

Demonstrating fitness for purpose of 2021 Census in England & Wales in the context of a transformed social statistics system



Jon Wroth-Smith

2021 Census Statistical Design

Fitness for Purpose

- Context
 - ‘Getting it right matters’
- Experience from Census 2011
 - Building on
- Looking forward to Census 2021
 - Legacy by design
- Fitness for purpose in a transformed social statistics system
 - Taking forward design

Context

*The UK's decennial population census is central to decisions in all areas of society – whether by businesses, councils, the health service or charities. It is the basis of population estimates; it underpins funding formulae; it provides insight into the wellbeing and needs of communities throughout UK society. It is also the most expensive single statistical undertaking – **getting it right matters, both in terms of providing value for this spending of public money, as well as in ensuring that users base these important decisions on trusted and high quality data.***

UK Statistics Authority (2015)

Context

Our aims for the 2021 Census are:

- to produce high-quality results
- to generate outputs that meet the needs of users
- to maximise online response rates for the census
- to produce timely outputs to maximise benefits
- to protect, and be seen to protect, confidential information
- to conduct it in a cost-effective way
- **to make recommendations for the approach to future censuses in England and Wales**



Context

- Regardless of how census-type statistics are produced, the UK Statistics Authority statements apply:
 - ‘Getting it right matters’
 - ‘Ensuring that users base these important decisions on trusted and high quality data’
- Administrative data is shared with ONS on the basis that statistics are produced in the public interest
- Digital Economy Act Principles and Code of Practice - Statistics
- Census 2021 will provide a legacy for ONS to demonstrate fitness for purpose using a transformed social statistics system

Experience from 2011

- To recognise the areas of success and to repeat them is important and understanding and appreciating the areas of improvement and learning from them is vital. Users were very positive about the overall conduct and quality of the data from each of the 2011 Censuses.
- The scale of this achievement contrasts with the dissatisfaction which accompanied the 2001 Census in England and Wales in some local areas.
- The success, particularly for ONS, reflects the deep reflection and learning it undertook after that census and throughout the development of the 2011 Census.

UK Statistics Authority (2015)

Experience from 2011 – Partnership

- Worked with a small number of local authority users to understand what validation analysis would give them confidence
- Identified administrative data of potential value
- Built this into validation plans
- Requested data from **all** local authorities where required
- Roadshows to explain approach for users

