Office for National Statistics

Using administrative data to set plausibility ranges for population estimates

Date: 27th March 2012
Coverage: Local Authorities in England and Wales
Theme: Population Change

Contents

Executive Summary

The purpose of this report is to present the research into producing plausibility ranges for local authority mid-year population estimates in England and Wales. The aim is to take the available administrative sources, explore the quality of these and the relationships between them and to identify ways in which they can be combined to give upper and lower limits within which the population estimates could reasonably be expected to fall. The research has been completed as part of the Migration Statistics Improvement Programme (MSIP). The report focuses on the methodology and results for children.

There are several data sources that can be used to create plausibility ranges for children. Information about these is provided using a quality framework which has been developed to provide a consistent presentation across different sources. The relationships between each of the data sources and the usual residence definition of the population which is used by ONS are also presented.

Two alternative methodologies for creating plausibility ranges are explained. A 'tolerance range' approach based on aggregate data and a data linkage approach using record level data. The options selected for each age group, with a justification, are provided and the results of calculating these ranges presented.

The results show that the plausibility ranges are narrowest for the youngest children where there is good correspondence between the administrative sources and widest for the oldest children where there is the greatest discrepancy between sources. There are some areas with particular characteristics, such as the presence of the foreign armed forces, where there are issues with the approach. However, generally the methodology is conceptually sound. In many cases where population estimates fall outside the ranges it is only by a small amount. However, there are groups of local authorities with estimates falling outside or close to either the upper or lower limit consistently across all the age groups.

Further work to calculate the ranges for future years and to compare them to the 2011 Census is required to evaluate their robustness. Annexes A to C present initial findings for other age groups where the research is not so far progressed. The work for these age groups also needs to be progressed.

At present the ranges can be used as an additional quality assurance tool for the mid-year population estimates. The research has generated a much greater understanding of the administrative sources and the current population estimates which will be used to inform 2011 Census quality assurance, the rebasing of the mid-year population estimates and the Beyond 2011 Programme. In the longer term improvements to the robustness of the ranges should be possible, which along with development of an adjustment framework could allow direct adjustments to the population estimates to be made where appropriate.

Any queries or feedback in relation to this research report can be done by e-mailing imps@ons.gov.uk

1. Introduction

This research report describes an initial exploration into the use of administrative data sources to set plausibility ranges, or upper and lower limits, within which the population estimates could reasonably be expected to fall. This research has been conducted as part of the Migration Statistics Improvement Programme (MSIP), a cross-government Programme set up to improve the methods and sources underpinning population and migration statistics.

1.1 Aims of the research

The purpose of this research is to:

- Determine whether it is possible to set upper and lower limits for population estimates based on the usual residence definition using administrative data sources
- Demonstrate some alternative approaches using the available sources of administrative data, and
- Develop plausibility ranges based on these approaches.

In the short term these ranges will provide ONS with an additional quality assurance tool to apply to the population estimates. In the longer term, following evaluation against the 2011 Census and assessment of the wider implications, ONS will seek to further improve the robustness of these ranges and to widen their scope such that they may be used as part of the population estimation process.

1.2 Overview of the report

This report focuses on the development and presentation of plausibility ranges for children aged up to 15. There are three key reasons why it was sensible to start with this age group:

a) The annual mid-year population estimates are based on the cohort component methodology. Essentially the population is aged on from the base year, births are added, deaths are subtracted and adjustments made for internal and international migration. It is important, therefore, to ensure that children are counted in the correct location to ensure that as they enter early adulthood and become more mobile their movements are captured from the best base possible.
b) The quality of the administrative sources for children is generally considered to be better than for older age groups. This is in part due to both the greater tendency for them to be registered with a GP and the existence (currently) of universal Child Benefit.
c) The sources of information available to ONS for this group include both aggregate and record level data. This has allowed ONS to demonstrate some alternative approaches to calculating plausibility ranges using the different types of data.

Section 2 of this report reviews the data sources used in the research, including an assessment of quality. Section 3 explores the methodology employed in developing the plausibility ranges, and presents evidence to support the chosen method, while section 4 presents and discusses the findings. Finally, section 5 draws out the conclusions and explores the next steps for this work.

Some initial work has been undertaken examining older age groups. However further work is required to understand the characteristics of the administrative sources before plausibility ranges are produced. A brief summary of this work is included at Annexes A to C.

One of the issues with this research is the lack of independent alternative sources against which to validate the plausibility ranges at this point in time. Whilst they will be assessed against the Census in future, a key aim of this research paper is to invite feedback from users to provide an early assessment based on local knowledge or on locally available sources of data. A feedback

form has been provided in Annex D. Users are encouraged to provide ONS with useful feedback on the methods used to generate the plausibility ranges and on the validity of the ranges proposed.

As the plausibility ranges are being published within a research report with the aim of giving users the opportunity to comment, they are considered to be experimental in nature.

Please either e-mail the feedback form contained in Annex D as an attachment to: imps@ons.gov.uk,

Or mail your response to:

Population Statistics Research Unit
Population and Demography Directorate
ONS Titchfield
Segensworth Road
Titchfield
Hants PO15 5RR

1.3 Relationship with other MSIP reports

ONS has published several research reports alongside this one. Two of these are closely associated with the plausibility ranges:

- A conceptual framework for UK population and migration statistics
  This report explores the context for population and migration statistics, develops the context for the production of these statistics in the UK, and examines the relationship between the concepts ONS is trying to measure, the data sources and processes required to measure them and the impact this has on the final outputs.
- Quality indicators and measures of uncertainty
  The quality indicators provide information about the relative size of certain groups of the population which are associated with being difficult components to measure. The measures of uncertainty will provide an estimate of the variability associated with the population estimates when they are published later in the year.

Both of these outputs provide information about the quality of the population estimates based on the existing methodology. In contrast the plausibility ranges are designed to give an independent upper and lower limit within which we might reasonably expect the population to fall. This is because they include alternative data sources and are based on directly estimating the limits, rather than deriving the population by estimating the components.

2. Data sources

2.1 Introduction

It is important to understand the origins, processes and quality associated with administrative data sources before they are used in official statistics. This section provides key quality information for each of the data sources used to produce plausibility ranges for children. The presentation has been standardised to allow effective comparisons between the quality of information available for each of the sources. Further discussion of the data sources in relation to the usual resident population is provided in Section 3.

The framework used here is based on that developed by Daas et al. (2009) in the Netherlands, a country with a long history of using administrative data sources. It covers three dimensions: the source, the metadata and the data. The subheadings under each of these and an explanation are provided in Tables 2.1 to 2.3 below. This analysis gathered the information available for each of the indicators for the administrative data sources that were used to create the plausibility ranges. For further details on the indicators, see Daas et al. (2009).

Some indicators such as unit non-response are more applicable to surveys. For several indicators, the information was missing or was not (centrally) available. Researchers within ONS have extensive experience with using some of the data sources so their expertise has been used to supplement published information.

Table 2.1: Data Source

1. Supplier	Name of the data sources, data source contact information, National Statistical Institute (NSI) contact person Reason for use of the data.
2. Relevance	Usefulness and importance of the data source for the NSI Potential statistical use of the data source Effect on response burden
3. Privacy and security	Legal provision and basis for existence of the data source Manner in which the data source is sent to NSI Are security measures required?
4. Delivery	Cost of using the data source Are frequency and terms of delivery documented? Formats in which the data can be delivered Does this comply with the requirements of NSI?
5. Procedures	Familiarity with the way the data has been collected and planned changes Ways to communicate changes to NSI Dependency risk of NSI (when not delivered according to arrangements made)

Table 2.2: Metadata

1. Clarity	Population unit definition Classification variable definition Count variable definition Time dimensions Definition changes
2. Comparability	Population unit definition comparison and comparability with NSI definition Classification variable definition and comparability with NSI definition Count variable definition and comparability with NSI definition Time differences and comparability with NSI reporting periods
3. Unique keys	Identification keys comparability with unique keys used by NSI Unique combinations of variables
4. Data treatments by data sources keeper	Population unit checks performed Variable checks performed Combinations of variables checked Extreme value checks Familiarity with data modifications Are modified values marked and how?

Table 2.3: Data

1. Technical checks	Can all the data in the source be accessed? Does the data comply with the metadata definition?
2. Over coverage	Non-population units
3. Under coverage	Missing units, selectivity
4. Linkability	Linkable units, mismatches, selectivity
5. Unit non-response	Units without data, selectivity
6. Item non-response	Missing values, selectivity
7. Measurement	External check, incompatible records, measurement error
8. Processing	Adjustments, imputation, outliers
9. Precision	Standard error on study variables
10. (Sensitivity)	Missing values, selectivity, effect on totals

2.2 Live Births

The registration of life events (births, deaths, marriages, and civil partnerships) is a service carried out by the Local Registration Service in partnership with the General Register Office (GRO) in Southport. The provision of life events data by GRO is formally defined by a service level agreement between ONS and GRO. The provision of births data under the agreement is monitored on a quarterly basis by the Fertility Management Group. The table below gives an overview of the quality indicators for this data source.

	Live Births
SOURCE	
Supplier	The registration of life events is a service carried out by the Local Registration Service in partnership with the General Register Office (GRO) in Southport. The provision of life events data by GRO is defined by a service level agreement (SLA) between ONS and GRO.
Relevance	High. Provides count of births used in population estimates for England and Wales at local authority (LA) level.
Privacy and security	The data are personal sensitive data and covered by the Statistics and Registration Service Act 2007 which makes it a legally required data delivery. Delivery via secure networks with each field having a different security level. Data accessible by designated users with specific levels of access permission.
Delivery	The data are loaded onto the Registration Online System for birth and deaths (RON) overnight.
Procedures	The provision of birth data under the agreement is monitored on a quarterly basis by the Fertility Management Group. From July 2009 almost all register offices were submitting data electronically using RON.
METADATA	
Clarity	Designated National Statistics. Births Logical Data Contents Specification documents specify definitions and coding.
Comparability	Data are presented using legally defined classifications and definitions on births and still births.
Unique keys	Provided at record level for outputs for NHS-Information Centre and approved researchers. Births records are split into a statistical file and a registry file. Each birth record has a RON-ID or system number which is unique to each registration event.
Data treatment by data source keeper	The RON enables Registrars to carry out additional validation checks at the point of registration such as validation of address and postcode.

DATA	
Technical checks	Data are checked on coding, validation, imputation, matching, and data integrity. Outlying values are investigated and verified, e.g. detailed checks are carried out on dates of births implying the age of the mother is <16 or >50.
Over coverage	Duplicates occur and are removed.
Under coverage	Low. Dataset routinely collects birth registration data which cover all births occurring in England and Wales. The cut-off date for inclusion in the annual dataset was births occurring in the reference year that were registered by 25 February of the following year in order to capture late registrations.
Linkability	Linkage possible on variables in the statistical file.
Unit non-response	Negligible. There is a legal obligation to register births and registrars are being chased to record births in hospitals. Eventually, all births get recorded when child is in contact with another administrative system, for example when they start school.
Item non-response	Live births with a missing postcode or residence outside England and Wales are imputed. Under the Population Statistics Act (PSA) certain confidential data items are collected at the registration of a birth. If any of these items are missing an appropriate value is imputed by ONS for the corresponding derived variable. Any remaining missing data items are re-imputed using the Canadian Census Edit and Imputation System (CANCEIS).
Measurement	The accuracy of information contained in the draft birth entry is the responsibility of the informant. Wilfully supplying false information renders the informant liable for prosecution for perjury.
Processing	Extensive checks are performed during processing. Adjustments are made based on investigation of unlikely or outlying values by verifying with the GRO.
Precision	The records on live births are considered to be highly accurate.
Other sources	http://www.ons.gov.uk/ons/guide-method/user-guidance/health-and-life-events/index.html

2.3 Child Benefit

Child Benefit (CB) data is derived from 100 per cent scans from the Child Benefit Computer System (CBCS). The data represent a snapshot at a point in time of all claimants and eligible dependents in England, Wales and Scotland present on the CBCS at 31 August, plus awards for new families and children. Her Majesty's Revenue and Customs (HMRC) have estimated that the take-up of CB is around 96 per cent for 2008-2009. The table below gives an overview of the quality indicators for this data source.

	Child Benefit
SOURCE	
Supplier	Data are derived from 100% scans from the Child Benefit Computer System (CBCS): all claimants and eligible dependents (children) in England, Wales and Scotland present on the 31 August of each year plus awards for new families and children made by 30 November of the same year. CB is administered by HM Revenue and Customs (HMRC). Delivery of data to ONS not legally required.
Relevance	High. Includes family counts, number of children in the family and their age and gender. Used for statistical purposes such as child population estimation by LA.
Privacy and security	Aggregate data. HMRC ensures that the data are non-disclosive by rounding all counts.
Delivery	Data from the August snapshot are available for use in February
Procedures	Data from CB are produced annually at the end of August and November. Changes to the eligibility for Child Benefit due for implementation in April 2013 will impact on coverage.

METADATA	
Clarity	National Statistics. Data are believed to be of high standard as they are based on a scan of the computer system used to administer and pay CB.
Comparability	Data sets prior to 2005 and from 2007 on were counted according to the number of individuals claiming Child Benefit; data for 2005 and 2006 were based on the number of families claiming Child Benefit. This was thought to affect 0.3% of claimants. An issue was identified with data for 2008; data from 2009 onwards are considered comparable and reliable.
Unique keys	Not applicable – aggregate data only are available. ONS has no agreement with HMRC for record level data
Data treatment by data source keeper	HMRC carry out extensive checks to ensure the accuracy, reliability, consistency, completeness and quality of the data. Procedures are constantly being updated and extended. If HMRC and the Department for Work and Pensions DWP cannot solve data problems, the data are not cleared for release and publication.

DATA	
Technical checks	ONS checks that the data may be accessed, are readable, consistent and that they comply with the metadata definition.
Over coverage	Due to liability for prosecution over coverage is unlikely.
Under coverage	Take up for under 1s estimated at 90 to 95 per cent. Take up increases with age but does not reach 100 per cent. Children between 16 and 20 are still eligible if undertaking further 'non-advanced' education; CB is subject to residence conditions.
Linkability	Not relevant as data received at aggregate data
Unit non-response	See under coverage above
Item non-response	For units where the postcode is missing CB is matched with tax credits and in the event of a discrepancy the postcode of the tax credit is used.
Measurement	Claimant information is self-reported. There may be a lag in updating address information as payment through bank accounts is not dependent on this being correct.
Processing	Not relevant as data received at aggregate level
Precision	No precision information as data is received at aggregate level
Other sources	http://www.hmrc.gov.uk/childbenefit/cb-key.htm

2.4 Patient Register

The Patient Register (PR) is a source covering all people registered with an NHS doctor within England and Wales. It contains approximately 58 million records and is used by ONS to calculate migration between local authorities within England and Wales. The table below gives an overview of the quality indicators for this data source. The data are collected by Primary Care Trusts and collated by NHS Connecting for Health.

	Patient register
SOURCE	
Supplier	NHS Connecting for Health. The data supplied are without names and addresses.
Relevance	High. Used as the basis of internal migration estimates, and in the small area population estimation process.
Privacy and security	Data are stored securely in accordance with a service level agreement (SLA) governing use of the data. A secure process is defined for collecting the data from the supplier, covered by the Statistics and Registration Service Act 2007 which makes it a legally required data delivery.
Delivery	An SLA specifies delivery arrangements and data are provided in text format within 1 month of the reference date. Annual snapshot is passed on to the ONS relating to 31 July each year for the production of population and migration statistics for the year ending 30 June each year.
Procedures	The data are collected as part of the GP registration process. The NHS National Back Office (NBO) undertakes work to resolve data quality issues including duplicate registrations.
METADATA	
Clarity	ONS checks that the dataset complies with the specification in the SLA and that all data in supplier files can be accessed.
Comparability	Data relate to one month after the population estimate reference date. This is to account for an assumed one month lag in patients re-registering with a GP after a move.
Unique keys	The NHS number provides a unique key.
Data treatment by data source keeper	There are standard procedures at the Primary Care Trust (PCT) to keep patient registers up to date when there are concerns that a patient has not visited their GP for some time; list cleaning is likely to vary geographically.

DATA	
Technical checks	Data in source can be accessed and complies with metadata.
Over coverage	3000 (approx 0.005 per cent) duplicate NHS numbers identified when the PR is longitudinally linked to itself. Some patients may remain on doctors' lists after having died or left England and Wales.
Under coverage	Individuals who are not registered with a GP will not be included in the dataset, likely to vary by age and sex, for example low registration of 16 to 29 year old males. List cleaning activities may accidentally remove patients who should remain registered.
Linkability	Whilst the NHS number provides a unique key it is not available on other datasets for linking. An alternative approach using Postcode, Date of Birth and Sex fields has been attempted with other data sources.
Unit non-response	See note on under coverage
Item non-response	<1% missing data <10 records with missing sex. Postcode and LA may be missing or imputed.
Measurement	Misreporting, lag in updating or erroneous list cleaning cannot be quantified at present. Incompatible records already dealt with via imputation.
Processing	Patient Register data is almost complete - in a sample of 600K records, 19 records did not have a valid sex code and 257 had a postcode which could not be matched on ONS postcode file. ONS has no information on amount of imputation done by PCTs before we receive the data.
Precision	N/A
Other sources	http://www.adls.ac.uk/department-of-health/gp-patient-register-dataset/?detail

2.5 School Census

The English School Census (SC) collects data on all pupils attending maintained schools in England, including a wide range of demographic information for each child. Data are collected for children aged 2 to 18 years, but has majority coverage in the 5 to16 year age range. The SC excludes pupils at Independent Schools, Pupil Referral Units, Early Years and Alternative Provision (including home schooling). It includes approximately 92 per cent of children. This percentage varies across local authorities. The table below gives an overview of the quality indicators for this data source.

	English School Census
SOURCE	
Supplier	Department for Education (DfE) collects data to ensure appropriate school funding for each local authority (LA). Data provided to ONS under the conditions set out in the Service Level Agreement (SLA) with DfE. Contact information available. Data delivery not legally required. Access allowed for improving population and migration statistics.
Relevance	High. School Census (SC) captures all state sector school children from which counts can be derived. It has complete coverage for this group at ages 5-15 and includes some pre-school and post-16. No response burden.
Privacy and security	The SC, managed by the DfE, collects information from schools in England (Education Act 1996 and Schools Standards and Framework Act 1998). Data are securely transferred to ONS for statistical purposes only. Access is restricted to limited number of named individuals.
Delivery	No charges made by DfE. Detailed arrangements for delivery of data in place including formats and data transfer procedures. Delivery of data is prompt. Information from main data collection updated annually. ONS only has access to variables specified in the Information Sharing Order to meet identified statistical needs.
Procedures	Data are collected from schools' Management Information System (MIS) into DfE's Centralised Data Collection and Management System. This validates data and queries are referred back to the school. Arrangements for regular contact with data supplier in place and form the basis for understanding any changes and unexpected delivery problems.

METADATA	
Clarity	Good understanding of key definitions and any changes over time. Population unit: children 5-15 attending state school Time dimensions: collected termly – ONS uses SC collected in the spring term
Comparability	Comparability varies and not always compatible with ONS definitions. Significant difference from mid-year in reporting period as School Census taken in January.
Unique keys	Each child has a unique Pupil Matching Reference number (PMR). Date of birth, sex and postcode can be combined to create almost unique combinations.
Data treatment by data source keeper	Checks are carried out within DfE's COLLECT system. Good understanding of validation and quality assurance processes and procedures. Changes due in 2013 when minimum age for leaving education will be raised to 17.

