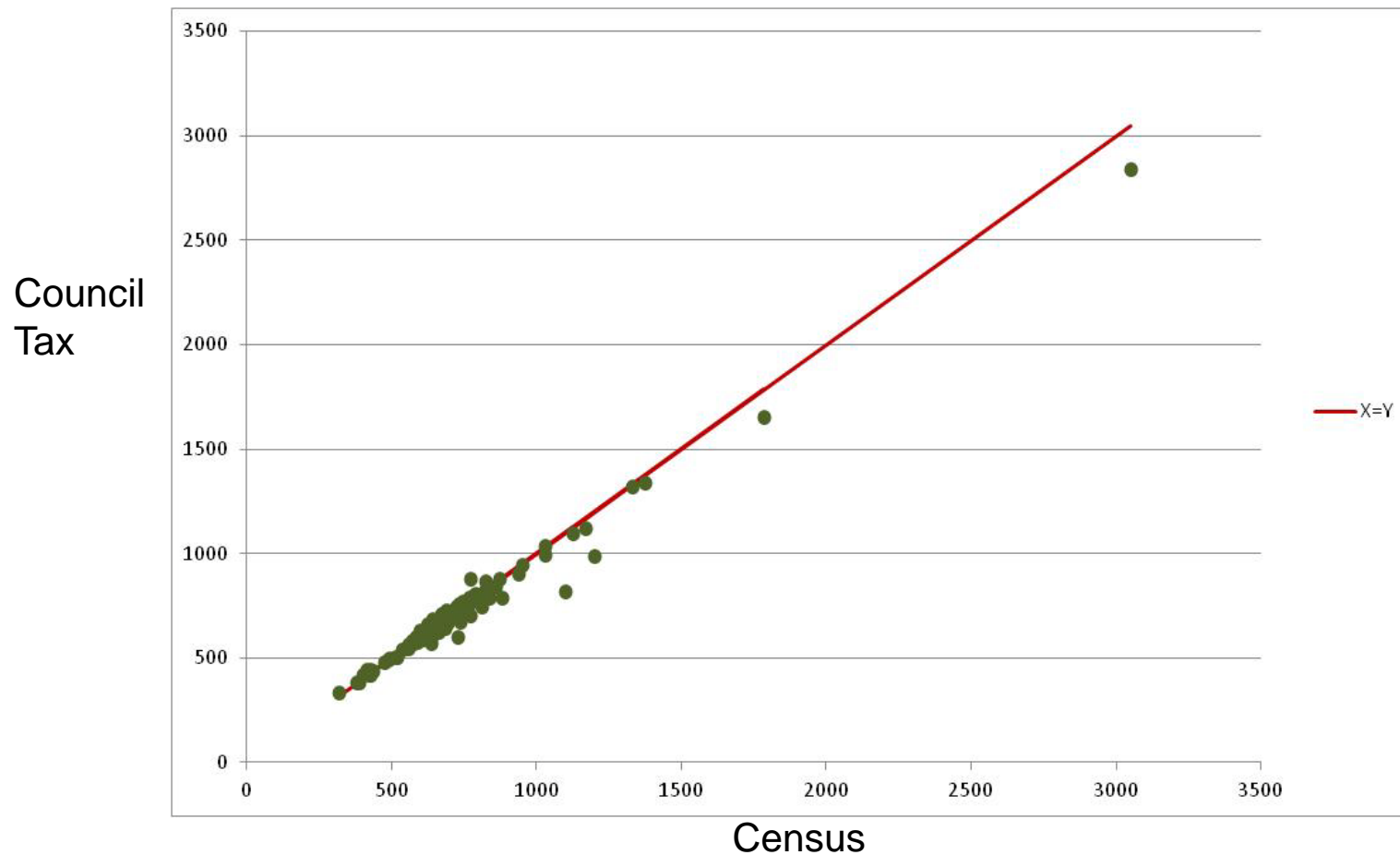
3. Inconsistency within a Local Authority (households)

- Council Tax (occupied) against Census Estimate
- Council Tax Class M (student hall) **included**



3. Inconsistency within a Local Authority (households)

- Council Tax (occupied) against Census Estimate
- Council Tax Class M (student hall) **excluded**