Figure 6 – Electoral registration data for counts of households by postcode

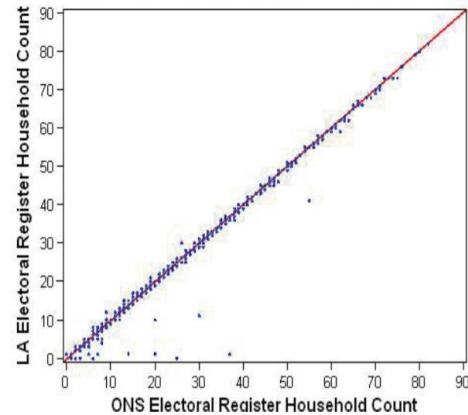


Figure 7 – Electoral registration data for counts of electors by postcode

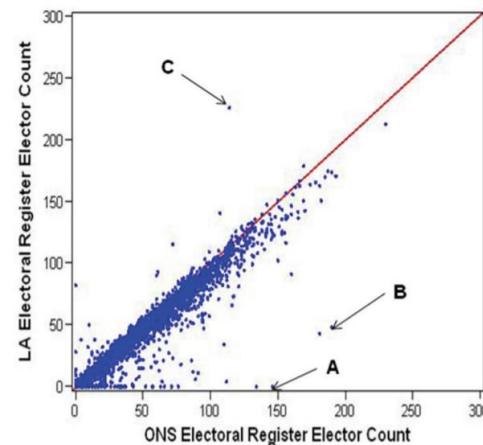
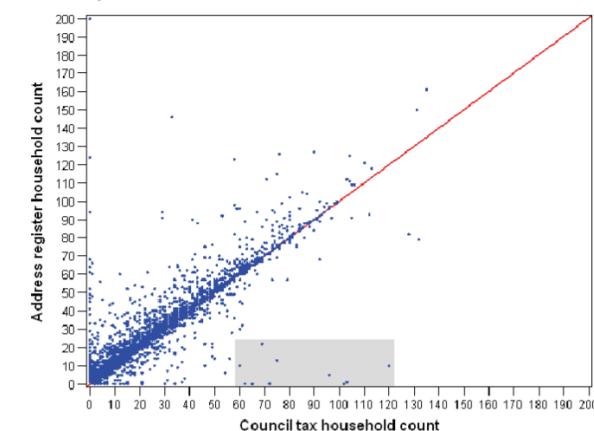
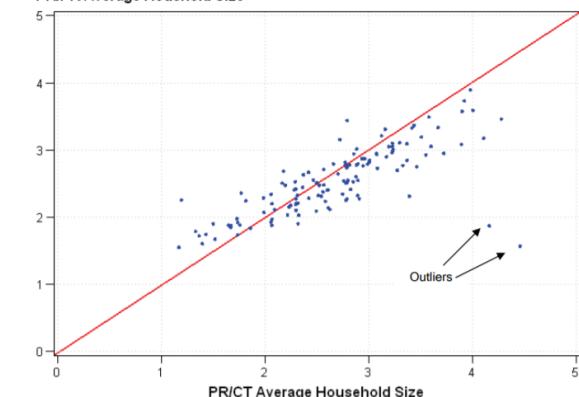


Figure 2 – Comparison of council tax and address register household counts for example LA



PR/PR Average Household Size



Experience from 2011 - Coherence

- Publication of a range of material alongside census estimates: confidence intervals, bias adjustments, response rates and validation/QA pack for each local area

Please allow a moment for the pack to update when selecting geographies

Select Region → Select Region
Select Local Authority → Select LA

2011 Census Quality Assurance Pack

Please Select a Local Authority

Help tomorrow take shape

Introduction
This Quality Assurance Pack is part of the first release from the 2011 Census of population for England and Wales conducted by the Office for National Statistics (ONS). It provides a snapshot of the usually resident population as at Census Day, 27 March 2011 and various comparators used in the quality assurance process. This is a summary version of the information that was made available to the Quality Assurance Panel during quality assurance of the census estimates.