DATA	
Technical checks	All data in the source can be accessed and match the metadata definitions
Over coverage	Unlikely as totals are checked against Published Admission Number. A small number of potential duplicate records are being flagged.
Under coverage	Approx 8% nationally for 5 – 15 year olds – children who do not attend state-maintained schools including those attending independent schools, pupil referral units and those who are home educated.
Linkability	Whilst the PMR number provides a unique key it is not available on other datasets for linking. An alternative approach using Postcode, Date of Birth and Sex fields has been attempted with the Patient Register.
Unit non-response	See note on under-coverage
Item non-response	ONS has received data from 2005 onwards and assessed the quality and level of item non-response by variable. ONS conducts geo-referencing on the data to append LA codes, however not available for all records because of processing decisions.
Measurement	DfE data collection site sense-checks data
Processing	Data validated through DfE's Centralised Data Collection and MIS. Number of corrections unknown, no imputation because data are being corrected by the individual school and not centrally.
Precision	Accuracy quite high – schools record in MIS which they use to monitor absence, achievement etc.
Other sources	http://www.education.gov.uk/rsgateway/schoolcensus.shtml

3. Overview of methods

3.1 Introduction

Administrative data provide a set of alternative sources of information from which estimates of the population can be derived. If the target population of the administrative source(s) can be harmonised with the usual resident population that ONS estimates it may be feasible to incorporate the administrative source(s) into the quality assurance process.

There are many papers showing significant academic and public sector interest in reconciling data sources in order to validate existing (official) statistics. These cover a wide range of data sources and applications. Some, such as Bernardi (2010), Thomsen et. al. (1998) and Tikabo (2011) are general methodological papers discussing various topics related to harmonisation and reconciliation. In particular Tikabo (2011) examines a methodology to create “tolerance ranges” to be used in quality assuring the 2011 UK Census.

Others, including Ruotsalainen (2005), and Longva et al. (1998), discuss the advanced use of administrative data to produce a register based Census in Norway and Sweden. A wide range of case studies can also be found – Johnson and Moore (2005), Zauberman et. al. (2009) and Hucks (2011).

All the papers identify that an essential step in the process of combining and harmonising data sources is the assessment of their quality and the feasibility of the reconciliation methods before using them in the production of official statistics. Harmonisation techniques vary dependent upon factors including data quality, type and the potential to link sources reliably. Where more than one data source is available, and both have known strengths and limitations, combining the sources can improve robustness.

This section sets out the initial conceptual comparisons made in order to understand the key similarities and differences between available data sources and the usual resident population. It then discusses where processes to improve reconciliation have been undertaken and explains the techniques used to create the plausibility ranges.

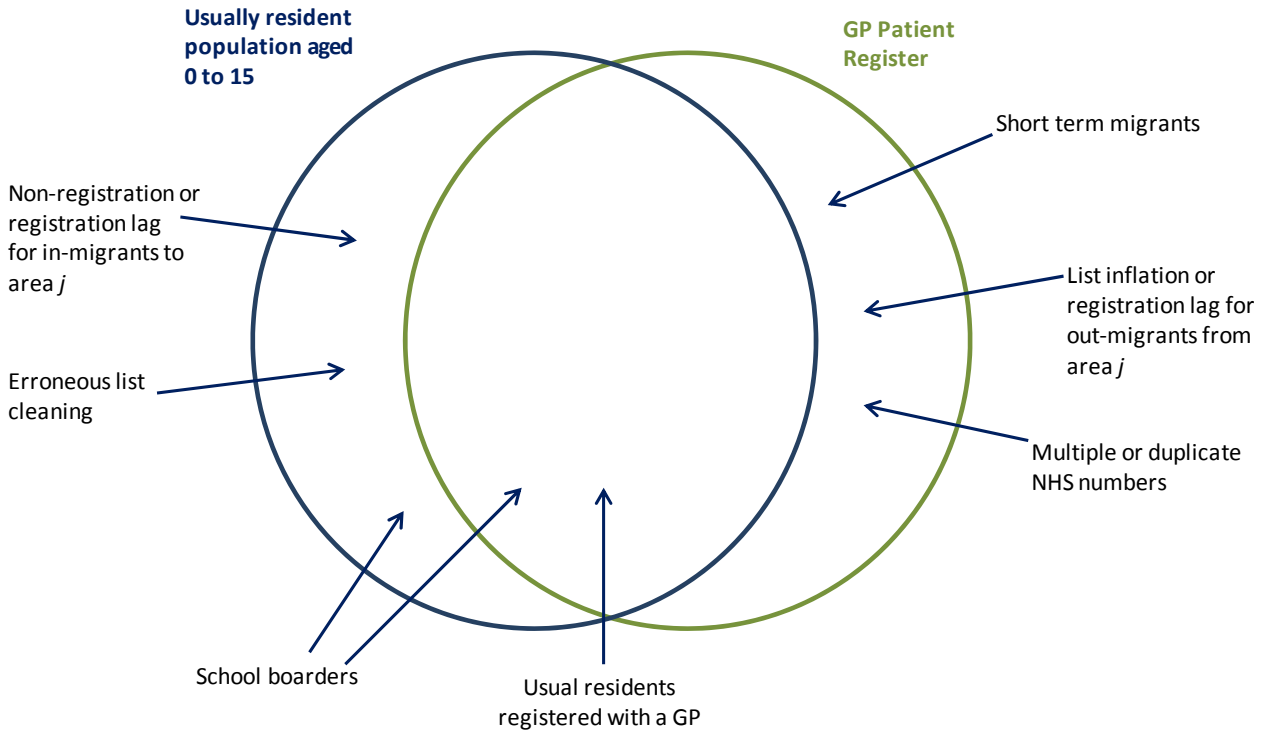
3.2 Initial comparisons and adjustments

In order to understand how administrative sources may be used to create plausibility ranges it is necessary to explore how they relate to each other. Additionally the coverage of the data sources needs to be evaluated against the definition of usual residence used in producing the ONS mid-year population estimates. The discussion below relates to children aged up to 15 only. Results of data comparisons after adjustments are presented in section 4.

3.2.1 Patient Register

Figure 3.1 shows how counts from the Patient Register relate to the usual resident population aged under 16. The area where the circles overlap in the centre of the diagram represents the group of people who are both in the Patient Register and considered part of the usual resident population. The area represented by the left hand side indicates people who are in the usual resident population but not in the Patient Register. The area represented by the right hand side indicates people who are included in the Patient Register but not in the usual resident population. Some of the groups represented relate to definitional issues – for example children at boarding school may appear in the Patient Register either where they attend school or where their parental home is. The others relate to specific issues associated with using the Patient Register to estimate the population at a particular time t, in a particular location j – for example some people may choose not to register with a GP, or may be registered with a GP at their previous address.

Figure 3.1: Relating the Patient Register to the usual resident population aged 0 to 15 in area *j* at time *t*



Patient Register data are available to ONS at individual record level. As a result two adjustments have been made:

1. Removal of any records with duplicate NHS numbers within a year where the number appeared as unique in other years. Table 3.1 shows the number of records removed from the Patient Register in each year between 2006 and 2009.
2. Removal of potential short term migrants. Migrants who register with a GP are allocated a 'Flag 4' against their record on the Patient Register (PR) if their previous address was outside the UK. To remove records of potential short term migrants from the PR, migrants who appeared with a Flag 4 in a given year and who are absent from the data in both the preceding and following year are identified. These records are removed from the data of the year in question. 61,260 short term migrants have been removed from the 2009 Patient Register used for this research.

Table 3.1: Number of duplicate records removed from the Patient Register by year

Year	No. of records removed
2006	1,825
2007	1,218
2008	577
2009	207

Source: ONS

There are several assumptions associated with the approach for short term migrants:

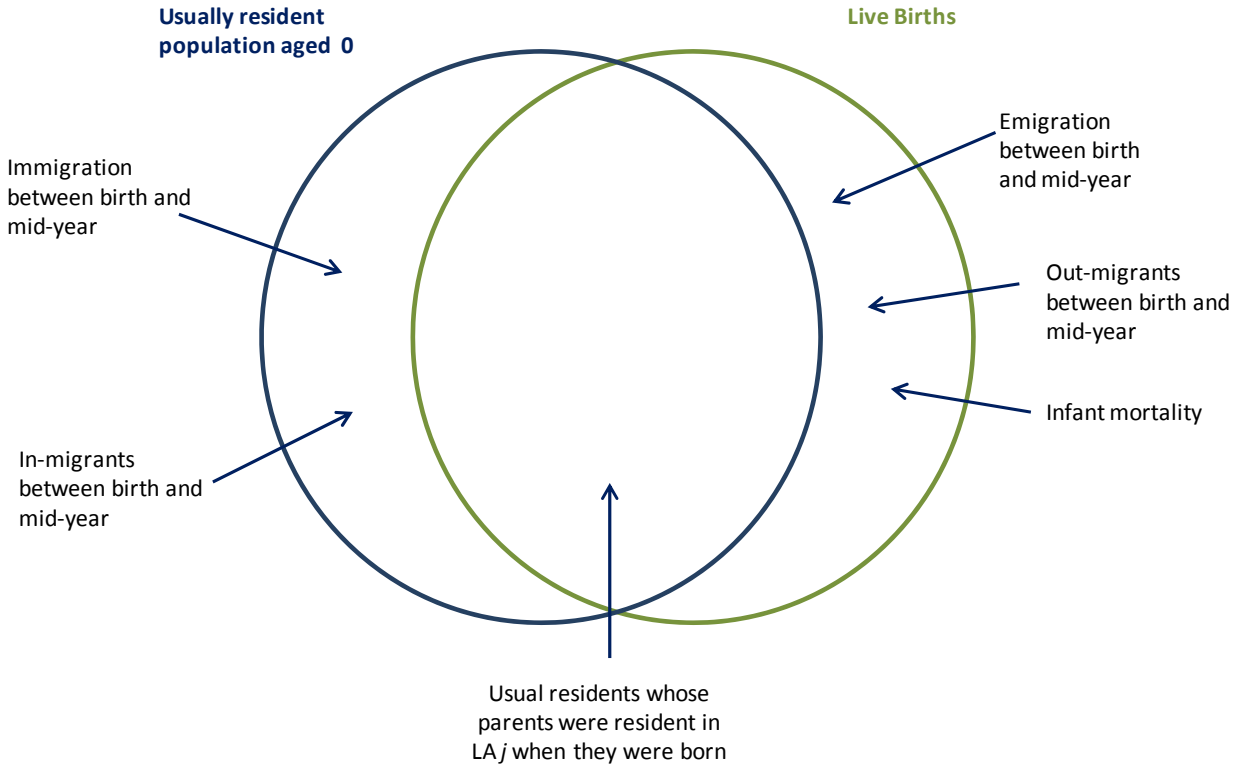
- that migrants register with a GP shortly after their arrival.
- that migrants not appearing in the Patient Register the year after their first appearance had left the UK

- that short-term migrants deregister or are promptly removed from GP lists after leaving the country.

3.2.2 Live Births

Figure 3.2 shows how the number of live births during the year to June 30 relates to the usual resident population aged 0 at mid-year. The central section represents those babies born during the year who have not moved (either within the UK or abroad). The right hand side represents those whose parents are resident in the area at the time of birth but that are no longer there – for example babies who have moved elsewhere with their parents. The left hand side represents those who have arrived in the area since their birth – for example babies who have immigrated to the UK with their parents.

Figure 3.2: Relating Live Births to the usual resident population aged 0 to 15 in area *j* at mid-year



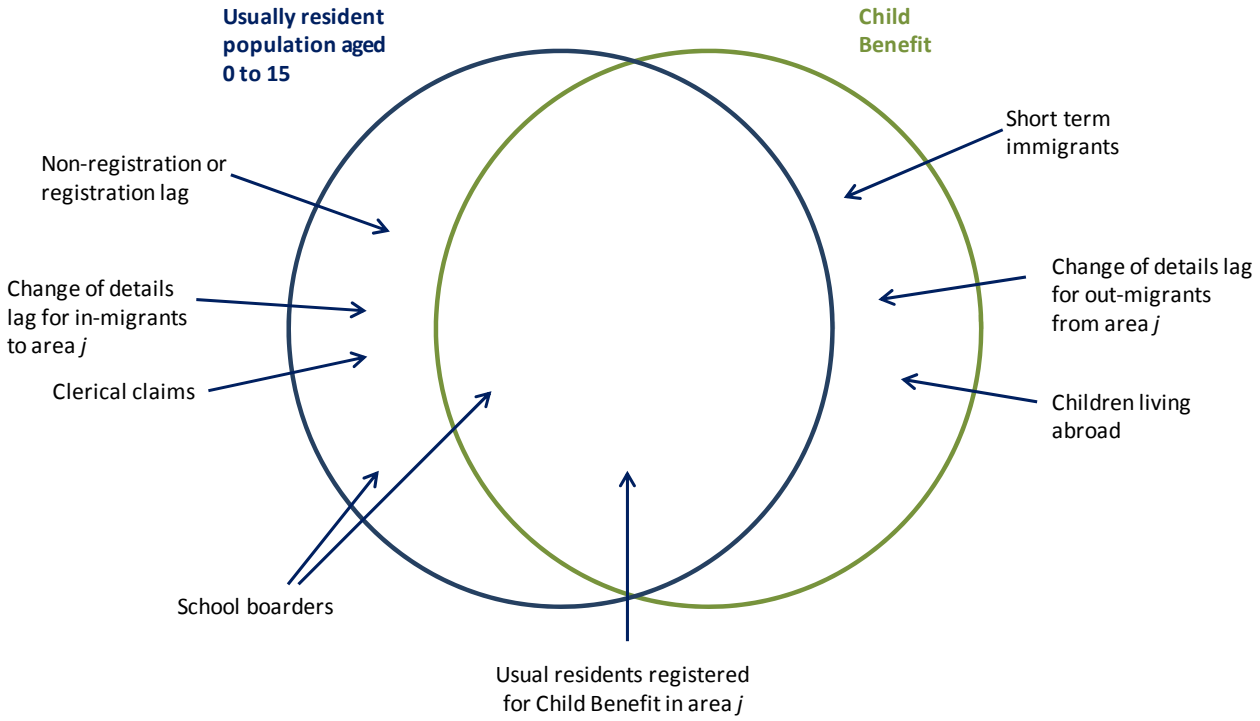
Live Births data are available at aggregate level. An adjustment to correct for infant mortality at the England and Wales level is possible by applying the infant mortality rate and removing this number from the total number of Live Births.

3.2.3 Child Benefit

Figure 3.3 shows how Child Benefit counts relate to the usual resident population aged 0 to 15. As before, the central section represents where the usually resident definition of the population overlaps with children on the Child Benefit dataset. The area on the right hand side indicates children in the Child Benefit dataset but not included in the usually resident definition – for example children who live abroad but whose parents live in the UK and are eligible to claim Child Benefit. The area on the left hand side represents children who are usually resident but not included in the Child Benefit dataset – for example children for whom no Child Benefit claim has been made.

Child benefit data are available at aggregate level only. No adjustments have been made to take account of differences in coverage from the usually resident definition.

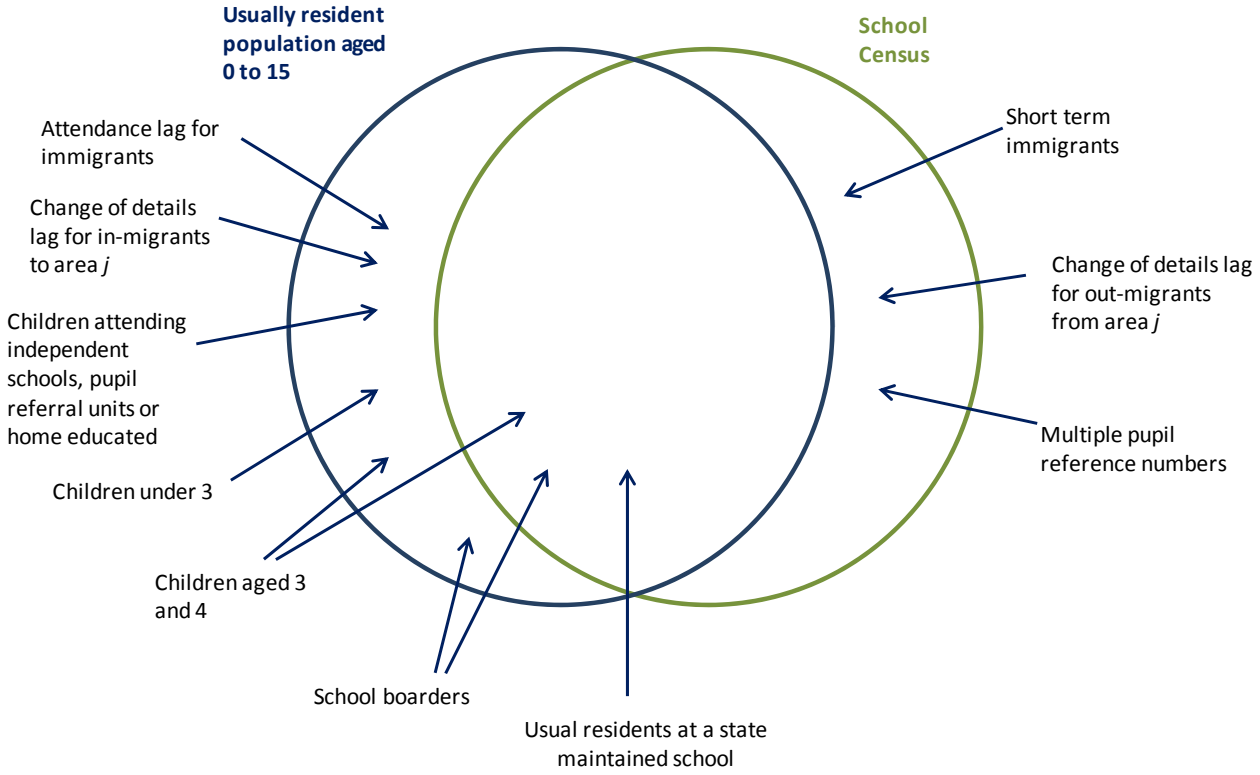
Figure 3.3: Relating the Child Benefit dataset to the usual resident population aged 0 to 15 in area *j* at time *t*



3.2.4 School Census

Figure 3.4 shows how School Census counts relate to the usual resident population aged 0 to 15.

Figure 3.4: Relating the School Census dataset to the usual resident population aged 0 to 15 in area *j* at time *t*



Again the central section represents where the usually resident definition of the population overlaps with children in the School Census dataset. The area on the right hand side indicates children in the School Census dataset but not included in the usually resident definition – for example short term immigrants. The area on the left hand side represents children who are usually resident but not included in the School Census dataset – for example children who attend independent schools.

Although the School Census is available at individual record level, no adjustments have been made for over coverage. In particular for short term immigrants it is not possible to identify new migrant pupils in the dataset in a reliable manner, particularly for 5 year olds. The School Census is deemed not suitable for use for 3 and 4 year olds as coverage for this group is very low.

3.2.5 School boarders

School boarders are treated as a 'special population' within the mid-year population estimates. This means that it is relatively simple to remove them from the estimates. As the diagrams above illustrate, it is not always clear where school boarders appear within the administrative data sources. As a result school boarders have been removed from the mid-year population estimates and the School Census for this research. In the Patient Register some school boarders are registered at the school address, others are not and are assumed to be registered at their parental address. It is not possible to identify and therefore remove school boarders reliably within the Patient Register or Child Benefit datasets. For Child Benefit the school boarders are most likely to be allocated to the local authority of the parent who is claiming, rather than the local authority where the child attends school.