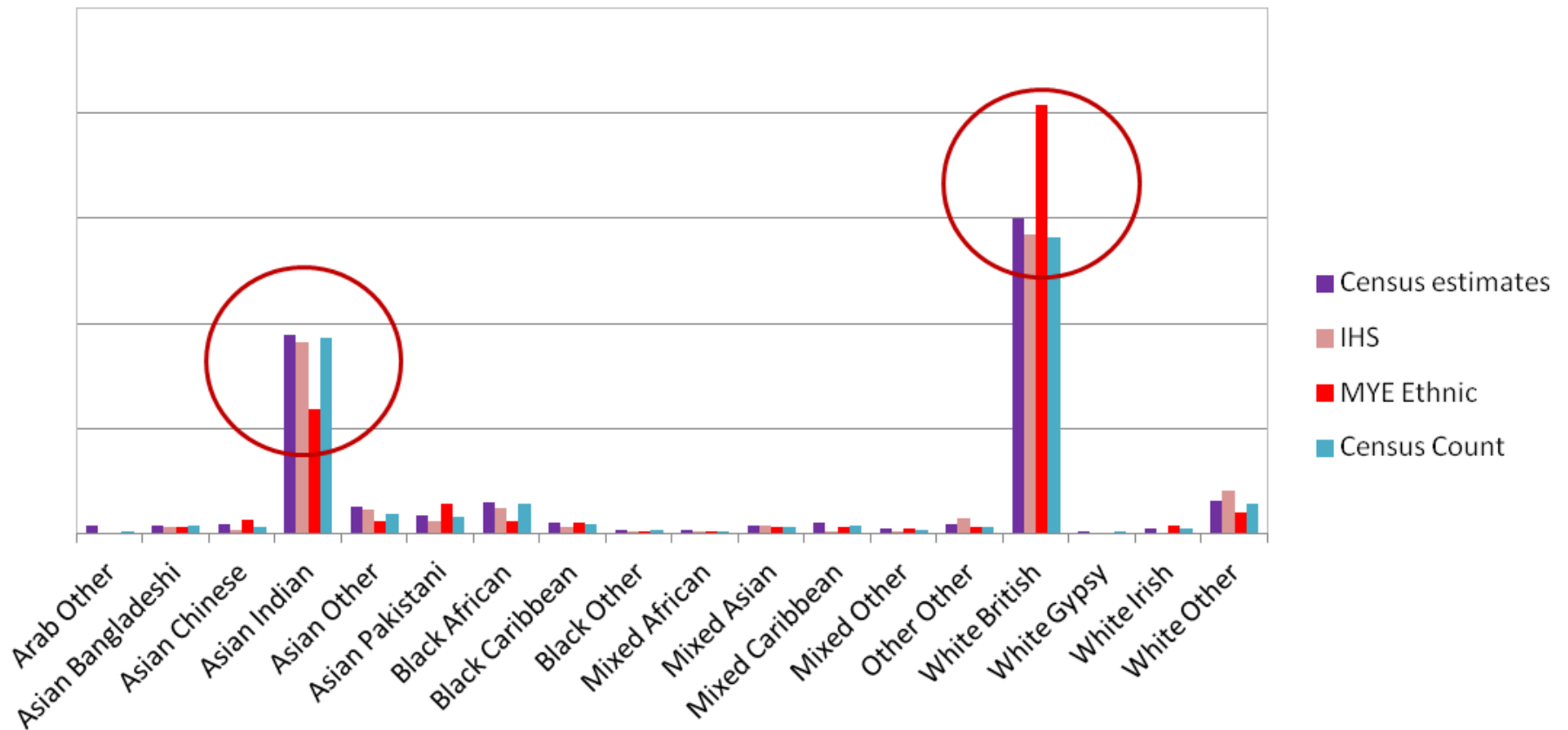


4. Consistency with Ethnicity Comparator

- Comparisons gave us particular confidence in the census estimates
- Cautious about the use of the check given potential quality issues of comparator data:
 - Integrated Household Survey (IHS) - Sample survey
 - Mid-Year Estimates by Ethnic Group - Based on 2001 Census ethnicity
 - School Census - Recorded by third party
- Compared well to comparators – particularly School Census estimates