Information is provided on the usually resident population at national, regional, county and constituent Local Authority District and Unitary Authority level, by age and sex. The census population estimates are provided by five year age band below national level. Comparisons within the pack are intended to the closest. There may be inconsistencies with comparator data published elsewhere because of rounding, disclosure control and processing to ensure consistency with census definitions.

This information pack is provided in accordance with the provisions of the Census Act 1920, and therefore only relates to the census in England and Wales

By selecting a local authority at the top of this page the pack provides an area specific summary of the material outlined on the contents page. Separate tables for all local authorities and regions also published as part of the first release. The tables can be

To be able to use the quality assurance pack please ensure that macros are enabled

Microsoft instructions for enabling macros in Excel 2003

Local Authority QA Pack:

Explanatory materials

Local Authority totals

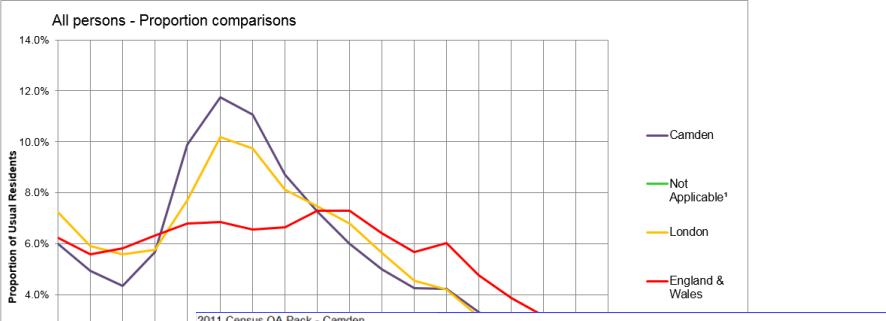
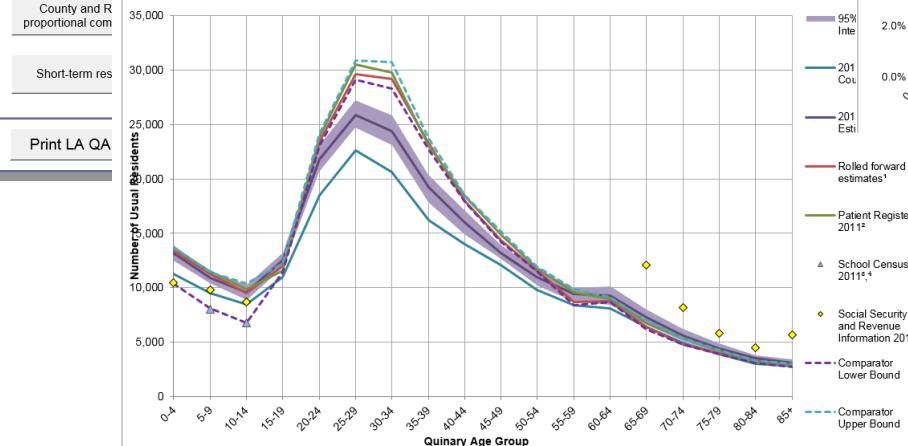
Age-sex population distribution