3.3 Development of plausibility ranges

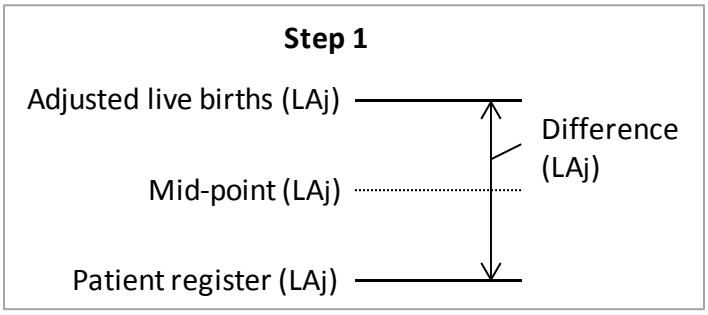
In order to develop the plausibility ranges several possible options were explored across the age groups dependent on the available data. The methodologies presented below are those used to produce the results presented in this report. Further details about the alternative options assessed are referenced where appropriate.

The options for calculating plausibility ranges are restricted according to how many and what type of data sources are available. For this research there were no ages for which only one data source was available, so assessment of single source ranges was not undertaken. As discussed in the introduction to this section it was considered most appropriate to use several sources together in order to use the combined strengths to overcome individual weaknesses.

3.3.1 Tolerance ranges based on aggregate data sources

Where one or more data sources are available at aggregate level only the 'tolerance range' approach proposed in Tikabo (2011) has been investigated. Figure 3.5 illustrates the methodology underlying this approach using adjusted live births and the Patient Register as example data sources.

Figure 3.5: Illustration of the steps to create aggregate data tolerance ranges

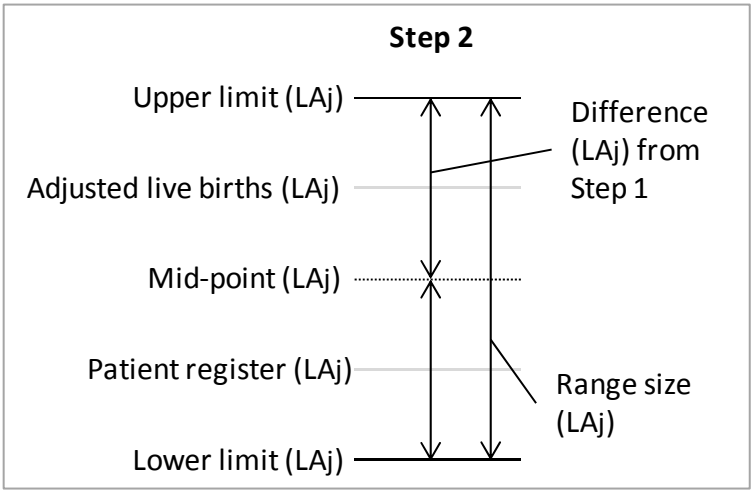


Step 1: Calculate the difference between the Patient Register count and the adjusted Live Births count at LA level

Step 2: Calculate upper and lower limits

Upper limit = Mid-point + Difference

Lower limit = Mid-point – Difference



Step 3

Range size (LAj)
=
2 x Difference (LAj)

Percentage range size (LAj)
=
Range size (LAj) / Mid-point(LAj)

Step 3: Calculate the size of the range as a percentage of the range mid-point

Three alternative options for constraining the minimum and maximum percentage range sizes are possible:

a) No constraint.
b) Setting a fixed minimum and maximum percentage range size.
c) Setting constraints on percentage range size based on the observed data.

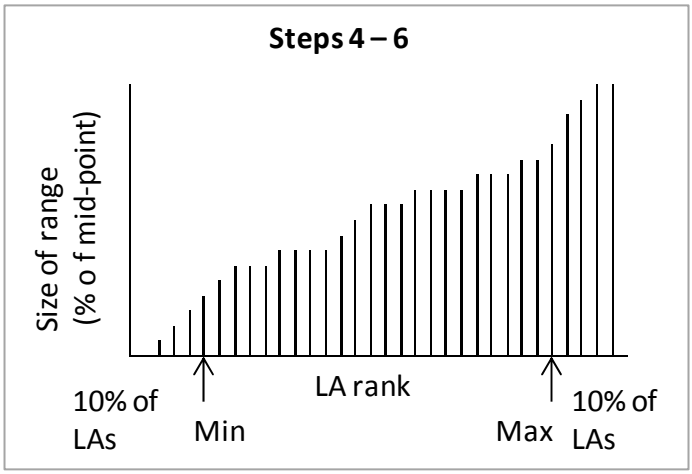
Option (c) is preferred as it provides a minimum range size to reduce sensitivity in cases where the correspondence between independent administrative sources is exceptionally good, and a maximum range size to increase sensitivity where the correspondence between sources is very poor. The minimum and maximum are based on the spread of the size of the ranges rather than a fixed cut off determined somewhat arbitrarily. Figure 3.6 below extends the illustration to demonstrate how option (c) has been implemented.

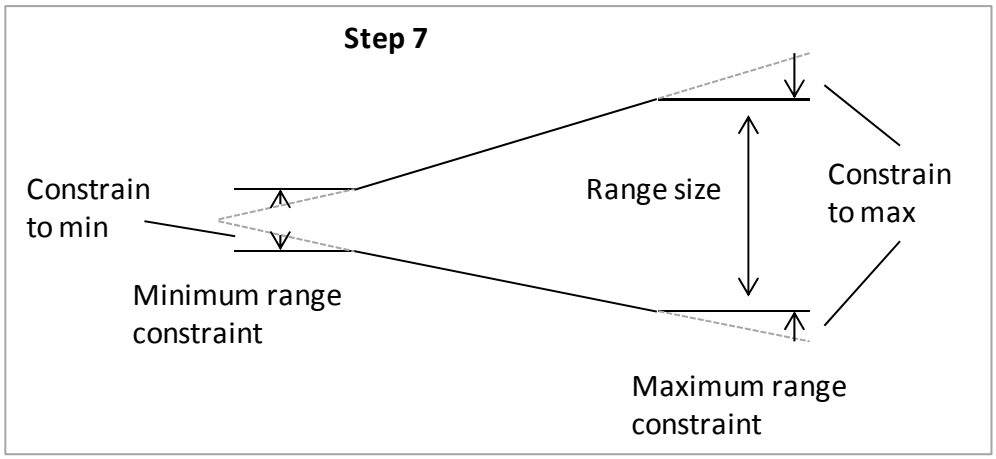
Figure 3.6: Illustration of the steps to constrain the percentage range sizes

Step 4: Rank all the LAs in order of the size of the percentage range calculated in step 3

Step 5: Calculate the minimum percentage range size so that 10 per cent of LAs fall below this point

Step 6: Calculate the maximum percentage range size so that 10 per cent of LAs fall above this point



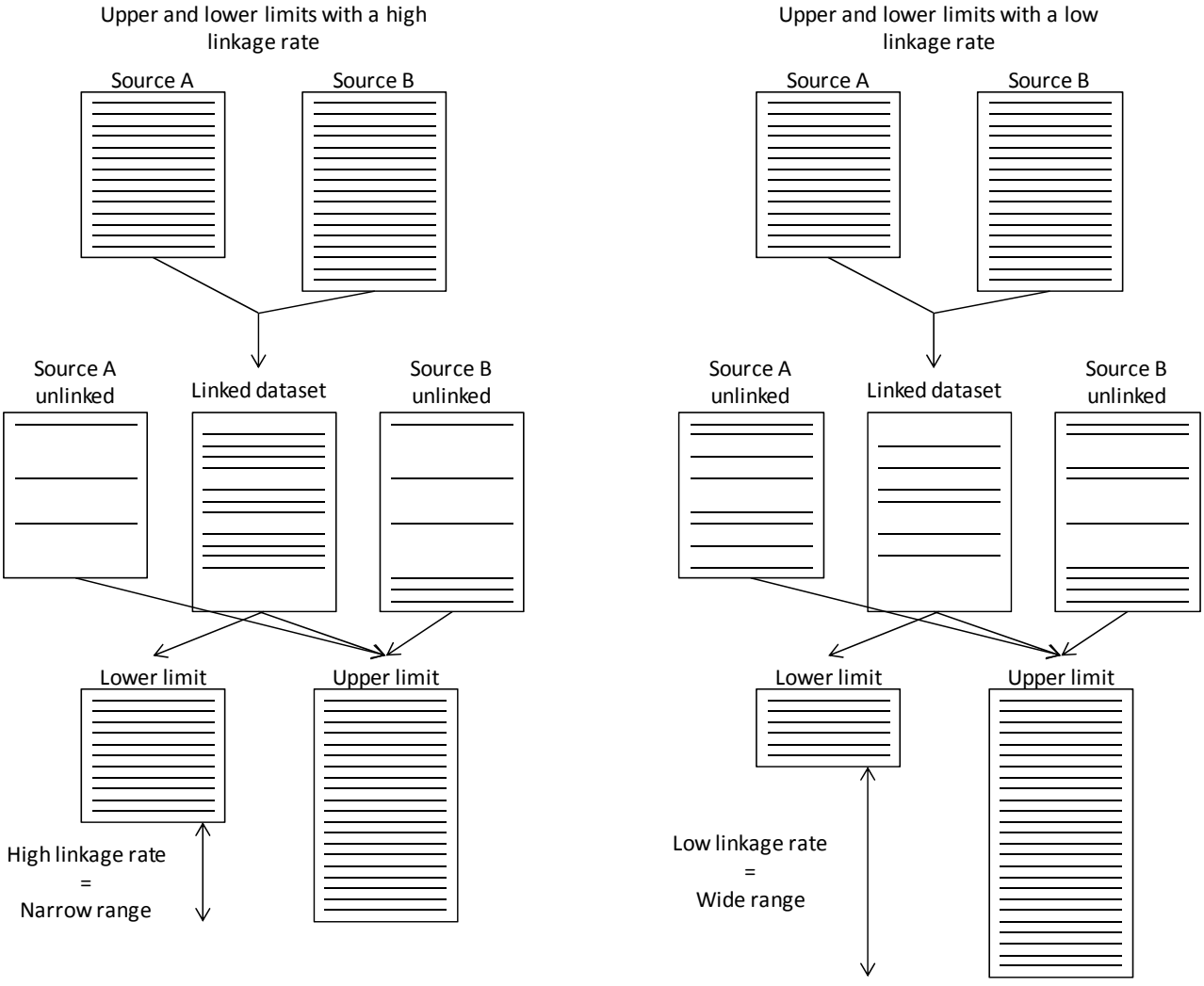


Step 7: Constrain the ranges that fall above or below the percentage range size minimum and maximum

3.3.2 Ranges based on individual record level data sources

Where two data sources are available at individual record level there are additional options for creating plausibility ranges. These are based on linking the data sources, and then assessing how the linked dataset can be used to provide either an upper or lower limit. The approach and the impact of linkage rates on sensitivity are illustrated in Figure 3.7 below.

Figure 3.7: Illustration of possible approach to ranges based on record level data



For the lower limit one very conservative option is to use only the linked records to provide a count. The extent to which records on the two sources are linked influenced the sensitivity, with a low linkage rate leading to a lower more conservative estimate. For the upper limit the most conservative option is to use the full extent of records from one source in combination with the unlinked records from the other. Again the sensitivity is dependent on the linkage achieved between the sources, with a low linkage rate leading to a higher and more conservative estimate.

3.3.3 Under one year olds

There are three data sources available to help estimate the population of babies aged under 1 at mid-year; Live Births, the Patient Register and Child Benefit. Although the Patient Register is a record level dataset, it is used in aggregate form as Live Births and Child Benefit are only available as aggregates. As a result the tolerance range approach illustrated in Figures 3.5 and 3.6 above is used for this age group.

Initial comparisons of the sources at England and Wales level show that the difference between Live Births (after adjustment for infant mortality) and the Patient Register was in the order of 1 per cent whereas Child Benefit is around 6 per cent lower. This is consistent with analysis of the metadata which indicates that Child Benefit take up increases with age, and is particularly low for babies in their first year. However comparison at local authority level reveals that although Child Benefit is often the lowest of the sources this is not always the case, and that the size of the difference between Child Benefit and the Patient Register varies considerably.

In order to avoid the Child Benefit data lowering the ranges in many areas, and to take advantage of the high quality of the Live Births total count for England and Wales, these two data sources are combined by distributing the Live Births across local authorities in England and Wales based on the Child Benefit data. This allows the local authority distribution of Live Births to be updated for migration between the time of birth and mid-year as migration within the UK for this age group is high. The Child Benefit distribution is used to provide an independent combined source to use in conjunction with the Patient Register to calculate the plausibility range. Figure 3.8 illustrates the additional step required before following the methodology outlined in Figures 3.5 and 3.6 above.

Figure 3.8: Illustration of the initial step to create plausibility ranges for children under 1

Step a: Adjust the Live Births dataset to allow for infant mortality at England and Wales level

Step b: Calculate adjusted live births for each local authority using the Child Benefit distribution

Steps a and b

Adjusted live births (LAj)
=
(Live Births E&W – Infant mortality E&W)
x
(Child Benefit LAj/Child Benefit E&W)

3.3.4 One to four year olds

For the 1 to 4 year old age group there are two data sources available, the record level Patient Register and aggregate Child Benefit data. At England and Wales level for this age group the Patient Register is 3 per cent higher than Child Benefit. However this varies considerably across local authorities, with 35 areas for males and 42 areas for females where Child Benefit counts are higher than those from the Patient Register. The tolerance range approach illustrated in Figures 3.5 and 3.6 above combining Patient Register and Child Benefit is used for this age group.

3.3.5 Five to fifteen year olds

3.3.5.1 Introduction

Two record level data sources are available for the 5 to 15 year old age group – the Patient Register and the School Census. Aggregate level Child Benefit data are also available. With the additional School Census data at record level it was possible to explore whether ranges based on record linkage were feasible. Both datasets contained unique identifiers which made longitudinal linkage of the datasets for each year relatively straightforward. However, these identifiers were not consistent between sources making it more complicated to link the longitudinal databases to each other.

The remainder of this section is divided into two parts describing the method and a short summary of the record linkage undertaken, and explaining the approach to create the plausibility ranges.

3.3.5.2 Record linkage

On both datasets the date of birth, sex and postcode for each record have been concatenated to create a consistent "matching variable". Where duplicates occur on a dataset the records are removed. The longitudinal databases are then compared to identify matches. Where a match is made the record is flagged and not considered against any further possible matches. The matching methodology is illustrated in Figure 3.9.

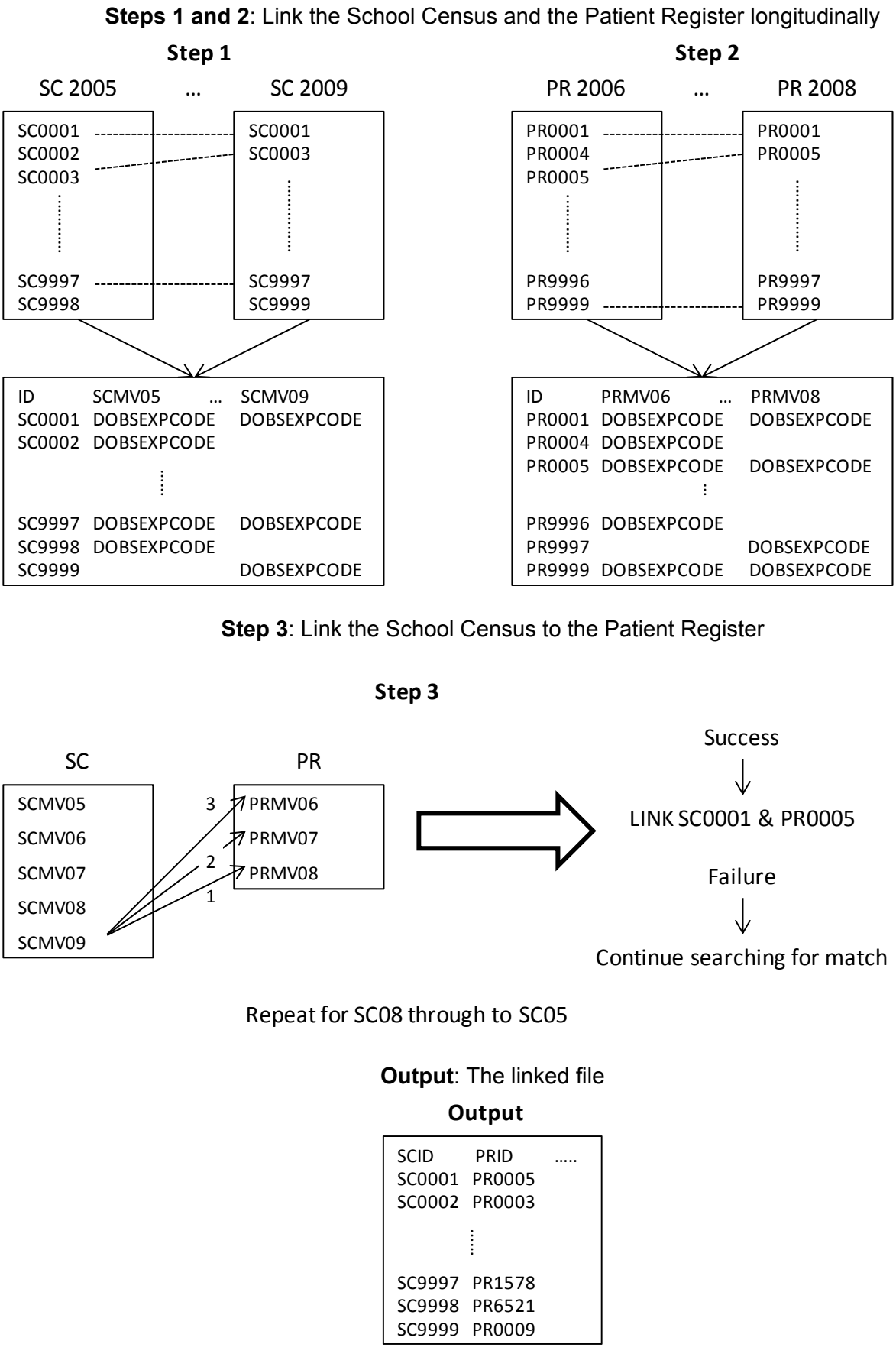
The matching process compared the 2009 School Census with the 2008 Patient Register first, then the 2007 Patient Register, and so on until all fifteen potential pairs of datasets had been examined. Subsequently the 2009 Patient Register data became available and the School Census for each of the years was compared to this and further matches made.

Summary information from the matching process is presented in Box 3.1 below. This shows that the linkage rate between the School Census and Patient Register is high, however there is considerable variability in the linkage rates between local authorities. The variability is much greater when the number of linked records is compared to the number of available records on the Patient Register. This result is expected as children at independent schools are not included in the School Census, but are likely to be included in the Patient Register. Areas with high numbers of children at independent schools feature prominently in the list of areas with the lowest linkage rates.

Box 3.1: Summary information about the School Census Patient Register matching process

Total number of records on longitudinally linked School Census file	9,949,914
Total number of records linked to the longitudinally linked Patient Register	9,131,296
Overall linkage rate (School Census)	91.8%
Linkage rate for 5 to 15 year olds (School Census)	94%
Local authority linkage rate range (School Census)	87% to 97%
Linkage rate for 5 to 15 year olds (Patient Register)	87%
Local authority linkage rate range (Patient Register)	48% to 95%

Figure 3.9: Illustration of linking the School Census to the Patient Register



The linkage rate is also very low for new Flag 4[1] records on the Patient Register, so some of the local authorities with low Patient Register based linkage rates are those with high numbers of new Flag 4 records.