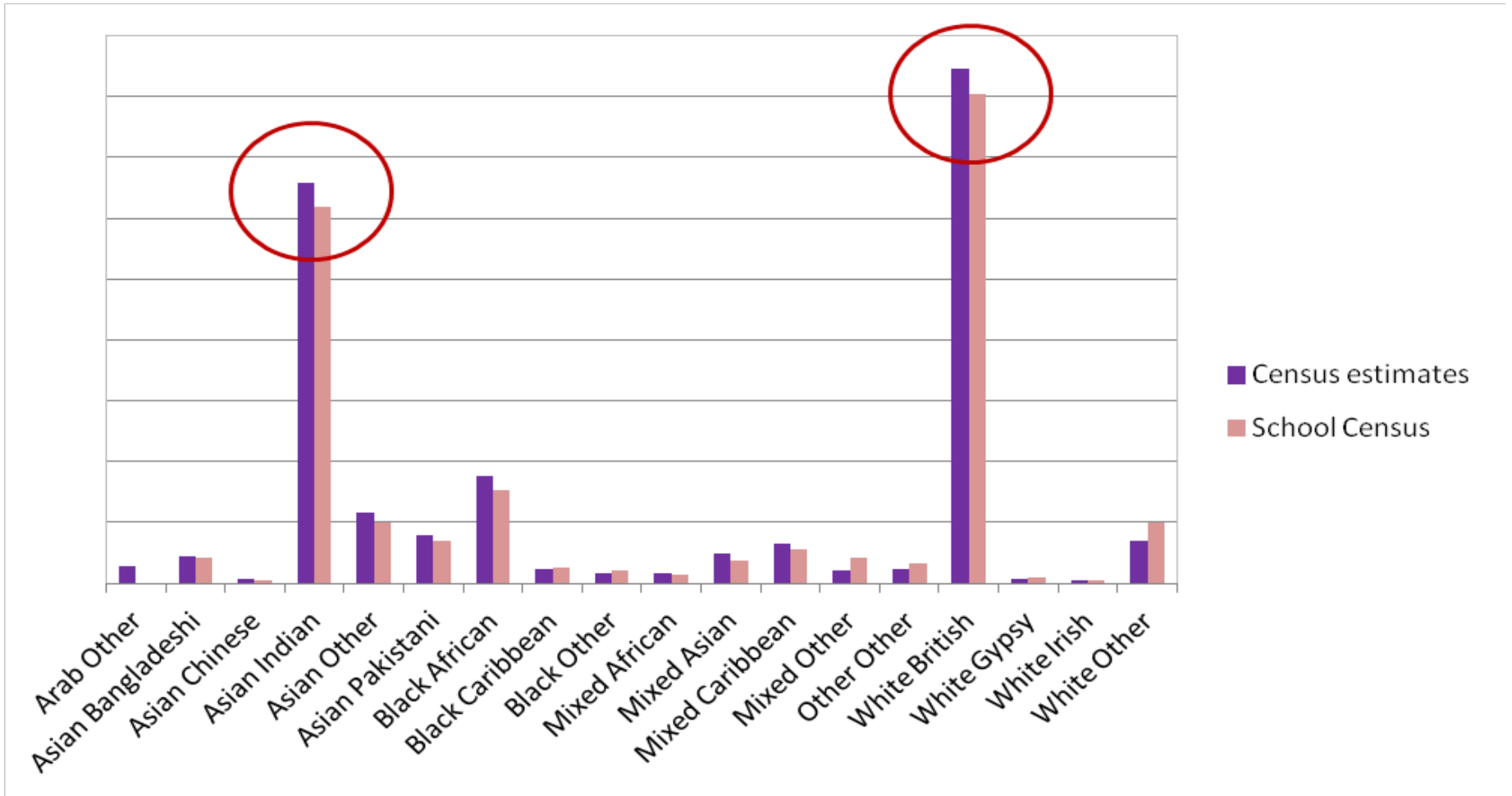
4. Consistency with Ethnicity Comparator

- All persons ethnicity



4. Consistency with Ethnicity Comparator

- Ethnicity of children of school age



Back to case study

Component	Action	Number
Raw Census count	Start	450,305
Dual system & Ratio estimation	Add	19,338
Bias adjustment	Add	6,136
Overcount	Subtract	-2,392
CE Adjustments	Add	703
National adjustments*	Add*	0
Census population Estimates	Finish to QA	474,090
Quality Assurance	Sign-off estimates	Yes