Above LA Geography:

County age-sex population distribution

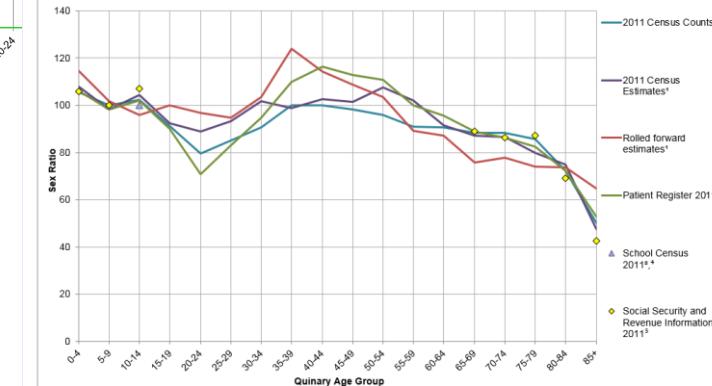
County age-sex distributions are kept in a separate sheet found at the link above. The counties sheet has its own print button and will not be printed when selecting the print button on this page. As counties are not a standard depository for either the estimation

All persons - Camden



2011 Census QA Pack - Camden

Sex ratios - Camden



Experience from 2011 – Transparency

- Ability to explain complex statistical techniques in a simple, transparent way so users understand difference between counts/estimates and the validation work undertaken
- 'Trout, Catfish, Roach' beginner's guide and 'completing the count' events

Trout, Catfish and Roach is a beginner's guide to understanding population estimates and the role of the Census Coverage Survey. By using the analogy of fishing, this booklet takes readers through a series of easy equations to explain the way we reach our census population estimates. A worked example can be found at the end of this guide.

Step one: counting trout



How do you find out how many trout there are in your pond?

You could drain it and count the fish I suppose, but it wouldn't do them much good. Perhaps if the pond was small you could try and catch them all. That would take a while. But there's nothing like fly fishing, so you decide to give it a go.

At the end of day one, you've caught 100 trout. Pretty good!

So how many are there in total? Still no idea, really. So you tag each one, put them all back (you've lovingly cared for them all of course), and carry on the next day.

The next day you manage to catch another 50; 25 with a tag, 25 without. So you've found another 25 and know for sure there are at least 125. Pretty good. That'll have to do.

THE TUB CENSUS

A yellow cylindrical container filled with various colored sweets (yellow, red, blue, green). Some sweets have small purple stickers on them.

Marked (Census)
Total 425
Colour 1 (Green) 223
Colour 2 (Purple) 202

QUESTIONS:

- How many sweets are in the tub?
- How many of each colour are in the tub?