There are 11 areas with linkage rates for 5 to 15 year olds below 90 per cent when compared against the available records for linking on the School Census. Of these areas nine have high numbers of new Flag 4 records on the Patient Register with low linkage rates (below 40 per cent of new Flag 4's linked to a School Census record). The final area, Richmondshire, has fewer children than the other areas in this group, a significant armed forces population and is one of the few local authorities where the number of children on the Patient Register is lower than the number on the Child Benefit dataset. The low linkage rate for this area is likely to be related to these factors.

3.3.5.3 Approach to plausibility ranges

With the wide range of data sources available to contribute to plausibility ranges for five to fifteen year olds there are a number of alternative options. These are outlined in Box 3.2 below. The option in bold is the one that has been selected.

Box 3.2: Plausibility range options for 5 to 15 year olds

Lower limit	Upper limit
1) School Census	1) School Census
2) Patient Register	2) Patient Register
3) Child Benefit	3) Child Benefit
4) Linked School Census Patient Register	4) Linked School Census Patient Register
5) Tolerance range lower limit	**5) Tolerance range upper limit**

Given the principle of using the combined strengths of data sources options 1 to 3 are discounted, although the possibility of using the School Census as a lower limit on its own is worth consideration as the source only covers 92 per cent of children. However, there are large differences in the coverage between local authorities so, while initially attractive, this approach is not appropriate.

The linked School Census and Patient Register has been selected for the lower limit for two key reasons. Figure 3.7 illustrates how a linked dataset may be used to provide a lower limit. A link means that an individual is present in both datasets at the same postcode. Given that a key concern regarding the quality of administrative data is that location information may not be kept up to date this is considered very important. The undercoverage of the School Census, coupled with the need to make a link to the Patient Register means that this estimate really should provide a lower bound for 5 to 15 year olds.

For the upper limit the tolerance range approach used for 1 to 4 year olds based on Child Benefit and Patient Register and illustrated in Figures 3.5 and 3.6 has been selected. Consideration has been given to using the linked dataset approach to producing the upper limit (see Figure 3.7, upper

[1] A Flag 4 indicates somebody who has registered with a GP whose previous address was abroad.

limit), however the extent of overcoverage using this approach is unknown. Also this approach uses unlinked records from both datasets, and as a result, while it does have some of the benefits of using combined data, it also suffers from some of the limitations related to using single sources, particularly issues regarding the quality of address information.

Ranges are calculated for the following three sub-groups within the 5 to 15 age range in order to improve sensitivity; 5 to 7, 8 to 11 and 12 to 15. These have been chosen to coincide with the key ages at which children change school to provide some resilience against any anomalies that may arise from changes in the relationship between the School Census and the total number of children at these ages.

3.4 Summary

A range of data sources are available for use in the creation of plausibility ranges for children. The methodology developed, with some justifications has been outlined above and is summarised in Table 3.2 below.

Table 3.2: Plausibility range approaches and age groups

Age group	Lower limit	Upper limit
Under 1	Tolerance range adjusted Live Births and Patient Register	Tolerance range adjusted Live Births and Patient Register
1 to 4	Tolerance range Child Benefit and Patient Register	Tolerance range Child Benefit and Patient Register
5 to 7	Linked School Census and Patient Register	Tolerance range Child Benefit and Patient Register
8 to 11	Linked School Census and Patient Register	Tolerance range Child Benefit and Patient Register
12 to 15	Linked School Census and Patient Register	Tolerance range Child Benefit and Patient Register

4. Results

4.1 Introduction

This section presents a comparison of the administrative data sources before introducing the plausibility ranges and comparing them to the population estimates.

Illustrative results are presented in this section. Results for individual areas are available in a separate Excel download. A full set of maps and graphs for all local authorities by age group and sex can also be found in a separate pdf download which includes a reference map to aid the identification of local authorities.

The *indicative* mid-year population estimates (iMYEs) excluding school boarders have been used for all comparisons. The immigration component used to calculate the iMYEs is based on the new regional distribution method, generally agreed to be an improvement over the method used to calculate the current official mid-year estimates[2].

4.2 National level comparison between administrative data sources

Figure 4.1 shows the relationship between the administrative sources by single year of age. Overall the administrative sources (Patient Register, Child Benefit and School Census from age 5) and the iMYEs show similar age distributions despite the differences in the levels.

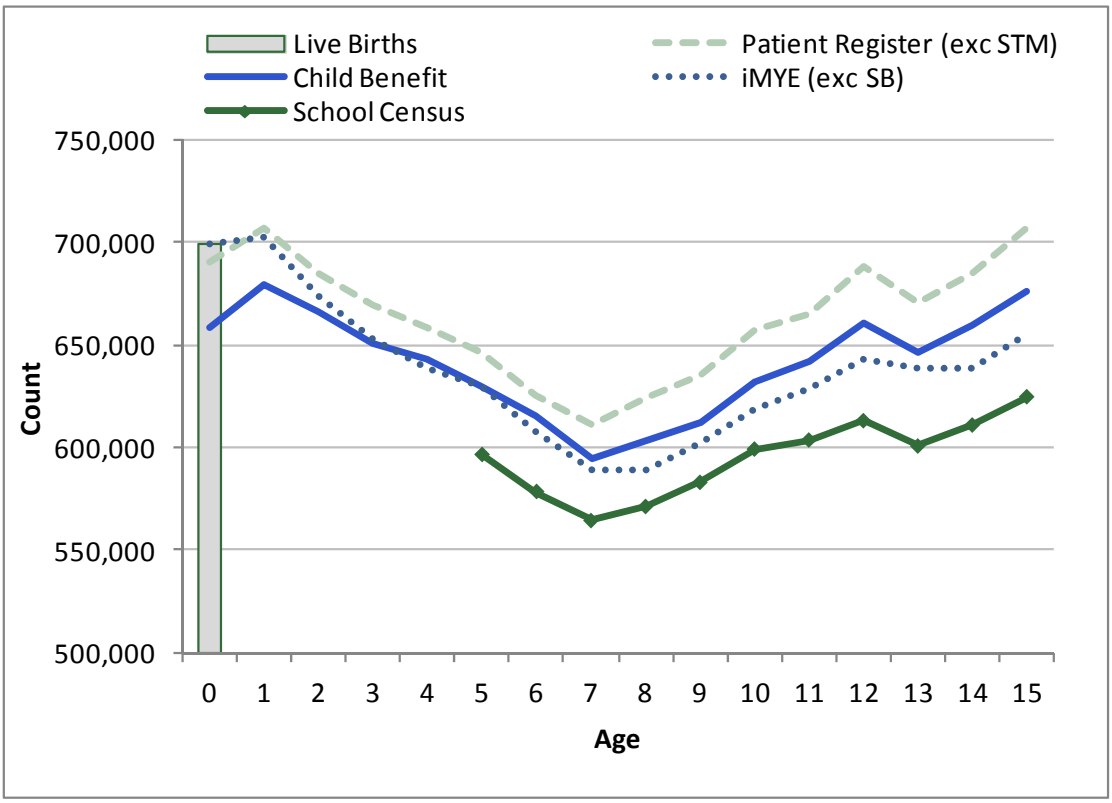
For children under 1 year of age the Live Births data provide the best coverage. This is likely to be due to the legal obligation to register. However, the local authority level counts are based on where babies' parents are resident when they are born, which may not necessarily be the same as where they live at 30 June which is what is required for the population estimates. Patient Register data are also considered to have good coverage of children under 1 as they tend to be registered quickly after birth and have regular interaction with the health service. Both these sources correspond well with the iMYE totals for England and Wales at this age. Child Benefit data give the lowest counts, due to the lag in take-up. HM Revenue and Customs estimate a take-up rate of between 90 and 95 per cent for under 1s.

Between ages two and seven the iMYEs are very similar to the Child Benefit counts. Patient Register counts are higher, and, from age five, School Census counts are lower as would be expected given the exclusion of certain groups from this source.

The iMYEs drop below the Child Benefit counts from age six. The gap between the School Census and other sources widens from age 11 when children usually transfer to secondary school, possibly reflecting a greater level of take up of private education from this age.

[2] http://www.ons.gov.uk/ons/guide-method/method-quality/imps/improvements-to-local-authority-immigration-estimates/overview-of-improved-methodology.pdf

Figure 4.1: Administrative sources by single year of age, England and Wales 2009



The differences between sources seen in Figure 1 shape the plausibility ranges. Where there are large differences between sources the ranges are wider. As a result the use of the School Census in the ranges for 5 to 15 year olds means that the ranges for this age group are generally wider than those for the younger ages.

4.3 Local authority level comparison

Figures 4.2 to 4.4 present comparisons between the administrative sources and the iMYEs by age and demonstrate some common themes that can be seen in the full set of local authority comparisons. Most areas show a relationship between the sources similar to the examples for Camden or Chesterfield in Figure 4.2. In Camden the Patient Register and iMYE correspond fairly well and the Child Benefit counts are generally lower reflecting Child Benefit take up below 100 per cent. The School Census counts are lower still highlighting the gap between children resident and attendance at state maintained schools. In Chesterfield all sources correspond very well.

Other patterns can also be observed, in particular local authorities where the iMYE gives the lowest estimate from age eight onwards (Figure 4.3) and age one onwards (Figure 4.4).

A complete set of charts for local authorities showing the relationships between the administrative data sources and the iMYEs by age are given in a separate pdf download.

Figure 4.2: 'Typical' relationships between administrative data sources

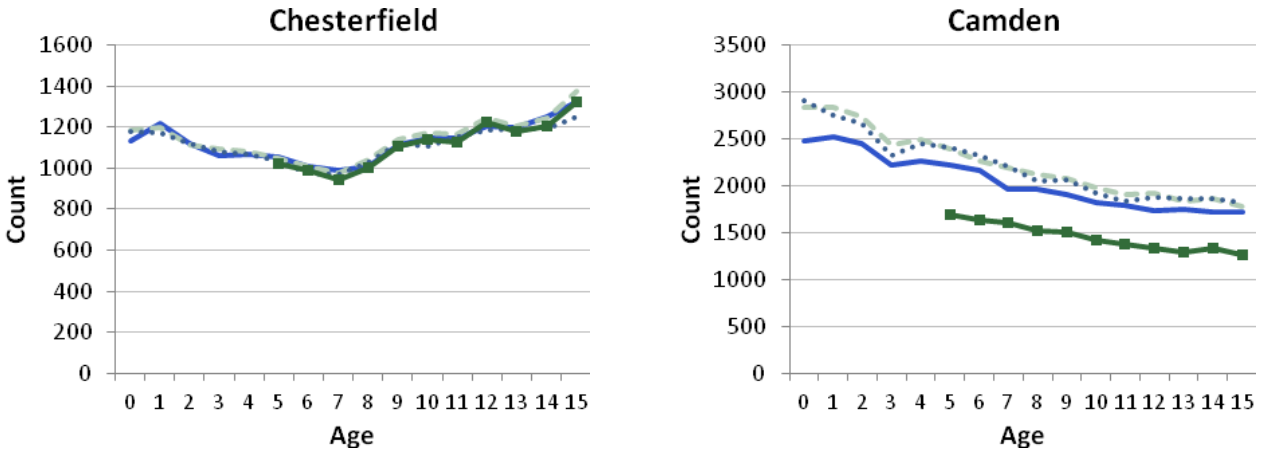


Figure 4.3: Relationship between administrative sources where iMYE is low from age 8 onwards

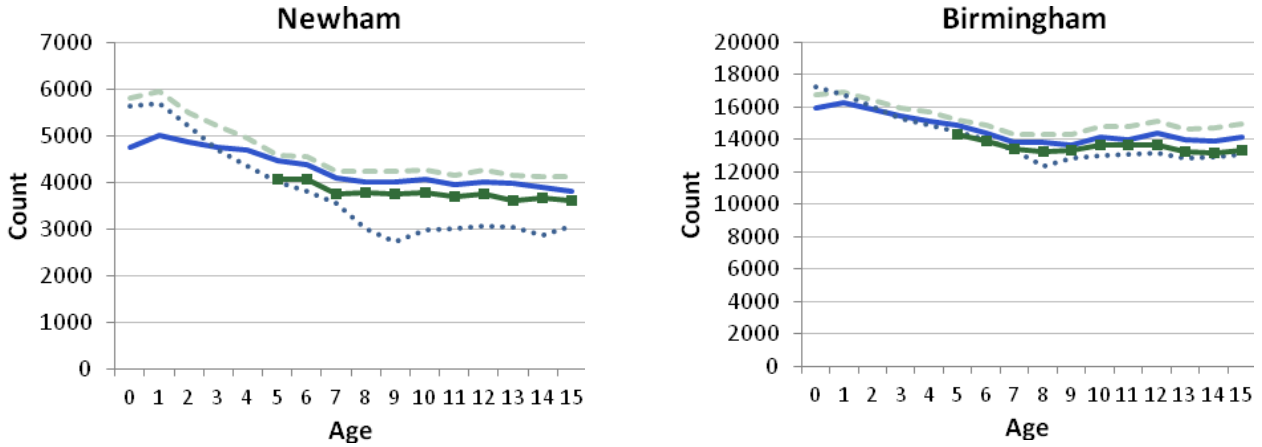
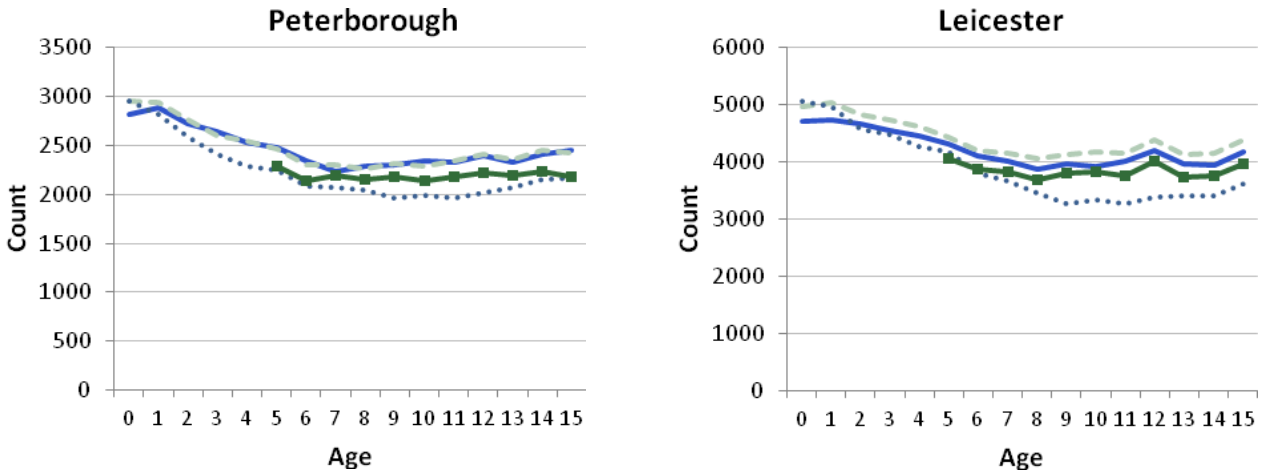
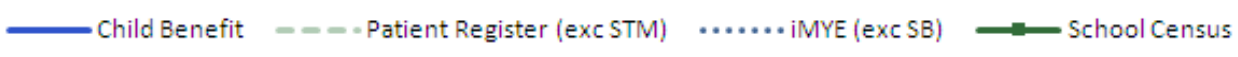


Figure 4.4: Relationship between administrative sources where iMYE is low from age 1 onwards



Source: Patient Register, Child Benefit and School Census data England, 2009



4.4 Plausibility ranges for children under one year old

4.4.1 Introduction

The 'tolerance range' approach is used for this age group to calculate both the upper and the lower limit. Live Births are adjusted for infant mortality and the local authority distribution revised to match that for Child Benefit to account for family moves between birth and the mid-year point to which the population estimates relate. The adjusted Live Births data and the Patient Register are used to create the ranges.

4.4.2 Relationship between the sources used

The adjusted Live Births exceed the Patient Register counts for the majority of local authorities. Differences between the two administrative sources are most likely to be caused by delays in the time it takes for changes in circumstances to feed through to them. For example, there may be a lag in initial registration with a doctor or re-registering with a new one following a move.

4.4.3 Plausibility ranges

The good correspondence between the counts in the administrative data sources for this age group result in narrow ranges. About one third of local authorities fall outside of the plausibility ranges, with more exceeding the upper limit than falling below the lower limit. Figure 4.5 shows where the iMYEs fall within and the extent to which they fall outside the plausibility range for females. Very few local authorities had iMYEs that fall outside the limits by more than 5 per cent of the mid point of the range.

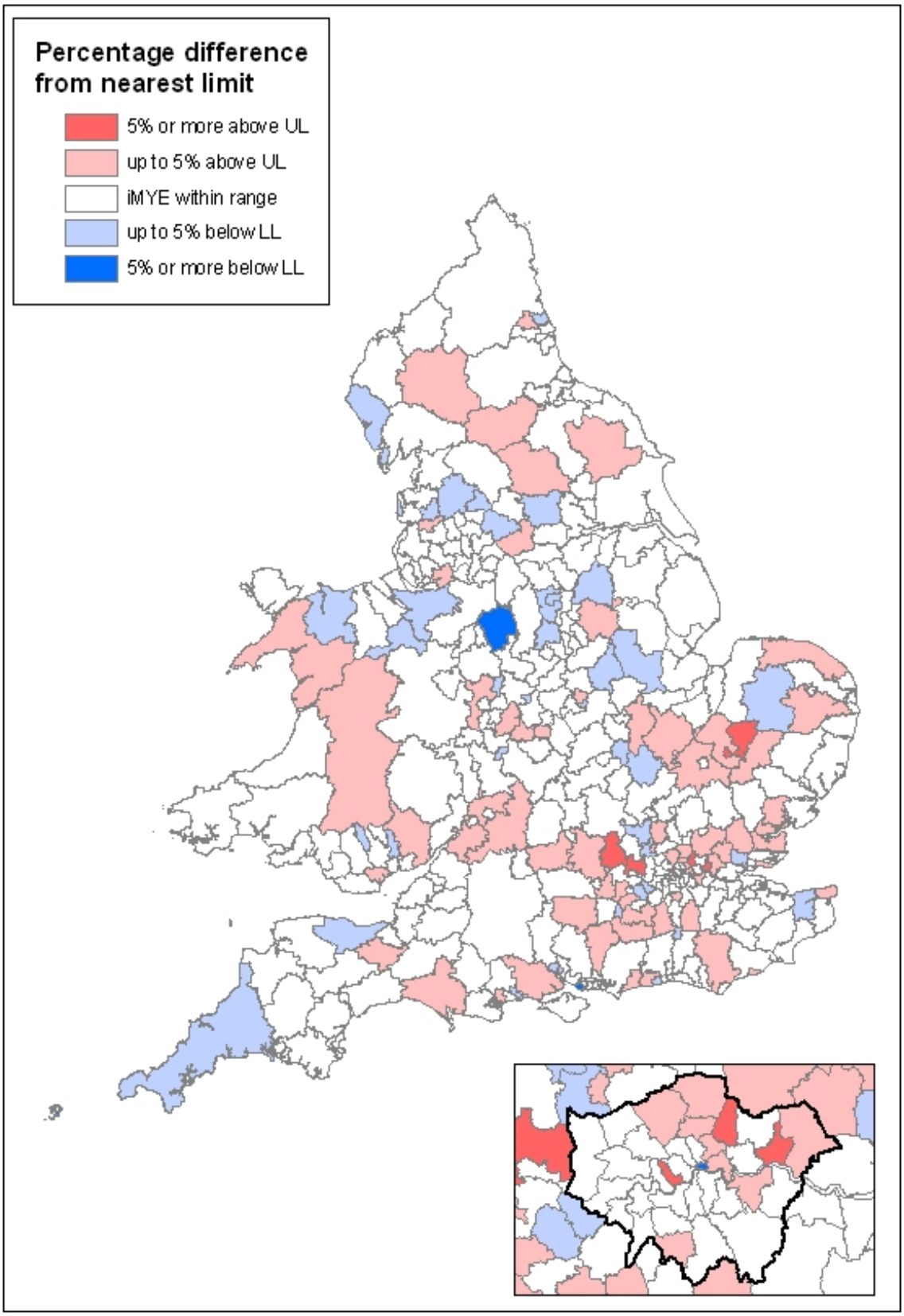
Estimates for both males and females in Forest Heath and Kensington and Chelsea both exceed the upper limit by more than 5 per cent of the mid point. This was also the case in the other age groups. Forest Heath has an important foreign armed forces population, and Kensington and Chelsea is characterised in part by foreign embassies with highly mobile populations. This is discussed further in section 4.7.