Summary and closing remarks

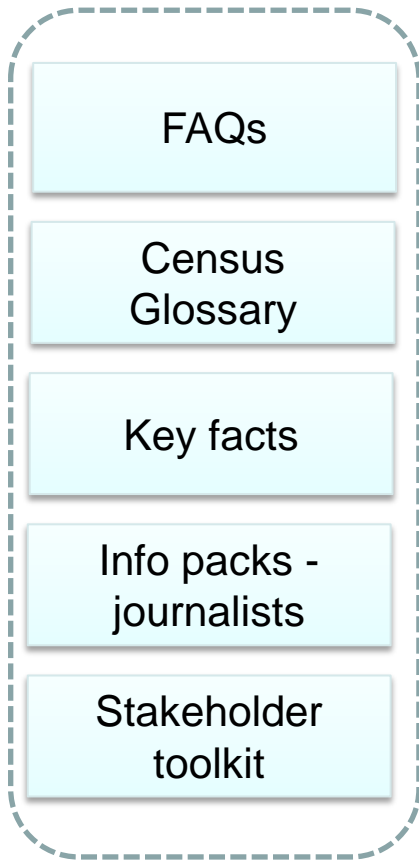
June/July 2012

The 2011 Census First Release

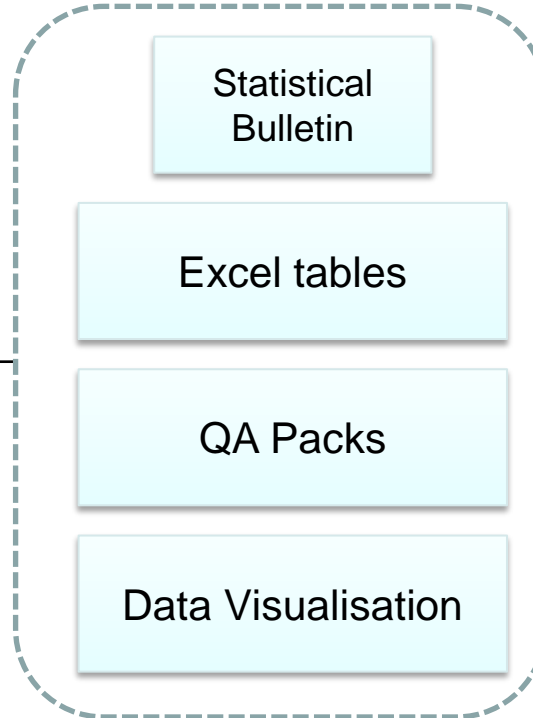
- Emphasis on usually resident census day population estimates and households
 - Coherence and clarity on:
 - How estimates were produced
 - the components of the estimates
 - Evidence used to ensure that the estimates were fit to publish
- To learn from the release of 2001 Census and bring forward those parts of that release to achieve the above
- Other materials aimed at different stakeholders

First release material – Overview

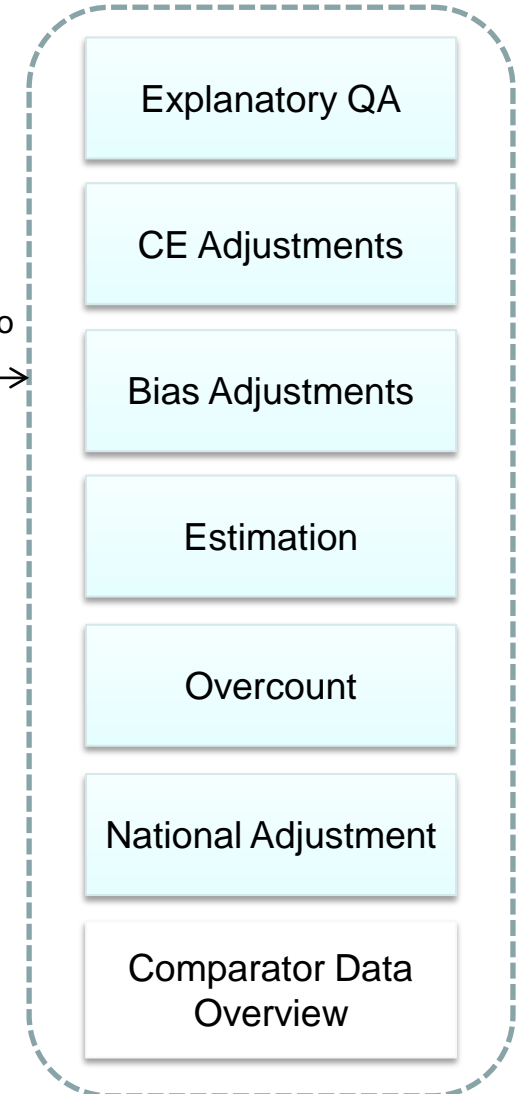
Media & Comms Material



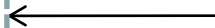
Release information



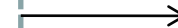
Explanatory Papers



Linked to



Linked to



Census first release - reminder

- Statistical bulletin and tables covering:
 - Usually resident population of E&W by LA by age/sex:
 - Short-term migrant population of E&W by LA
 - Household estimates by LA
- Commentary to highlight key inter-censal and geographic changes
- A range of explanatory material covering topics presented today:
 - Dual System Estimation and bias adjustment
 - Alternative household estimate
 - CE adjustments
 - National adjustment
 - Overcount
 - QA

Census first release – Stakeholder toolkits

- What is it?
 - an online communications toolkit
 - frequently asked questions (FAQs)
 - key messages
 - editorial content
 - guidance on branding and logos
- Who is it for?
 - Users to answer questions from their customers
 - Users to communicate own messages about census outputs.
- Updated as new content is made available

Key points from today

Building confidence:

- transparency in the methods
- Simple demonstrations of complex methods to improve understanding
- Detailed methods based on local information
- Consistent application of methods across country
- Extensive QA
- Wide range of materials explaining the methods

What comes next?

- **16 July 2012 - 1st release of Census results**
- **September 2012 – 2011 MYE (census based)**
- **October 2012 – Census Advisory Group meetings**
- **October 2012 - Short-term 2011 census based population projections**
- **November 2012 - Census outputs and dissemination roadshows**
- **November 2012 to February 2013 – 2nd release of census results**

Thank you

**Please complete your delegate
feedback form**

**We hope you have a safe journey
home**