Help tomorrow take shape © Office for National Statistics

Experience from 2011 – Independent Assurance

- Users may not understand technical detail or have time to go into all elements of statistical design
- Independent assurance of methods ahead of census provided by **panel of academic experts**
- Expertise drawn from difference fields
- Published all papers put to panel and final report with recommendations
- ONS published response

'We are reasonably optimistic that, having taken account of our recommendations to develop, document and consult on specific aspects of methodology, the 2011 Census in England and Wales will provide population estimates that can guide resource allocation and social policy in the right direction for the next ten years' (Plewis et al, 2010)

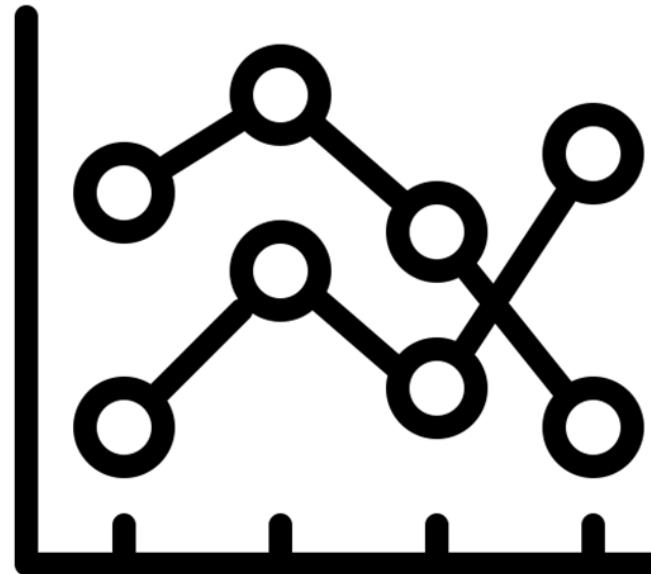
Looking forward to 2021

High level design goals and principles:

- Results that reflect the country we live in today by meeting quality targets which are as in 2011:
 - nationally accurate as measured by a confidence interval of +/-0.2%, with bias less than 0.5% for England and Wales
 - high-quality locally with 95% confidence intervals for all local authorities of +/-3%
 - minimal variation within local authority area
 - response rate targets of 94% nationally and 80% locally in all local authorities, to support these quality levels
- Outputs that are timely and easy to use (first statistics published within a year with detailed statistics published within two years)

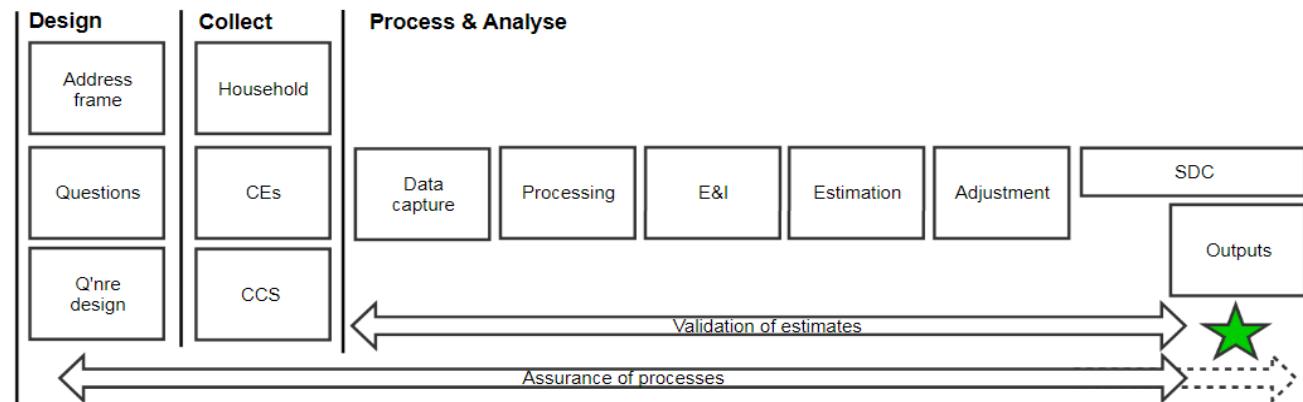
Looking forward to 2021

Clarity - Quality Assurance and Data Validation



Looking forward to 2021 – Quality Assurance

- Revisited the Generalised Statistical Business Process Model (GSBPM 5.1, Section 8.2)
 - Quality is a complex and multi-faceted concept with several quality dimensions.
 - Quality management production of statistics in an institution is normally set within an overarching framework
 - A fundamental role in quality management is played by the set of quality control actions should be implemented within the sub-processes to prevent and monitor errors and sources of risk



Looking forward to 2021 – Quality Assurance

- ‘Impact Maps’ provide a framework for the QA of each Census process
- Identify potential quality issues, diagnostics, mitigations and contingencies

Pages / ... / Process and Analyse Reports 

 Edit  Save for later  Watch  Share ...