Other local authorities for which population estimates are more than 5 per cent higher than the upper limit of the plausibility ranges for both sexes are Barking and Dagenham and Waltham Forest.

Estimates for both males and females in the City of London and the Isles of Scilly both fall below the lower limit by more than 5 per cent of the mid point, however as both of these local authorities have only very few children the absolute differences are very small. Gosport and Staffordshire Moorlands are also below the lower limit by more than 5 per cent of the mid point for females, but not for males.

Plausibility ranges for all age groups for individual local authorities are provided in the Excel spreadsheet that accompanies this report.

Figure 4.5: Comparison between mid-year population estimates and plausibility ranges, females, under 1 year old, local authorities in England and Wales, 2009



4.5 Plausibility ranges for children aged one to four years old

4.5.1 Introduction

The 'tolerance range' approach is used for this age group for both the upper and the lower limit, based on Child Benefit and Patient Register data.

4.5.2 Relationship between the sources used

Figure 4.6 shows the areas where differences between the Patient Register and Child Benefit exceed 2.5 per cent of the mid point between the sources. Patient Register counts exceed the Child Benefit counts in most local authorities. However, in some areas the Patient Register counts are lower than those based on Child Benefit. For example Richmondshire, which has a large Home Armed Forces base, has base medical facilities[3] which are believed to be used by families who may therefore not be present on the Patient Register. Similarly, an armed forces base in Wiltshire offers medical facilities to families living on the base[3] contributing to the low Patient Register counts in this local authority. However, this is not the case in all armed forces areas as arrangements vary between military establishments. Harrogate, another armed forces area, has a high number of children on the Patient Register in comparison to Child Benefit. Given that it also neighbours Richmondshire further work is needed to investigate the impact of the presence of armed forces on administrative data sources.

Other areas with a Patient Register count higher than Child Benefit are clustered around London and the South East, but also include some smaller cities such as Gloucester, Oxford, Norwich and Cambridge. Possible explanations for this could include:

- Different levels of take up of Child Benefit
- Delays in notifying changes of address for Child Benefit
- Delays in re-registering with a doctor following a move

Further research is required to identify how each of these might influence the differences observed.

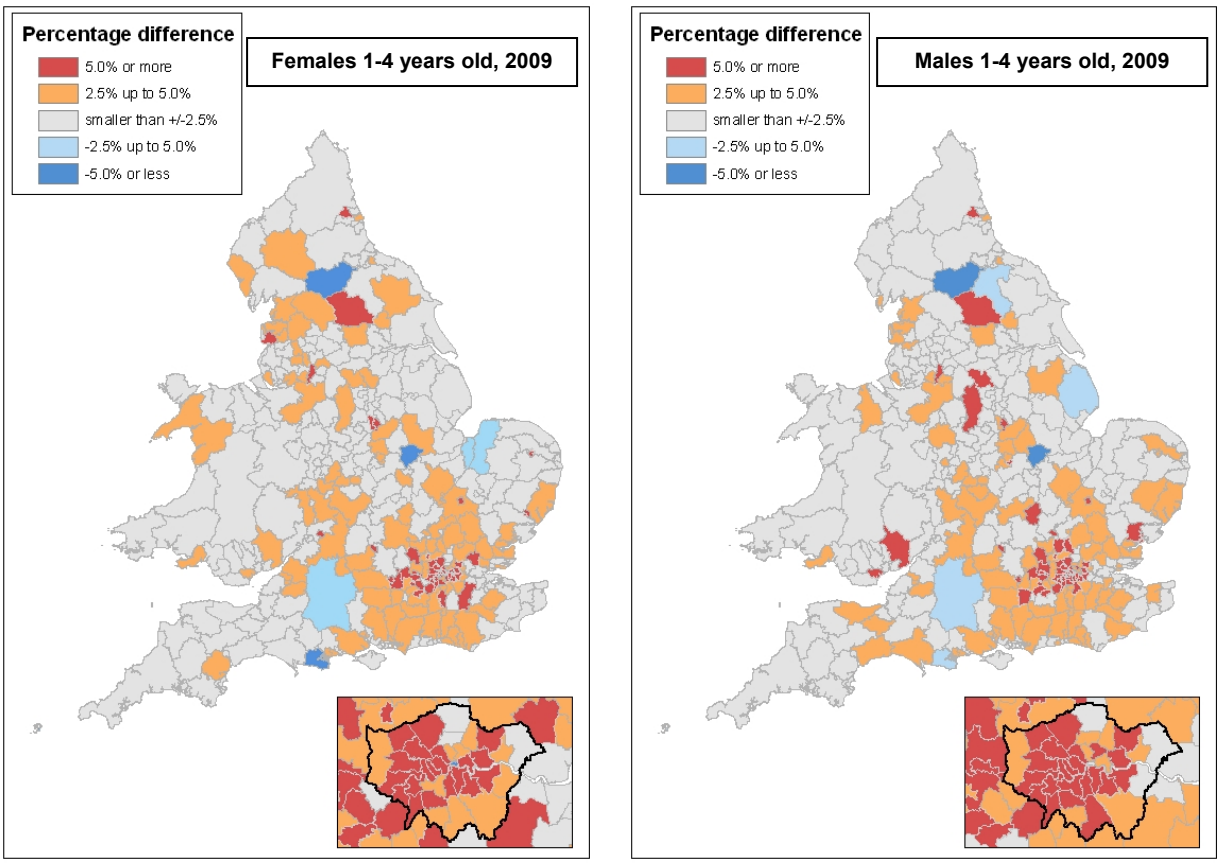
4.5.3 Plausibility ranges

Figure 4.7 shows the local authorities that have population estimates for males in this age group that are within and outside the plausibility ranges. Although potential issues have been identified with the administrative sources in home armed forces areas, these do not seem to influence the results. As well as Forest Heath and Kensington and Chelsea, estimates for East Cambridgeshire in this age group exceed the upper limit for both sexes by more than 5 per cent of the mid point. Like Forest Heath, East Cambridgeshire has an important foreign armed forces population which may have an impact on the relationship between the iMYE and the range.

The iMYE for males in this age group in Peterborough is over 5 per cent of the mid point below the lower limit. Although the estimate for females is also below the lower limit, the difference is not so extreme.

[3] Catterick and Salisbury Plain Garrison guides published by Ministry of Defence

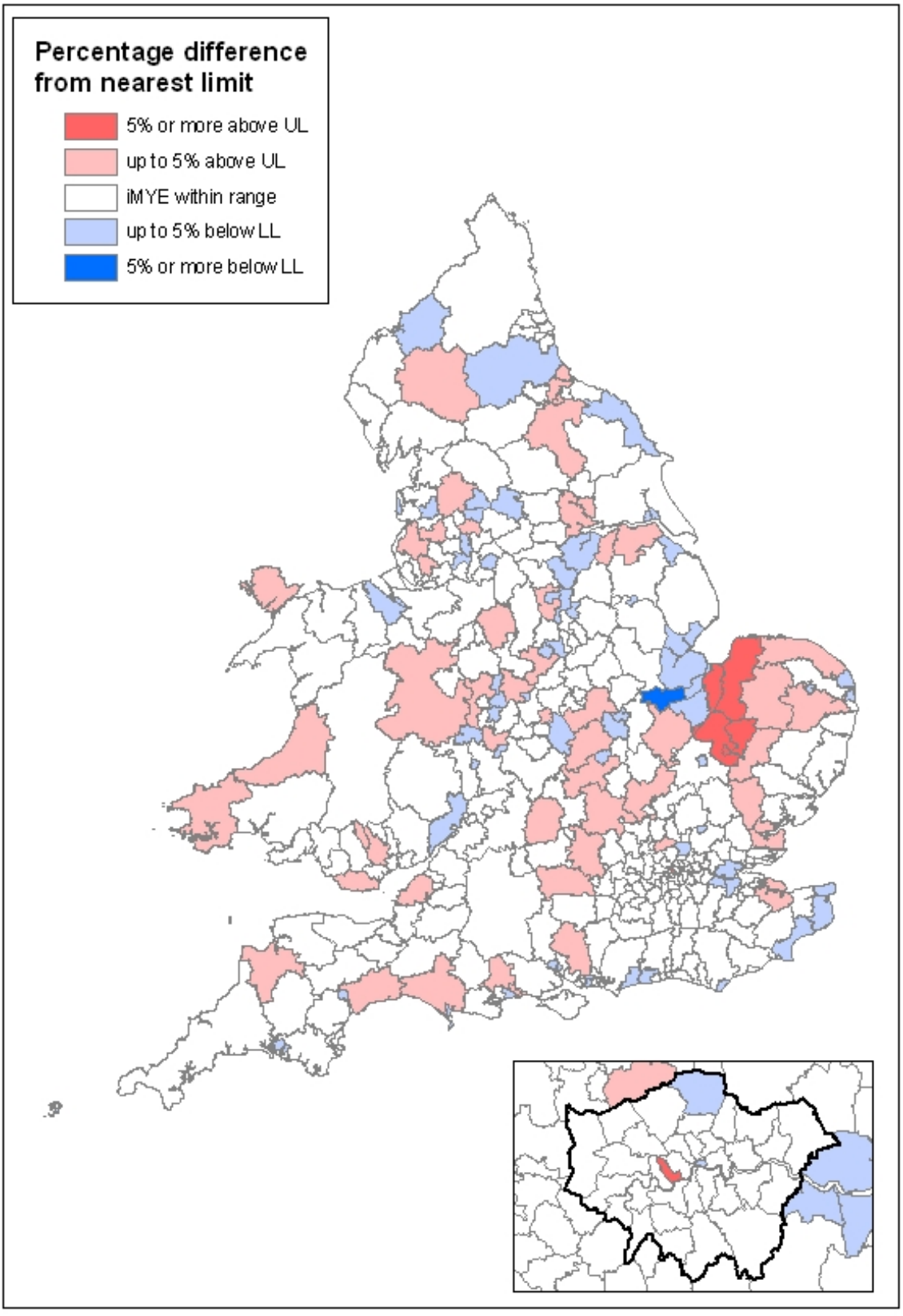
Figure 4.6: Difference between Patient Register and Child Benefit, ages 1 to 4, England and Wales, 2009



How percentage difference in these maps is calculated:

The percentage difference is calculated by subtracting the Child Benefit count from the Patient Register count at LA level. The difference is then divided by the mid-point between the two sources to give a percentage at LA level.

Figure 4.7: Comparison between mid-year population estimates and plausibility ranges, males aged 1 to 4 years, local authorities in England and Wales, 2009



4.6 Plausibility ranges for children aged 5 to 7, 8 to 11, and 12 to 15 years old

4.6.1 Introduction

The record linkage approach is used to calculate the lower limit for these age groups. The Patient Register and the English School Census have been linked as described in section 3.3.5.2, and the linked records provide the count for the lower limit. It has not been possible to calculate a lower limit for local authorities in Wales as access to the Welsh School Census at record level has only recently been available.

The 'tolerance range' approach is used to create the upper limit for these age groups based on Child Benefit and the Patient Register.

4.6.2 Relationship between the sources used

The issues discussed in section 4.5.2 regarding armed forces presence on the Patient Register are also relevant to the 5 to 15 year old age group.

Figure 4.1 shows that the gap between the School Census and the Patient Register increases with age. This is likely to be due to increased attendance at independent schools, as small step changes can be seen at ages 11 and 13 in addition to a gradual widening of the gap.

Figure 4.8 below shows the differences between Patient Register and Child Benefit for males in the age groups 5 to 7, 8 to 11 and 12 to 15. These maps demonstrate a marked increase in the differences between these sources at local authority level between the 5 to 7 age group and the 8 to 11 age group. The pattern looks different again for the 12 to 15 age group. This may be a sign that not only is there a wider gap between the Patient Register and Child Benefit for 8 to 11 and 12 to 15 year olds than for 5 to 7 year olds (see Figure 4.1) but that the gaps become more variable geographically. There are several reasons why this may be the case, and they are linked with the fact that the older a child is, the more likely they are to have moved at least once at some point during their life. Two examples of why the gap may widen are:

- The lag in re-registering with a doctor and therefore updated details being on the Patient Register may differ from the lag in notifying HMRC of a change in address as this does not impact on the continuation of payment of Child Benefit.
- There may be differences in where children in this age group are included on the administrative sources as those attending boarding schools will appear on the Child Benefit dataset at the address of the parent or guardian claiming the benefit whereas they may appear either at the parents' address or the boarding address on the Patient Register.

4.6.2 Plausibility Ranges

Table 4.1 shows the numbers of local authorities falling above, within and below the plausibility ranges for this age group. As noted above the lower limits for the 5 to 15 age group are based on the linkage between School Census and Patient Register. This gives a conservative minimum, and so estimates for few local authorities fall below them. Table 4.1 also shows that between the 5 to 7 age group and the 8 to 11 age group there was a shift from more areas falling above the upper limit to more areas falling below the lower limit.

Figures 4.9 and 4.10 show the local authorities that have population estimates for males and females that are within and outside the plausibility ranges in these age groups.

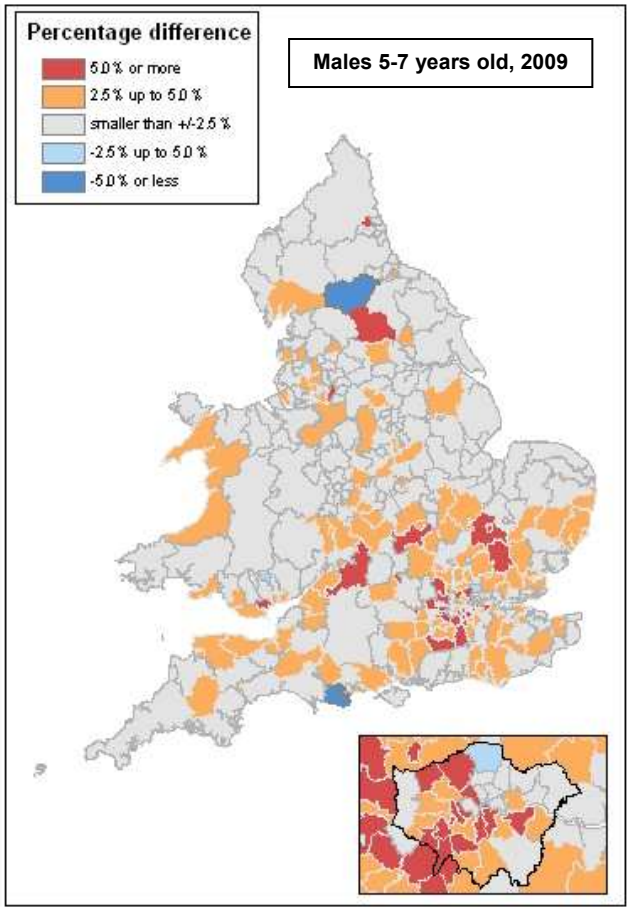
As for 1 to 4 year olds estimates for Forest Heath and Kensington and Chelsea fall above the upper limit by more than 5 per cent of the mid point for all 5 to 15 year old age/sex groups. Estimates for East Cambridgeshire also fell above the limit by this amount for most of them.

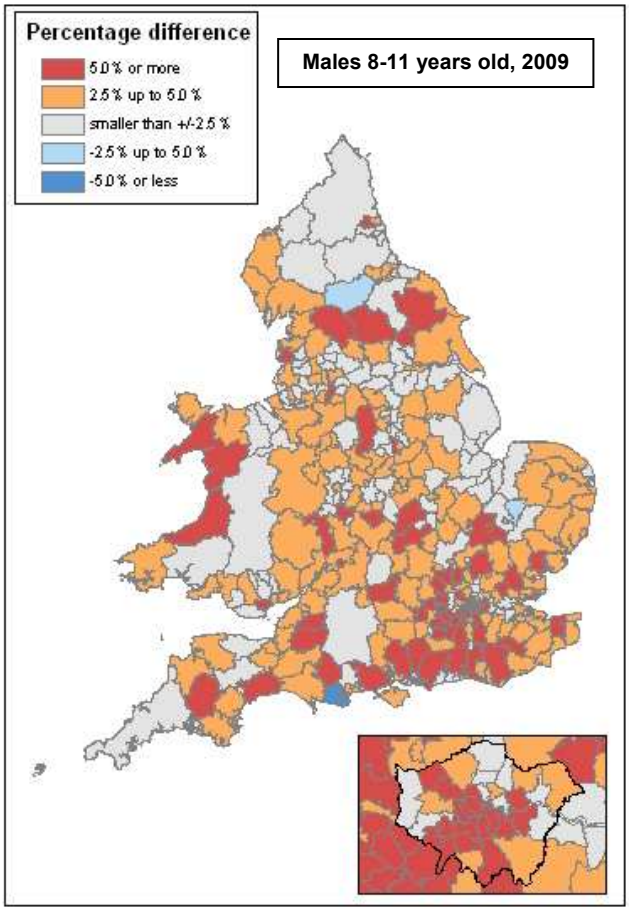
Both Leicester and Newham have estimates falling below the lower limit by more than 5 per cent of the mid point for both the 8 to 11 and 12 to 15 age group for at least one of the sexes. In addition Great Yarmouth (males), Nottingham, Reading and Slough (both) had estimates below the lower limit in the 8 to 11 age group.

Table 4.1: Number of Local Authority iMYEs that are within and outside the plausibility ranges, England only

	Males			Females		
	5 to 7	**8 to 11**	**12 to 15**	**5 to 7**	**8 to 11**	**12 to 15**
Higher	64	31	11	77	47	17
Within	261	273	296	247	259	289
Lower	1	22	19	2	20	20

Figure 4.8: Difference between Patient Register and Child Benefit, ages 5 to 15, males, 2009





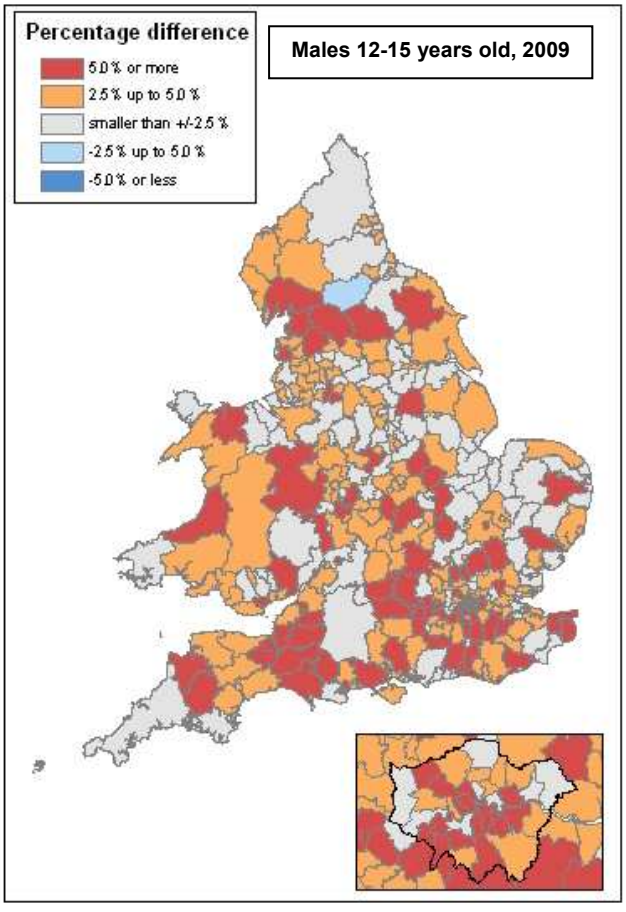
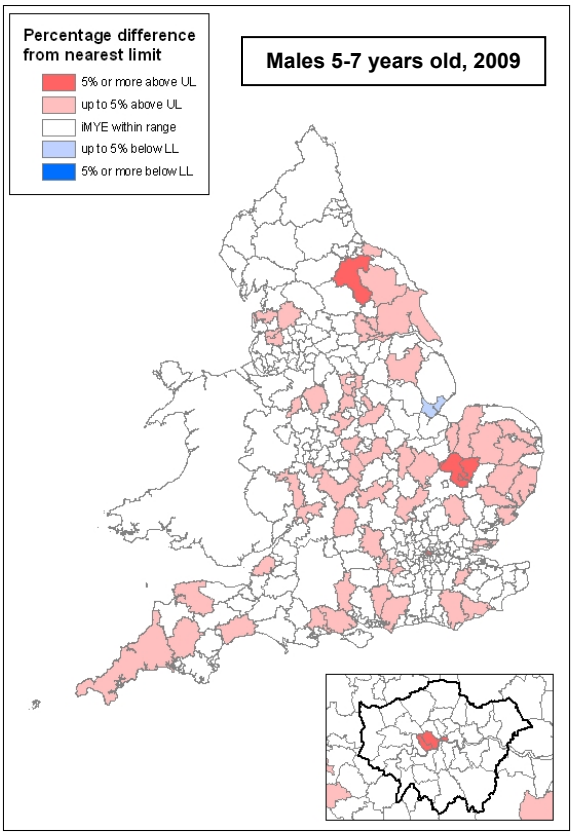
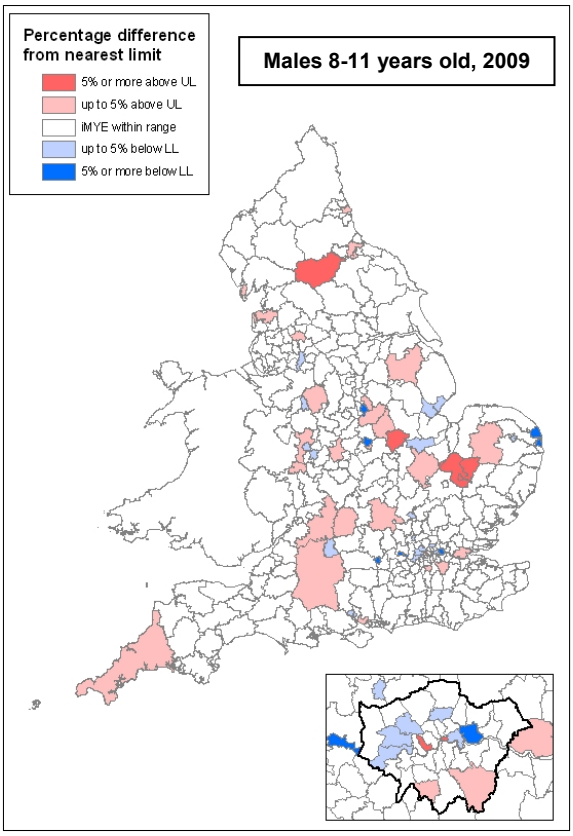


Figure 4.9: Comparison between mid-year population estimates and plausibility ranges, males aged 5 to 15 years, local authorities in England only, 2009





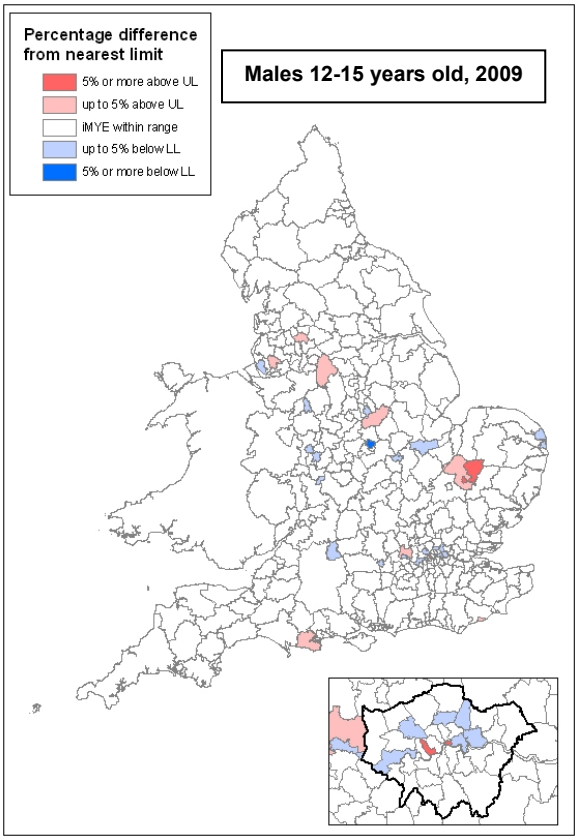
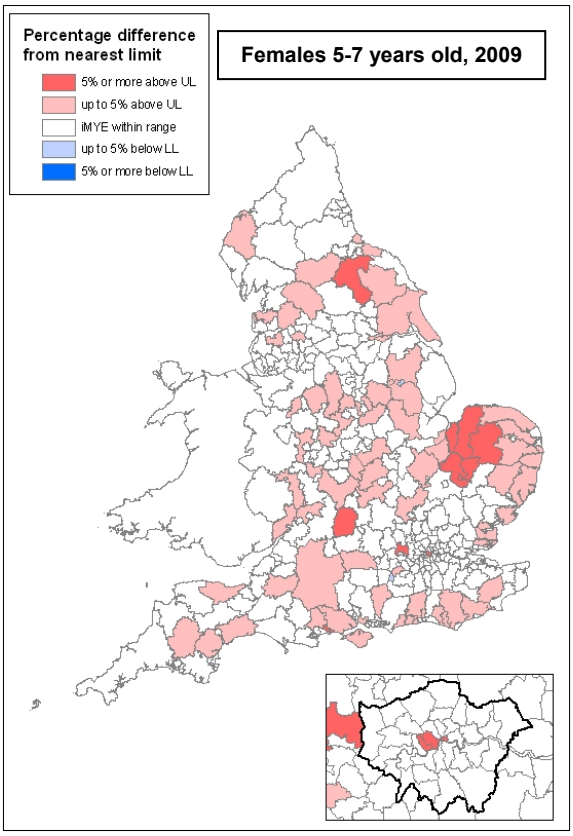
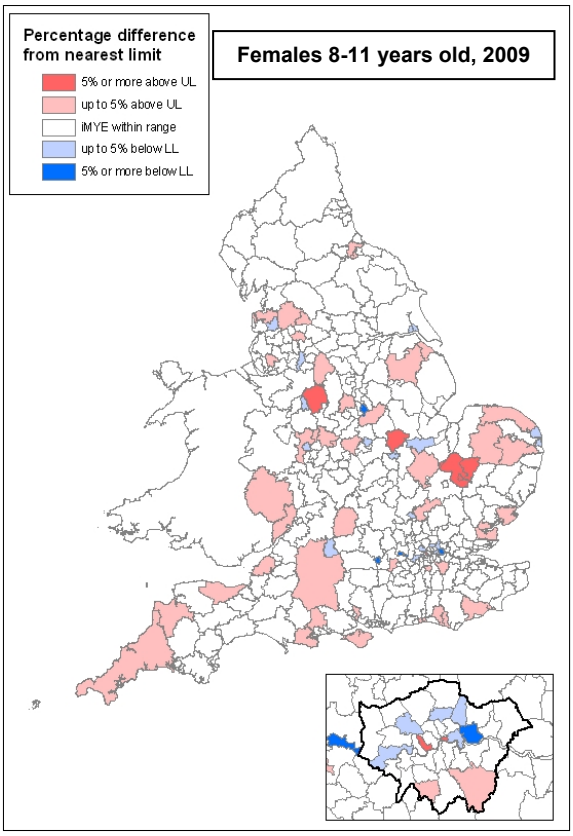
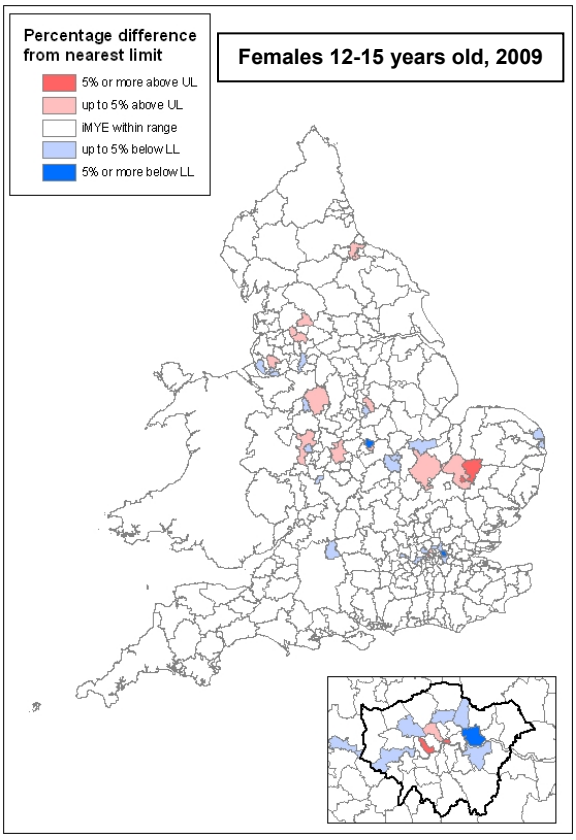


Figure 4.10: Comparison between mid-year population estimates and plausibility ranges, females aged 5 to 15 years, local authorities in England only, 2009







4.7 Across Age Groups Analysis

In order to get a broader picture of the plausibility ranges it is important to look at them across all age groups. Many of the observations made for individual age groups are driven by the relationship between the administrative data sources used for the relevant age group, for example narrow ranges for children under 1 result from good correspondence between the sources. However some local authorities fall outside the plausibility ranges due to their particular characteristics. Understanding how the administrative data sources and iMYEs interact helps to interpret and improve plausibility ranges.

Three local authorities had iMYEs higher than the upper limit for all age and sex groups; Forest Heath, East Cambridgeshire and Kensington and Chelsea. In Forest Heath and East Cambridgeshire the presence of foreign armed forces influences this as the foreign armed forces are not included in the administrative data but are accounted for in the iMYEs. Figures 4.11 and 4.12 show the plausibility ranges and iMYEs for Forest Heath and East Cambridgeshire respectively. These show that the greater presence of foreign armed forces in Forest Heath mean that the extent to which the iMYEs fall outside the range is much greater.

Figure 4.11: iMYE by age and gender plotted against the national upper and lower limits of plausibility, Forest Health 2009

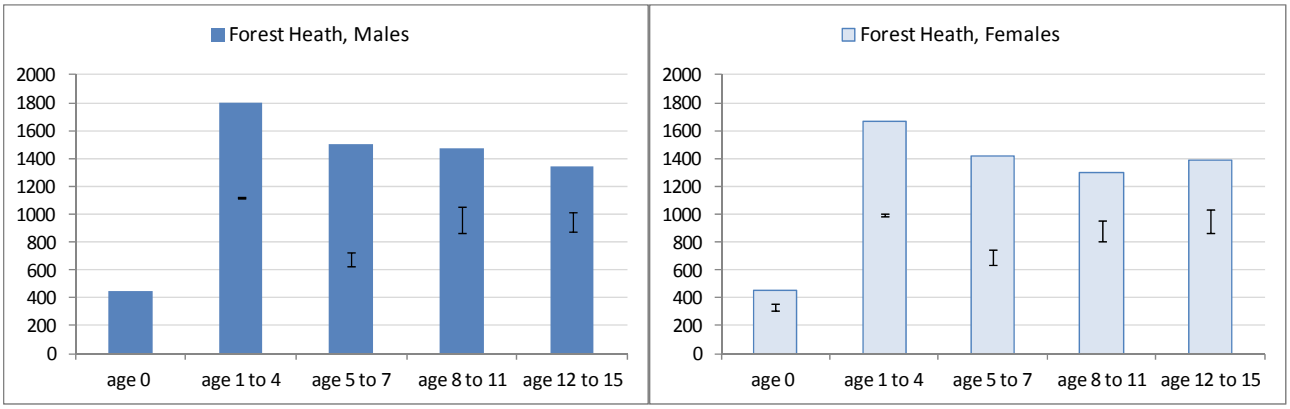
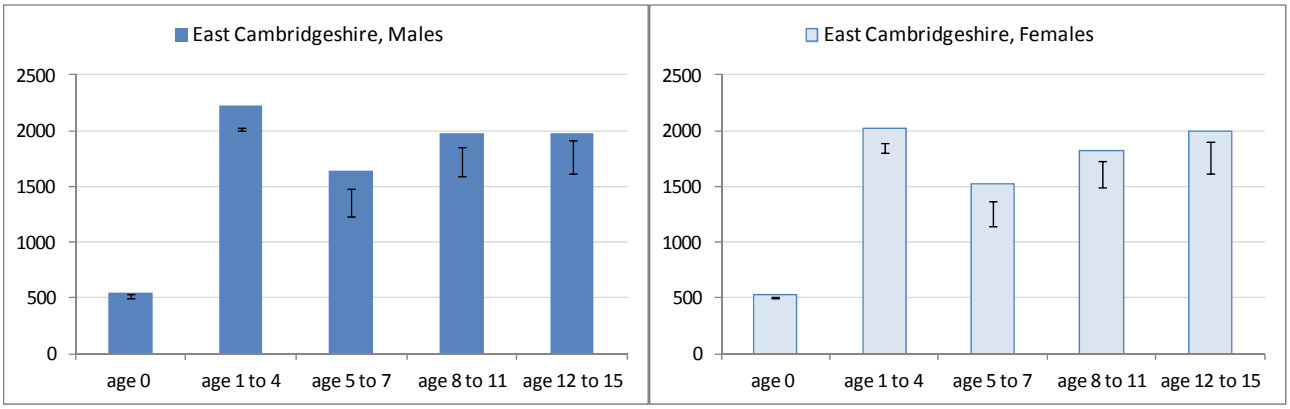


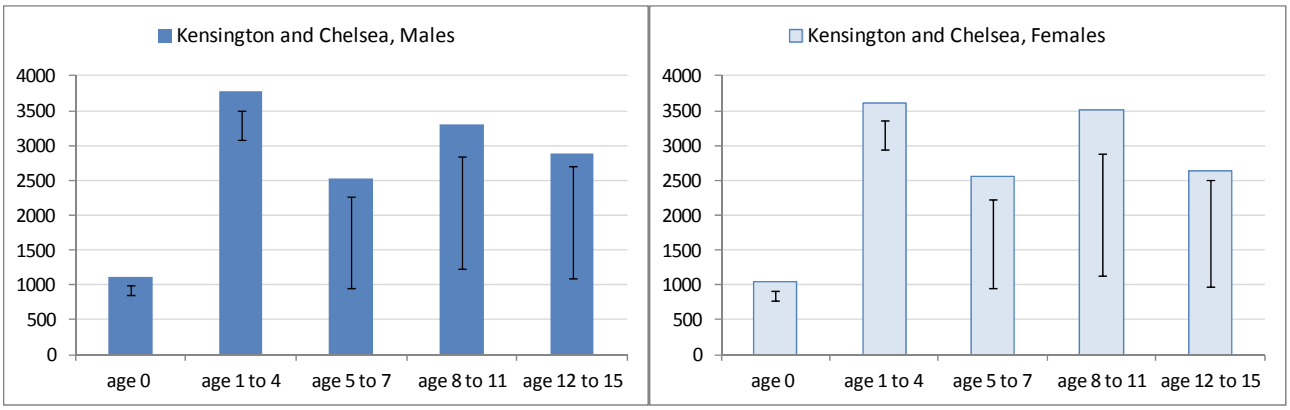
Figure 4.12: iMYE by age and gender plotted against the national upper and lower limits of plausibility, East Cambridgeshire 2009



The situation in Kensington and Chelsea is different. This local authority encompasses smaller areas ranking as low as 1,685 and as high as 26,022 on the 2010 Index of Multiple Deprivation where 1 is the most deprived area and 32,482 is the least deprived. This diversity may have an influence on the extent to which the population is included in the administrative sources. For example, the presence of foreign embassies with transient populations who may not use the

services which the administrative data sources were derived from has the effect of giving a much lower administrative count than population estimate. The result is that the plausibility ranges for Kensington and Chelsea are very wide, with the iMYEs consistently falling above them, as shown in Figure 4.13.

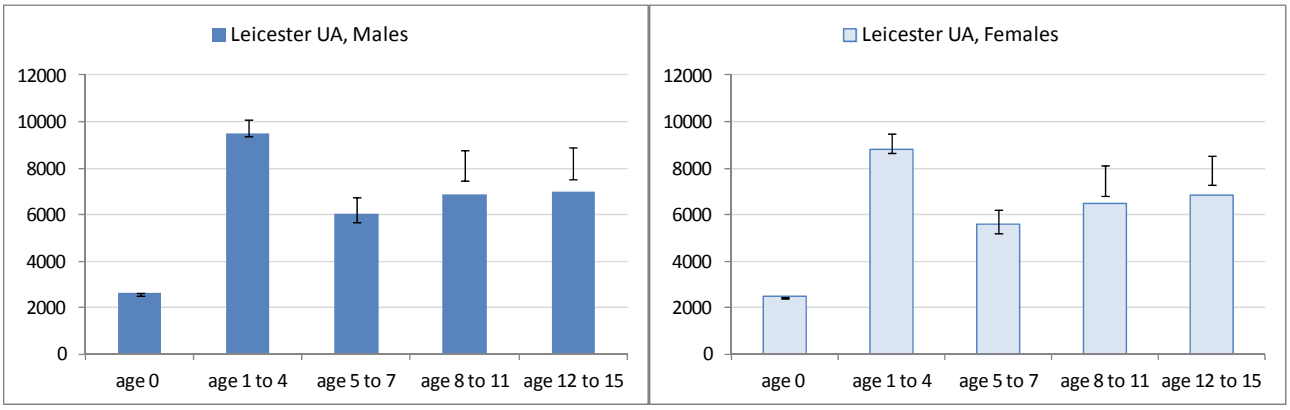
Figure 4.13: iMYE by age and gender plotted against the national upper and lower limits of plausibility, Kensington and Chelsea, 2009



These extreme examples highlight areas where it was difficult to create plausibility ranges based on currently available administrative sources that miss a part of the 'usual resident population' as defined in the iMYEs. Further research is required to investigate alternatives, or develop a methodology that could allow for these factors.