Resolve Multiple Records (RMR) - Process - Impact Map

Created by Alexandra Turda; last modified by Hayleigh Hibbs on Jul 15, 2019

Resolve multiple records for the same address. Structure data

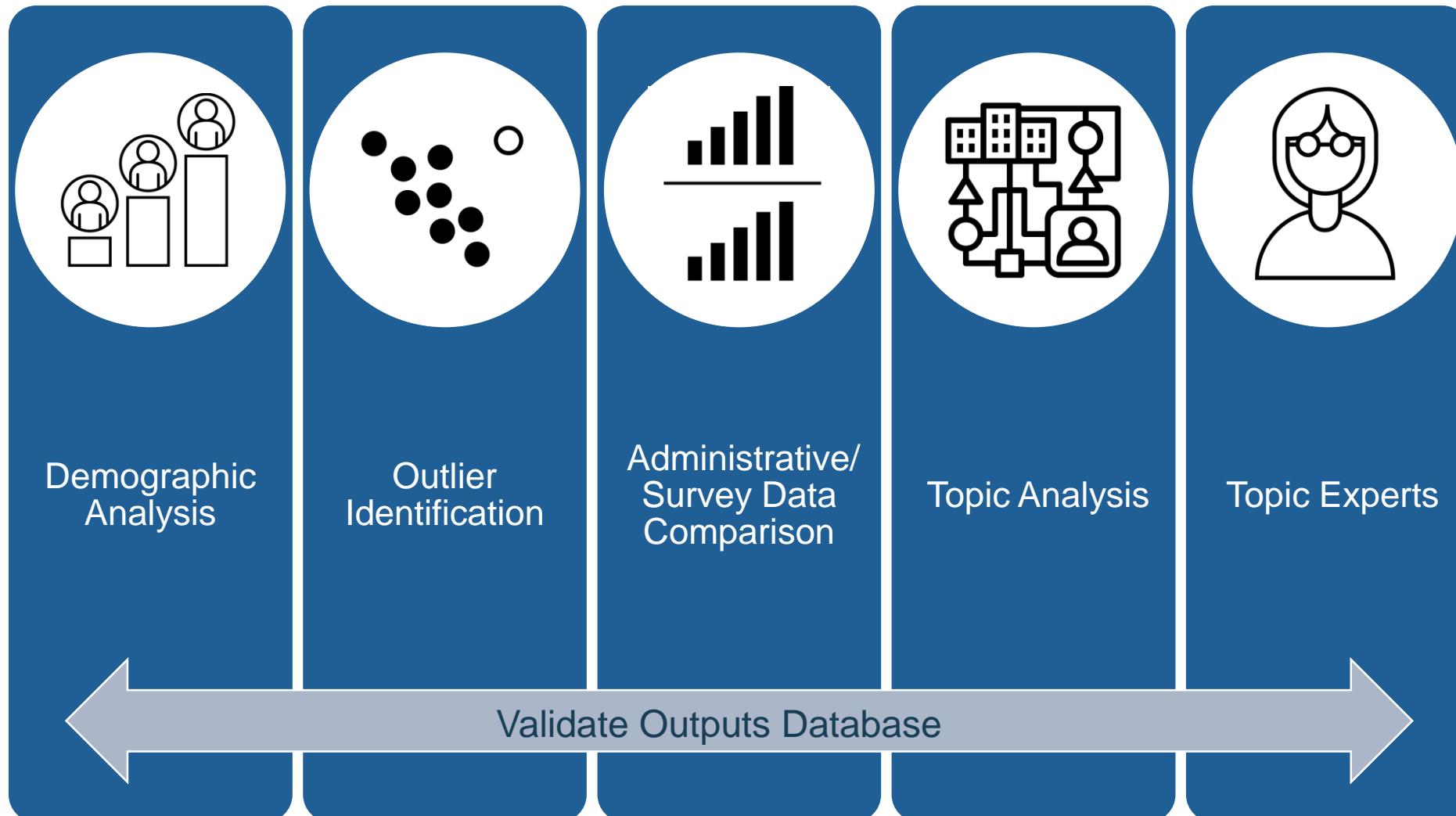
Impact Map

Why (the process)	Where (the issue)	How (can error be seen)	What (are we going to do)	Concerns	Action	Recommendation
Module 1 - Resolve Communal Establishments (CEs) (OUT OF SCOPE FOR COLLECTION REHEARSAL): <ul style="list-style-type: none">• Resolves multiple CE responses to one.• Removes Household (HH) responses that are from the same address as a CE response.• Moves individuals from discarded HHs to the retained CE at the address.• Adds individuals captured on HH Continuation and Individual forms at the same address as a CE response to the retained CE response at the address.	<ul style="list-style-type: none">• Removing 'too many' or 'too few' records.	<p>For each Local Authority and batch for the day:</p> <ul style="list-style-type: none">• No. of HH responses disregarded as identified as belonging to an address from which a CE response form was received.• No. of CE responses disregarded as identified as a duplicate CE response.• No. and % of HH visitors removed and no of questionnaires they were removed from.• No. of people added to CE records from HH questionnaires, no of CE records they were moved to and number of HH questionnaires they were moved from.• No. of people added to CE records from HH continuation questionnaires, number of CE records they were moved to and number of HH continuation questionnaires they were moved from.• No. of people added to CE records from individual questionnaires, and number of CE records they were added to.• No. of individuals with a residents ID that has been disregarded.	<ul style="list-style-type: none">• Monitor diagnostics - check that method is working as specified.• Produce values for number of households and percentage of households ID'd as containing multiple responses (HH, dummy and CE responses).• Metadata during processing - how long module took to run, whether it worked etc.	<ul style="list-style-type: none">• LS can be used as independent (1%) check on RFP and RMR processes.		
Module 2 -Identify Duplicate Household Responses (looser rules for collection rehearsals): <ul style="list-style-type: none">• Works out how many HHs there should be at an address when there are multiple HH responses received for an address.• Achieved through deterministic rules that are used to understand which HH responses are to be merged.	<ul style="list-style-type: none">• Getting this wrong impacts matching because the HHs won't match up.• RMR function fails.	<ul style="list-style-type: none">• Over count of respondents, more than expected.				
Module 3 - Resolve Household Responses (looser rules for collection rehearsals): <ul style="list-style-type: none">• Resolves multiple HH responses to one HH response	<ul style="list-style-type: none">• Getting this wrong impacts matching because the HHs won't match up.	<ul style="list-style-type: none">• Over count of respondents, more than expected.				