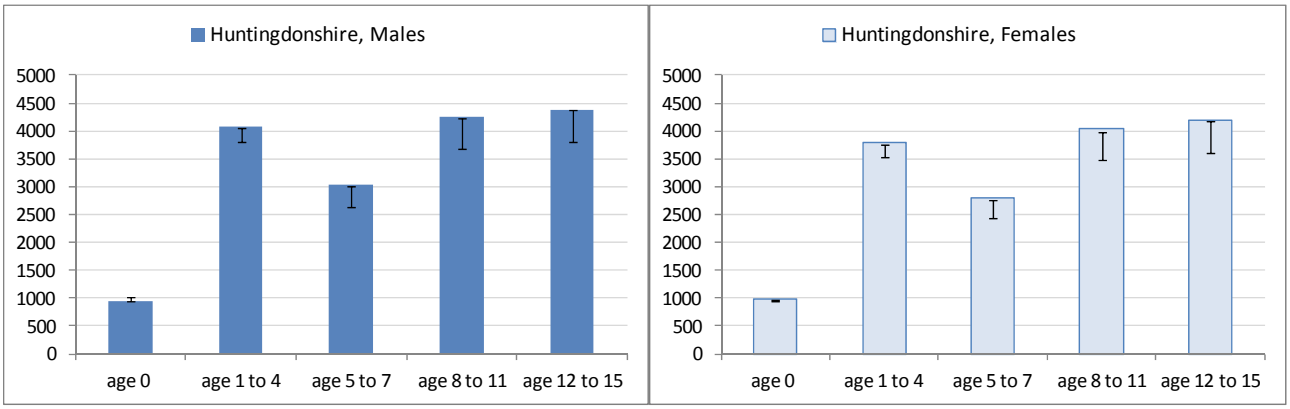
There are several local authorities that have iMYEs that fall close to or below the lower limit. Figure 4.14 showing the estimates and ranges for Leicester provides a good example of these where a low iMYE appears for children aged 8 and over. In view of the conservative nature of the lower boundary for this age group these results indicate the need for further investigation of both the iMYE and administrative sources.

Figure 4.14: iMYE by age and gender plotted against the national upper and lower limits of plausibility, Leicester, 2009



A further group of local authorities exceed the plausibility ranges for most age/sex groups. An example is Huntingdonshire, where Figure 4.15 shows that the iMYE falls inside the ranges just once across the 10 plausibility ranges by age group and sex. However, iMYEs for areas in this group are not generally far away from the limits. More detailed research into the reasons for these results is recommended.

Figure 4.15: iMYE by age and gender plotted against the national upper and lower limits of plausibility, Huntingdonshire, 2009



4.8 Summary

This section has presented a comparison of the administrative data sources at national level which shows that:

- The Patient Register provides a higher count for children aged up to 15 years than the other administrative sources and the iMYEs.
- Child Benefit counts are clearly lower than iMYEs for ages 0 and 1 due to the lag in take up, but are clearly higher for ages 8 to 15.
- The gaps between the administrative sources and the iMYEs become wider with age, particularly after age 8.

At local level the patterns are much more variable and depend on particular local circumstances. The analysis shows that in some areas with a significant armed forces presence there may be an effect on the relationship between the sources.

Plausibility ranges have been presented for five age sub-groups and for both males and females. Exploration of these has shown that:

- Large differences between the administrative sources lead to wide plausibility ranges and estimates for fewer areas falling outside the limits.
- For individual age/sex groups many areas fall within the ranges.
- In many cases where the iMYEs fall outside the ranges this is by only a small amount.
- The presence of foreign armed forces in Forest Heath and East Cambridgeshire mean that the iMYEs are higher than the upper limit in these areas.
- The diversity of the population and the presence of foreign embassies in Kensington and Chelsea lead to wide ranges for this local authority.
- There is a group of local authorities which tend to have estimates close to or below the lower limit across all age/sex groups.
- There is also a group of local authorities which tend to have estimates close to or above the upper limit across all age/sex groups.

5. Conclusions

This work demonstrates that it is possible to use administrative sources to set upper and lower limits for the population estimates in a systematic way. This report presents approaches to creating plausibility ranges for population estimates for children. The ranges have been produced using administrative data in combination to provide upper and lower limits within which the estimates might reasonably be expected to fall. Two alternative methodologies have been developed that are conceptually sound. However, the results of comparisons between the administrative sources and of the ranges with the indicative mid-year population estimates show that there are some local authorities with particular characteristics where further investigation and evaluation is required.

This research is a significant step towards being able to use administrative sources more directly in the population estimation process. Throughout the course of the project results of use to 2011 Census quality assurance, the Beyond 2011 Programme, and post-Census revisions to the 2002 to 2010 population estimates have been shared to ensure that they are taken into account. In addition, the ranges can be used as an additional quality assurance tool for the population estimates in future.

Considerable further work is required before plausibility ranges could be used to adjust population estimates, however. The most important aspects to address are outlined below.

Currently the ranges have only been calculated for 2009. The calculations should be repeated for 2010 and 2011 as the necessary data become available and evaluated for stability over time. For example to check whether areas that fall outside the ranges based on 2009 calculations also fall outside for the later years.

At present there are no independent data sources that are sufficiently well understood and of good enough quality to evaluate the plausibility ranges. However, once 2011 Census estimates become available the ranges should be compared to these in order to understand the issues in those areas falling above or below. In combination with the analysis over time this will help to determine the robustness of the ranges.

The Census should also be used to explore the administrative data sources in cases where the presence of particular populations has been identified as a problem for calculation of the ranges, for example home and foreign armed forces.

The plausibility ranges presented are only for children up to age 15. Summaries of the work to explore ranges for other age groups have been presented in annexes to this report. The 2011 Census results along with new sources should be used to help develop ranges for these other age groups.

An adjustment framework will be required in order to ensure that population estimates remain coherent. This will need to address questions such as whether adjustments should be made in such a way to constrain to national totals.

Annex A: Evaluation for the young adult population

A.1. Introduction

This section reviews the available data sources and potential for creating plausibility ranges by local authority, age and sex for people aged 18 to 24 within England and Wales. The sources for this age group are the Patient Register, the Lifetime Labour Market Database (L2) and data from the Higher Education Statistics Agency (HESA). This age group presents particular challenges as people are most likely to migrate around this age.

A.2. Sources and assumptions

Three data sources are available for this age group, the Patient Register, the Lifetime Labour Market Database (L2) and data from the Higher Education Statistics Agency (HESA).

The Patient Register includes all people registered with a GP in England and Wales. The data are supplied to ONS by NHS Connecting for Health. They are used in ONS as the basis for calculating estimates of internal migration. Record level data are available and use the NHS number as a unique identifier.

Further information regarding the quality of the Patient Register is included in sections 2.4 and 3.2.1. The quality of the Patient Register is not considered to be as good for this age group as for children partly as a result of their greater mobility.

As for the other age groups short term migrants have been removed from the Patient Register dataset prior to analysis using the methodology outlined in section 3.2.1.

The Lifetime Labour market Database (L2) is a 1 per cent sample of National Insurance numbers and activity associated with them from interactions with HMRC and DWP. L2 data are supplied at aggregate level by local authority, age and sex. More information about the L2 processing is included in section B.2 of this report.

The Higher Education Statistics Agency (HESA) is an independent body collecting information on student records and producing outputs for government, funding bodies and the public. The HESA student record microdata provided to ONS contains data on all students registered at a government funded higher education institution following a course that will lead to the award of a qualification. Whilst the information from this source is considered to be of reasonable quality, it remains unclear exactly what point in time the data refer to. Additionally, the data source only covers a part of the population aged 18 to 24 so needs to be combined with another source in order to increase its usefulness.

A.3. Method

With two record level data sources, Patient Register and HESA, as well as the aggregate L2 data, there are a variety of options available for creating plausibility ranges for this age group. These are evaluated below.

The simplest option would be to take two sources and use one as the upper limit and one as the lower. However, it does not meet the key principle of combining data sources to improve quality. This is particularly important for this age group given their mobility and the resulting high likelihood of a lag in administrative sources recording a change of details.

Options that combine the data sources include:

a) A tolerance range approach using two sources from the Patient Register, HESA and L2

b) Linking Patient Register and HESA data, and using combinations of linked and unlinked data to give upper and lower limits
c) Combinations of (a) and (b), for example using linked HESA data to adjust the Patient Register for student moves then using this in a tolerance range approach with the L2 or setting upper and lower limits as a fixed percentage around the adjusted Patient Register.

Initial analysis focused on establishing whether the quality of the data sources was sufficient to develop ranges at this stage.

Section B.5 concludes that the L2 data was of insufficient quality to be used to create plausibility ranges at local authority level for the working age group. These conclusions also apply to the 18 to 24 age group and consequently the L2 data are not evaluated further in this section.

Exploration of the quality of the Patient Register and HESA data for the 18 to 24 year old age group focuses on attempting to link the two data sources. Although both have unique identifiers, these are not common across the datasets, so a similar approach to the one used for Patient Register and the School Census explained in section 3.3.5.2 has been adopted.

The Patient Register has been linked longitudinally. At the time of matching, only one year of HESA data was available. The matching variable has been created on each dataset by concatenating date of birth, sex and postcode. The HESA dataset contains two matching variables, one based on the term time address, the other on the domicile or home address. Deduplication on the matching variable has been undertaken on both datasets. The Patient Register was matched with the HESA term time matching variable and, in the case of no match, with the home address matching variable.

An adjusted Patient Register dataset has been created by 'moving' students to where they are living according to the HESA data. This is done by setting the postcodes to be those based on the HESA term time address. This is based on the assumption that some students are slow to register with a doctor after moving to university. In those cases where a link has been made on the home address and no term time address is available the students are allocated to the postcode associated with their university campus. There are limitations with this approach as not all student accommodation is in the same local authority as the campus with which it is associated, and some students do not move away from home to attend university.

A.4. Results

The England and Wales total for 18 to 24 year olds in the Patient Register is 146,345 higher than the indicative mid-year population estimate (iMYE) total with most of the difference explained by the female Patient Register count being 140,000 higher than the iMYE.

At local authority level differences between the Patient Register and the iMYEs range from Brent where the Patient Register is more than 6,000 higher for this age group to Manchester where the Patient Register is more than 16,000 lower. The local authorities where the Patient Register is lower are predominantly university towns.

When the adjusted Patient Register data are compared with the iMYEs at local authority level the results are inconclusive. In some areas the gap narrows, for example Manchester where the difference between the adjusted Patient Register and the iMYE closes the gap by 7,000. However, a sizeable gap still remains. In Leeds a gap of over 15,000 is reduced by less than a thousand after the adjusted data are used.

One of the reasons for this variability in the results might be explained by looking at the England and Wales totals for the Patient Register before and after adjustment. This shows that the adjusted Patient Register is almost 260,000 lower than the unadjusted version. This discrepancy is

most likely to be as a result of the deduplication process, however further investigation is required to validate this.

A.5. Conclusions

This research shows that there is still a substantial amount of work to do for this age group. Whilst matching has been undertaken between the Patient Register and the HESA data to establish quality, some uncertainty remains regarding the matching process and this needs to be followed up.

As the results become available from the 2011 Census further work to evaluate both the Patient Register and the HESA data for this age group will be possible.

The absence of a reliable data source to compare with the Patient Register at local authority level for this age group is a problem. This could be addressed by following up alternative data sources, for example further education, not in employment, education or training (NEET) or the Department for Work and Pensions Customer Information System (CIS).

Annex B: Evaluation for the working age population

B.1. Introduction

This section reviews the available data sources and potential for creating plausibility ranges by local authority, age and sex for females aged between 25 and 59 and males aged between 25 and 64 within England and Wales. The sources for this age group are the Patient Register and the Lifetime Labour Market Database (L2).

B.2. Sources and assumptions

Two data sources are available for this age group, the Patient Register and the Lifetime Labour Market Database (L2).

The Patient Register includes all people registered with a GP in England and Wales. The data are supplied to ONS by NHS Connecting for Health. They are used in ONS as the basis for calculating estimates of internal migration. Record level data are available and use NHS number as a unique identifier.

Further information regarding the quality of the Patient Register is included in sections 2.4 and 3.2.1. The quality of the Patient Register is not considered to be as good for this age group as for children.

As for the other age groups short term migrants have been removed from the Patient Register dataset prior to analysis using the methodology outlined in section 3.2.1 of the main report.

The Lifetime Labour market Database (L2) is a 1 per cent sample of National Insurance numbers and activity associated with them from interactions with HMRC and DWP. These interactions typically result from periods of employment or claiming benefits and pensions. Rules can be applied around the activity information to derive a flag that marks the person holding the National Insurance record as present or absent from the UK, and thus derive estimates of the resident population. Investigation and refinement of these rules is ongoing.

UK Nationals[4] and migrants are treated separately from one another as a gap of a year in activity for UK Nationals may not indicate absence from the UK. Migrants with gaps in activity are thought more likely to be absent from the UK, for example because they may lack support networks or ties. Therefore for UK Nationals 10 week's activity either side of a 1 year gap is counted as a continuous period of residency.

For both UK Nationals and migrants, where there are 52 weeks or more activity from tax or benefits across 1 or more consecutive years then people are counted as resident in the years with activity. There are two exceptions to this, which allow for people who show no activity on the L2 but may still be resident in the UK:

- Women aged 40+ with 15 or more years past caring for a dependent child are classed as resident as they are unlikely to show activity or move abroad. A similar assumption is made for men aged 50+ with 15 years or more past caring.
- Individuals in receipt of a state pension paid in the UK are assumed to be resident.

[4] UK Nationals are defined as those whose National Insurance number (NINo) was activated at age 15 years and 9 months. This group could potentially include some people who arrived as migrants aged 15 or younger. Migrants are defined as those who have their NINo activated at a later age and have a foreign country of origin.

The "special populations" in the mid-year population estimates (armed forces, prisoners and school boarders) may or may not appear in the L2. However, they affect specific areas and do not significantly affect the majority of local authorities. For over 230 areas the contribution of the special populations to the mid-year estimate is less than 0.5% of the population in that area. Key points to note about special populations in the L2 are:

- The prisoners special population includes only those who have served at least six months of their sentence at the mid-year point. As they do not have activity during time in prison they may be absent from the L2 data (except in years coincident with starting or ending sentences).
- Home Armed Forces and their dependents should be present on the L2 but cannot readily be identified.
- Foreign Armed Forces are not on the L2 because they are not eligible for benefits and pay tax on their earnings via their home country.

People who do not interact with DWP or HMRC services are not captured by the L2 data. Examples could include:

- people who do not pay tax in the UK or National Insurance;
- people on incomes below the Lower Earnings Limit for National Insurance contributions.

L2 data are supplied at aggregate level by local authority, age and sex. UK national short term migrants are removed from L2 stocks through the whole time series. However, this is dependent on activity information being available to do this. Foreign national short term migrants are removed in all years up to 2007. For 2008 and 2009 they are not removed and assumed to be resident as activity information is not available from later periods to determine otherwise. Further work would be required to correct for this issue.

Given that the L2 is based on a one per cent sample, confidence intervals around the L2 estimates have been calculated. Confidence intervals have been derived using a similar approach to that often used for survey data. The approach to calculating confidence intervals is given in Box B.1 below.

Box B.1: Calculation of confidence intervals around L2 estimates of population

Confidence intervals are calculated using a similar approach to that often used for survey data, with the standard error (S.E.) for a given local authority, age group and gender calculated as follows.

$$S.E. = (1/\text{Sample proportion}) * n * \sqrt{[(p * (1-p))/ n]}$$

Where

n = the number of records in the 1% L2 extract before any records are removed and rules or assumptions are applied

p = the L2 estimate for a particular local authority, age group and gender divided by n

Sample proportion is 0.01 (i.e. 1%).

The standard error is multiplied by 1.96 to create 95 per cent confidence intervals.

If **c** = the estimated resident population for a given local authority, age group and gender, then:

Lower limit of the confidence interval = c – 1.96 * S.E,

Upper limit of the confidence interval = c + 1.96 * S.E.

B.3. Methods

Only the Patient Register data is available at record level for this group, so approaches must be based on combining aggregate data.

The following methods for calculating plausibility ranges are possible:

a) Using the Patient Register and L2 in a tolerance range approach such as that presented in section 3.3.1.
b) Using L2 confidence intervals as upper and lower limits.
c) Some other method of combining the sources and taking account of the confidence intervals.

However, initial analysis focuses on establishing whether the quality of the data sources is sufficient to develop ranges at this stage.

B.4. Results

The L2 sample sizes by age, sex and local authority are an important factor in determining the quality of the estimates.

In the 25 to 29 and 30 to 34 age groups over one third of local authorities have L2 sample sizes of less than 30 males and 30 females indicating that in some LA/age/sex groups there could be issues with the quality of the L2 data (Figures B.1 and B.2). For both males and females, the 40 to 44 and 45 to 49 age groups have the lowest percentage of local authorities with small samples.