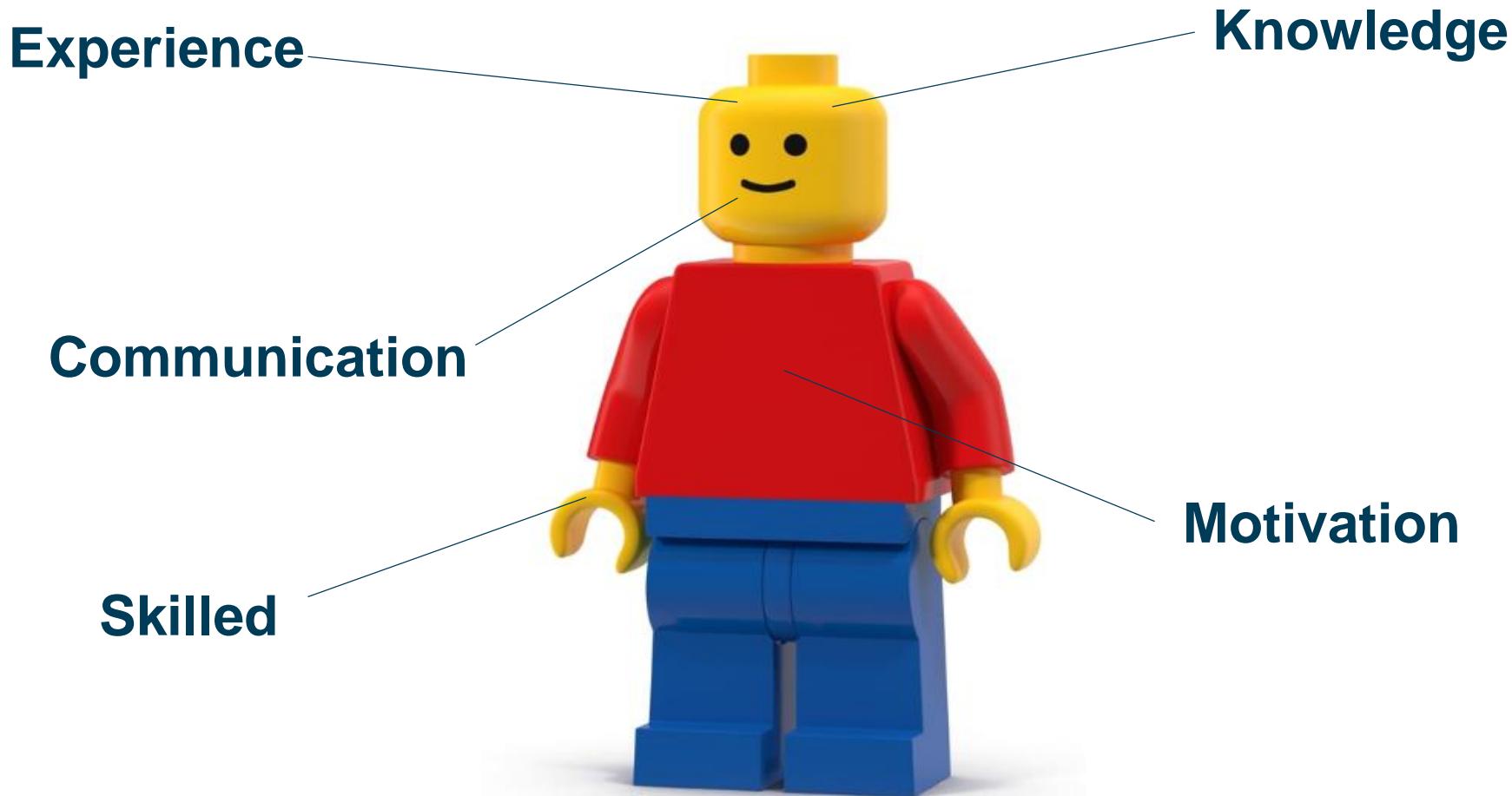
Looking forward to 2021 – Validate Outputs (Database)

- Revisited the Generalised Statistical Business Process Model (GSBPM 5.1, Section 6.2):
 - Checking that the population coverage and response rates are as required;
 - Comparing the statistics with previous cycles (if applicable);
 - Checking that the associated metadata, paradata and quality indicators are present and in line with expectations;
 - Checking geospatial consistency of the data;
 - **Confronting the statistics against other relevant data (both internal and external);**
 - Investigating inconsistencies in the statistics;
 - Performing macro editing;
 - **Validating the statistics against expectations and domain intelligence.**

Looking forward to 2021 – Validate Outputs (Database)



Looking forward to 2021 – Developing ONS Topic Experts



Fitness for purpose in a transformed social statistics system

- Regardless of how census-type statistics are produced..... '*getting it right matters*'
- Which elements of demonstrating fitness for purpose in Census 2021 can be taken forward to a transformed social statistics system with administrative data at its core?
- User confidence in fitness for purpose
 - Partnership
 - Accessible methods
 - Consistency
 - Technical confidence

Fitness for purpose in a transformed social statistics system

- Quality Assurance
 - GSBPM – ‘prevent, monitor errors and sources of risk’
 - Use of impact map approach in partnership with data owners
 - Identify potential quality issues
 - Diagnostics for issues arise
 - Mitigations for issues arising
 - Contingencies for issues arising
- Validation tools
 - Demographic analysis, outlier detection, comparison with survey/admin data, topic analysis/experts
 - **Challenge** – comparisons to admin data when same sources are at the heart of our methods