Figure B.1: Percentage of local authorities with L2 sample sizes less than 30, males, by quinary age group, 2009

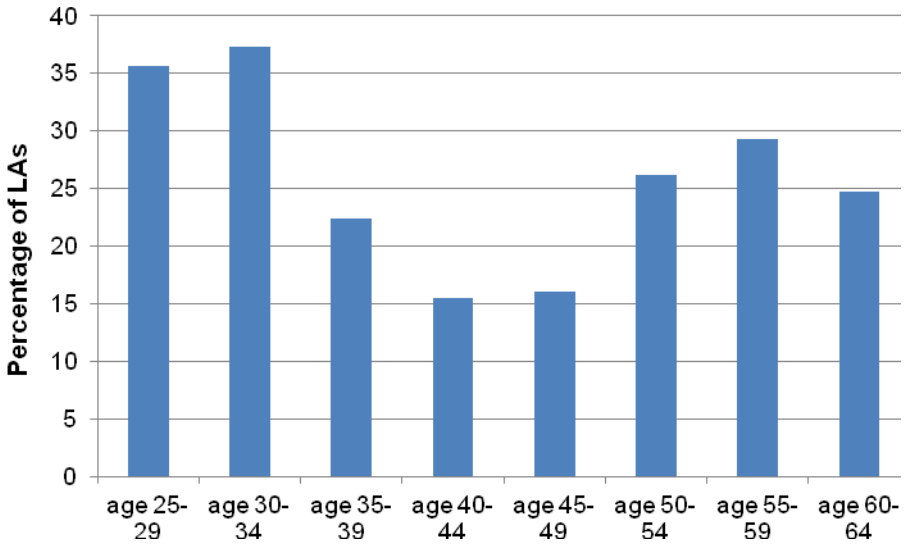
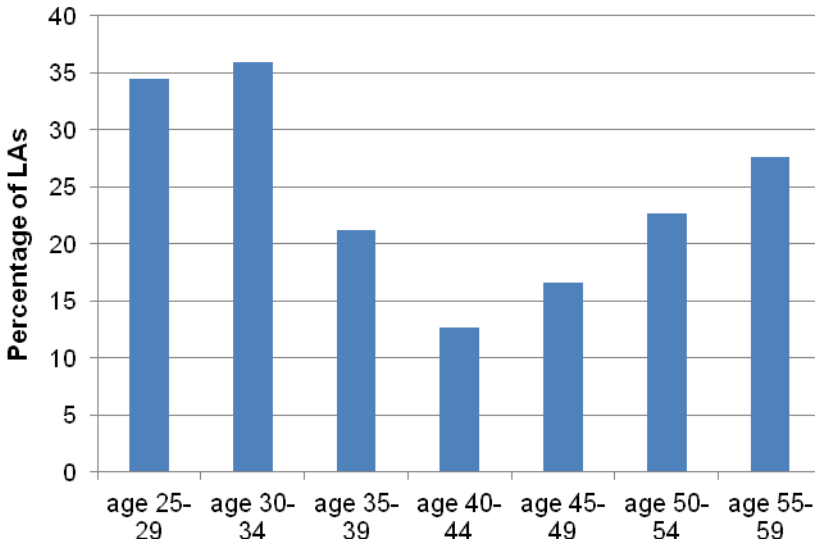


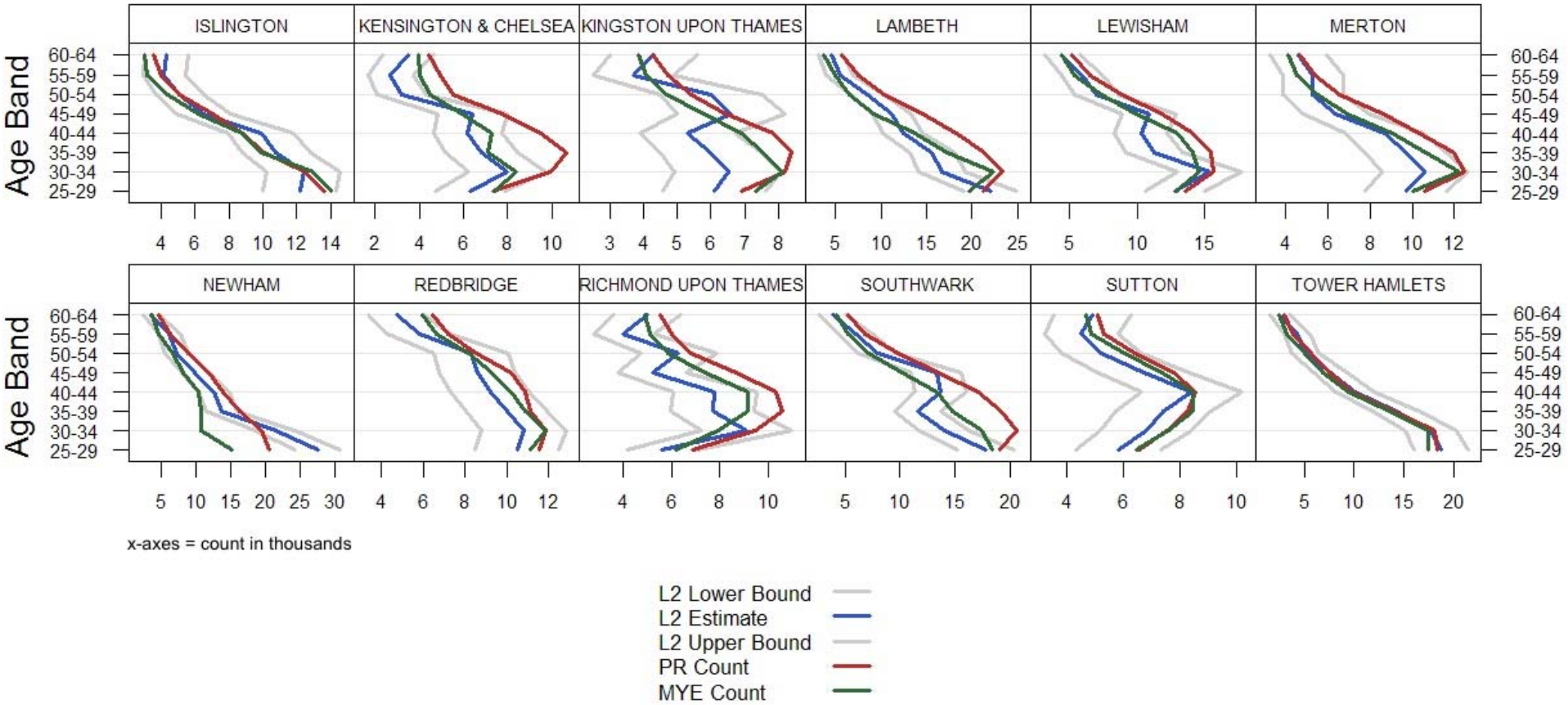
Figure B.2: Percentage of local authorities with L2 sample sizes less than 30, females, by quinary age group, 2009



Charts such as those shown in Figure B.3 help to summarise the data available by local authority, age group and sex. A selection of London local authorities is shown to highlight the diversity of relationships between the data sources. Key points from these charts are that:

- Generally the sources are better aligned for the older quinary age groups within the age range
- Local authorities with smaller populations display greater variability across the age distribution, and wider confidence intervals than those with larger populations
- For some local authorities there are some particular age groups where there are large differences observed between the sources (e.g. Newham)
- Some local authorities (e.g. Islington) show good agreement between all three data sources
- There are limitations with both the Patient Register and L2 data.

Figure B.3


ISLINGTON
KENSINGTON & CHELSEA
KINGSTON UPON THAMES
LAMBETH
LEWISHAM
MERTON
NEWHAM
REDBRIDGE
RICHMOND UPON THAMES
SOUTHWARK
SUTTON
TOWER HAMLETS
Age Band
60-64
55-59
50-54
45-49
40-44
35-39
30-34
25-29
x-axes = count in thousands
L2 Lower Bound
L2 Estimate
L2 Upper Bound
PR Count
MYE Count

Comparison of the indicative mid-year population estimate (iMYE) to the L2 shows that for younger quinary age groups the range of differences observed is wider than for older quinary age groups (Figures B.4 and B.5). Also more areas display an iMYE higher than the L2 in the older quinary age groups than in the younger ones.

Figure B.4: Number of local authorities by absolute difference between MYE and L2, males, 25 - 29

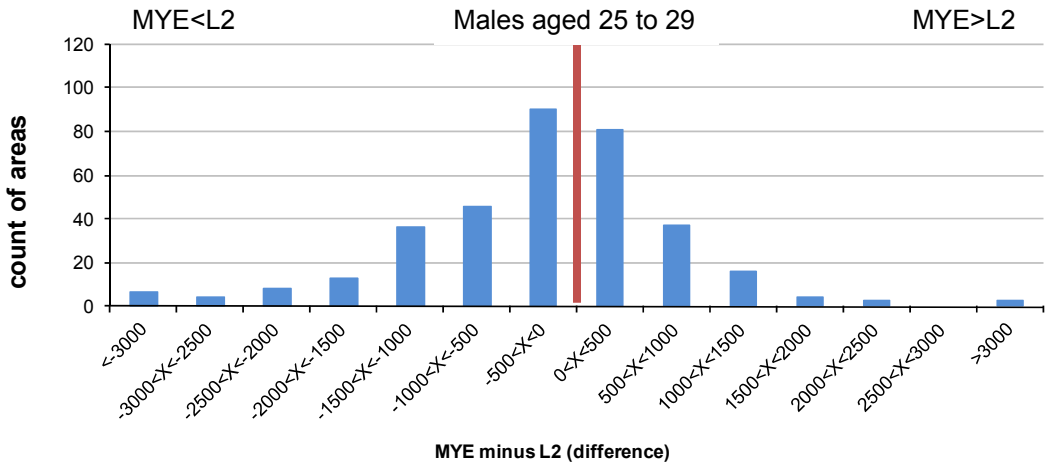
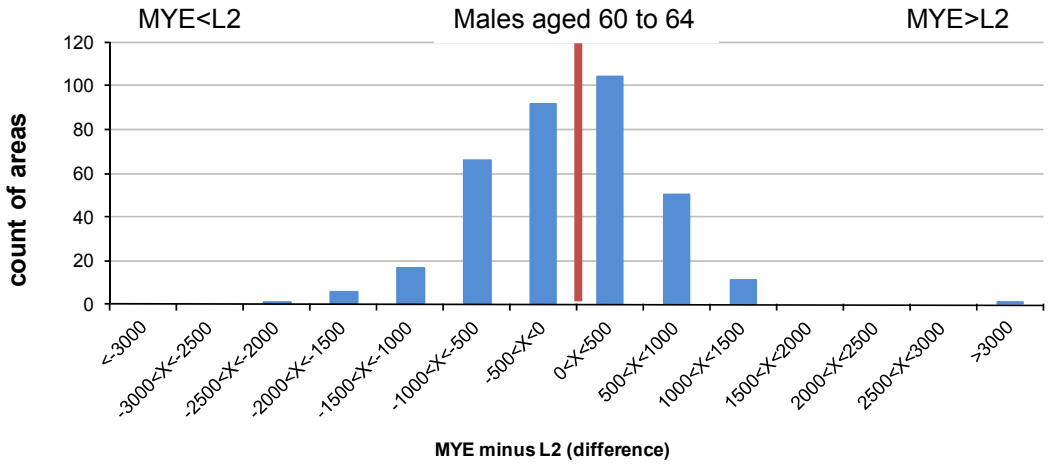


Figure B.5: Number of local authorities by absolute difference between MYE and L2, males, 60 - 64



5. Conclusions

There are limited data sources available for this very wide age group. For both sources there are concerns regarding their quality, particularly in relation to coverage and the quality of location information.

The results demonstrate that there are a number of different issues with using the L2 data for creating plausibility ranges.

- At local authority and quinary age group level the sample size is not good enough.
- The inclusion of short term migrants in the latest available tax year means that further work is required
- The quality of the L2 address information remains unclear.

As a result it is important to understand how these sources compare to the 2011 Census before the work is progressed further. When record level Customer Information System (CIS) data from

the Department for Work and Pensions becomes available to ONS further options will be available to help understand these issues in more detail including record linkage between the CIS and the Patient Register. This would increase the range of alternative approaches available for this age group.

Annex C: Evaluation for the over-retirement population

C.1. Introduction

This section reviews the available data sources and potential for creating plausibility ranges by local authority, age and sex for males aged over 60 and females aged over 65 within England and Wales. The sources considered for this age group are the Patient Register and the Work and Pensions Longitudinal Study.

C.2. Sources and assumptions

Two data sources are available for this age group, the Patient Register and the Work and Pensions Longitudinal Study (WPLS).

The Patient Register includes all people registered with a GP in England and Wales. The data are supplied to ONS by NHS Connecting for Health. They are used in ONS as the basis for calculating estimates of internal migration. Record level data are available and use NHS number as a unique identifier.

Further information regarding the quality of the Patient Register is included in sections 2.4 and 3.2.1. The quality of the Patient Register is not considered to be as good for this age group as for children.

As for the other age groups short term migrants have been removed from the Patient Register dataset prior to analysis using the methodology outlined in section 3.2.1 of the main report.

The WPLS is administered by the Department for Work and Pensions (DWP). The Pensioner Client Group includes all DWP clients of state pension age, regardless of the benefit or programme they interact with. It therefore includes all those claiming state pensions and other pension benefits such as pension credit and attendance allowance.

C.3. Methods

Only the Patient Register data is available at record level for this group, so approaches must be based on combining aggregate data.

Plausibility ranges could be based on the Patient Register and Pensioner Client Group in a tolerance range approach such as that presented in section 3.3.1. For this age group alternatives are limited to ranges that do not use the data sources in combination.

However, initial analysis focuses on establishing whether the quality of the data sources is sufficient to develop ranges at this stage.

C.4. Results

Figures C.1 and C.2 show comparisons of the indicative mid-year population estimates (iMYEs), the Patient Register and the Pensioner Client Group (WPLS) by age group and sex. The age groups at the extremes show the largest differences between the sources.

For females aged 60 to 64 and males aged 65 to 69 there are large differences between the sources which may be caused by the level of pension take up in these age groups.

The Patient Register is the highest source for most age groups. Exceptions are females aged 80 to 84, males aged 85 to 89 and all people over 90. The WPLS is the lowest source for most age groups. Exceptions are males aged 70 to 74 and 75 to 79 and females aged 85 to 89.

The scales for these charts disguise large relative differences between the iMYE and the administrative sources for the over 90 age group. In total the iMYE is over 29 thousand higher than the Patient Register with this difference split equally between males and females, despite the number of females in this age group being more than double the number of males. The iMYE is also over 25 thousand higher than the WPLS in the age group, with the difference for males being almost double that for females.

Figure C.1: Comparison of iMYE, Patient Register and Pensioner Client Group, males, England and Wales, 2009

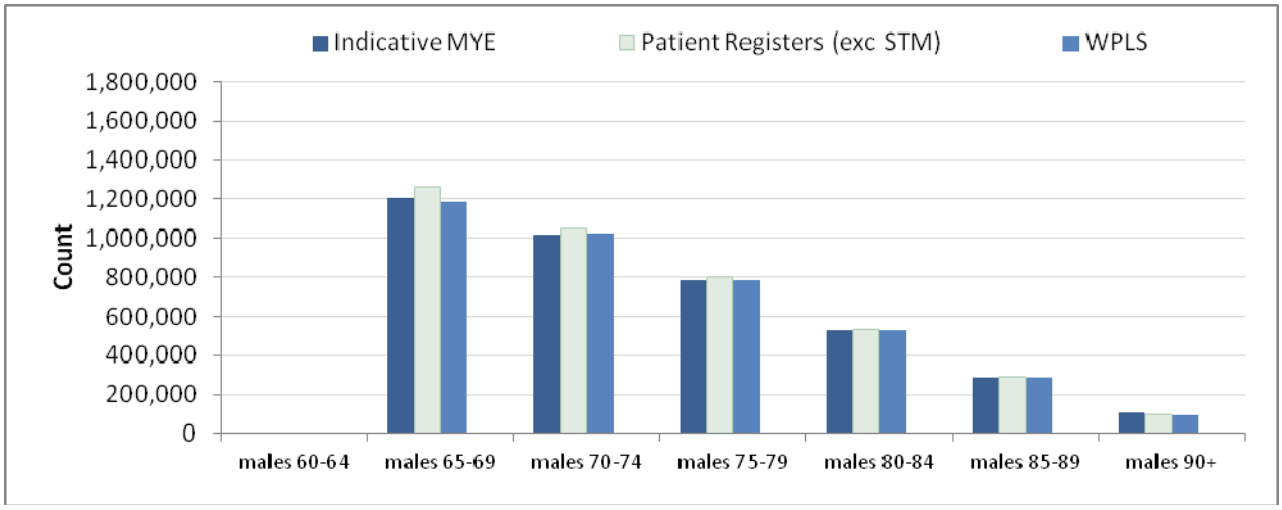
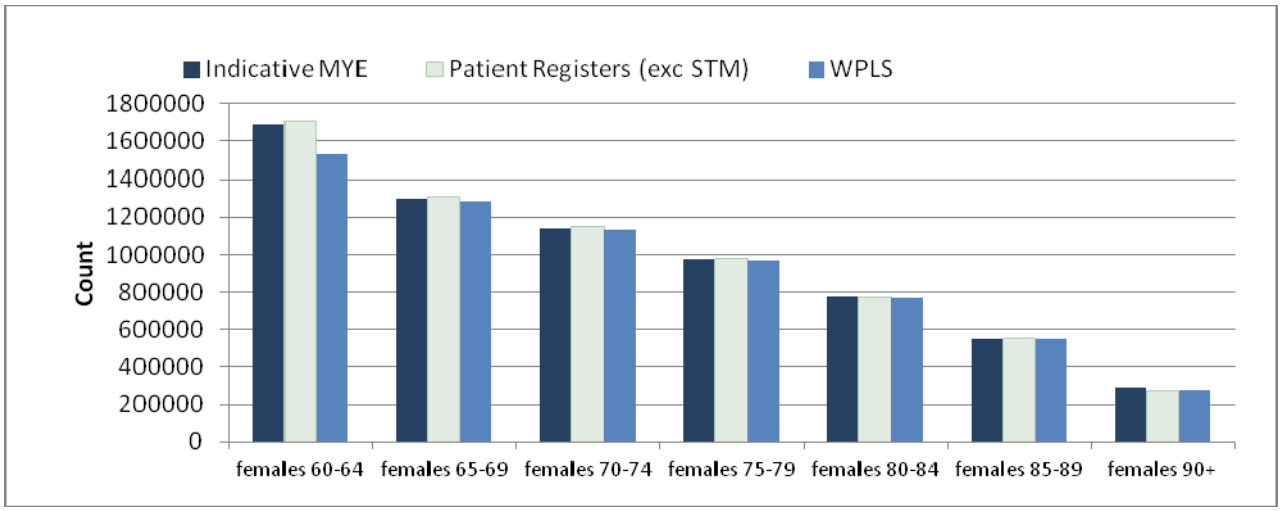


Figure C.2: Comparison of iMYE, Patient Register and Pensioner Client Group, females, England and Wales, 2009



C.5. Conclusions

Although the analysis shows differences between the sources for people over retirement age, these are relatively small and it is likely that the tolerance range approach would produce narrow ranges. As a result it is also likely that many areas would fall outside the ranges. Upper and lower limits should be calculated using the tolerance range approach set out in section 3.3.1 and a comparison made with 2011 Census results when they become available. Further research should be undertaken to investigate the apparently high iMYE for people aged over 90. It may also be appropriate for the 60 to 69 age group to be included with the working age group to take advantage of a wider range of administrative sources and reflect forthcoming increases to the state pension age.

Annex D: Feedback form

Plausibility ranges for population estimates – Feedback Form

ONS is seeking feedback regarding the plausibility ranges from users of the Local Authority (LA) mid-year estimates to assist with their validation and further development. In particular ONS is interested in how the ranges compare to local knowledge, whether from administrative data or anecdotal evidence.

Please complete the following questionnaire and return to imps@ons.gsi.gov.uk as soon as convenient but no later than **30 June 2012.**

Do the administrative sources appear to be in-line with your knowledge of them? (Please see line charts provided, which show 3 administrative sources and the mid-year estimates).
Do you have any thoughts on the approaches we have used? For example - do you have alternative suggestions for combining data? - do you have any comments on the data used, or further refinements that could be made? - are there other data sources which you think could be used in creating plausibility ranges?

What are your thoughts on the plausibility ranges for your LA (s)? - are they wide, are they narrow? - do you have any comments on the position of the indicative mid-year estimates relative to the plausibility ranges?
Do you have any other comments on the plausibility ranges?

Thank you very much for your help.

Annex E: References

Daas, P., Ossen, S., Vis-Visschers, R. and Arends-Toth, J. (2009) Checklist for the Quality evaluation of Administrative Data Sources, Statistics Netherlands, the Hague/Heerlen.

Bernardi, A., Cerroni, F. and De Giorgi, V. (2010) Analysis on Economic Fiscal Data for Statistical uses, Seminar on Using Administrative Data in the Production of Business Statistics , Rome, 18th-19th March 2010.

Hucks L & Axwesco F (2011) WaterAid in Tanzania, East Africa data reconciliation workshop, WaterAid, WHO, UNICEF.

Johnson, B. and Moore K. (2005) Consider the Source: Differences in Estimates of Income and Wealth From Survey and Tax Data, Online at: http://www.irs.gov/pub/irs-soi/johnsmoore.pdf

Longva, S., Thomsen, I. and Severeide, P.I. (1998) Reducing Costs of Censuses in Norway Through Use of Administrative Registers, *International Statistical Review*, 66 (2): 223-234

Ruotsalainen K. (2005) Combining enterprise data to employment data in register-based employment statistics. Paper for the Sienna Group on Social Statistics , Helsinki

Thomsen, I. and Kleive Holmoy, A.M. (1998) Combining Data from Surveys and Administrative Record Systems. The Norwegian Experience, *International Statistical Review*, 66 (2): 201-221

Tikabo, M. (2011) Setting tolerances for age-sex check of Census 2011 data. Office for National Statistics (internal paper)

Zauberman, R., Robert, P., Nevanen, S. and Didier E. (2009) L'acteur et la mesure, Le comptage de la delinquance entre donnees administrative et enquetes. *Revue francaise de sociology*, 50 (1):31-62