

# Intruder Testing

**Census 2021 England and Wales**  
**Risk and Utility in the Create a Custom Dataset System**  
Sam Trace

# Background

- Key [Census 2021 White Paper](#) promise 'Every person's identity will be protected, not only through secure handling and storage of their data, but also by ensuring that our statistical publications do not identify individuals'
- Since 2011, there has been exponential growth in information publicly available about individuals
- There is an all new customizable system for Census 2021
- Census 2021 has new methods protecting the data

# Statistical Disclosure Control (SDC) methods

- Targeted Record Swapping – identifying people and Households that stand out in the data, swapping them with a similar record in a nearby area.
- Cell Key Perturbation - this adds noise to the figures, making slight changes to cell counts
- Disclosure rules – automated rule-based checks run by the system, which decide if there is a low enough disclosure risk to allow the release of a dataset.

How do we check these have done enough?

# Intruder Testing

- [Intruder testing](#) is where ‘friendly’ intruders try to identify people in the data to check the risk level
- Census 2011 outputs were [intruder tested](#) before release
- It is a practical check to see if the methods worked
- The point of the exercise is to try and find out if it is possible to identify individuals in the data

# Legal Standard for outputs

- There must be ‘sufficient uncertainty’ about any identification from a small count
- Identifications made with publicly available information in combination with the data are included
- Testers do not need to be specialist hackers
- Methods must cover the ‘means likely reasonably to be used’

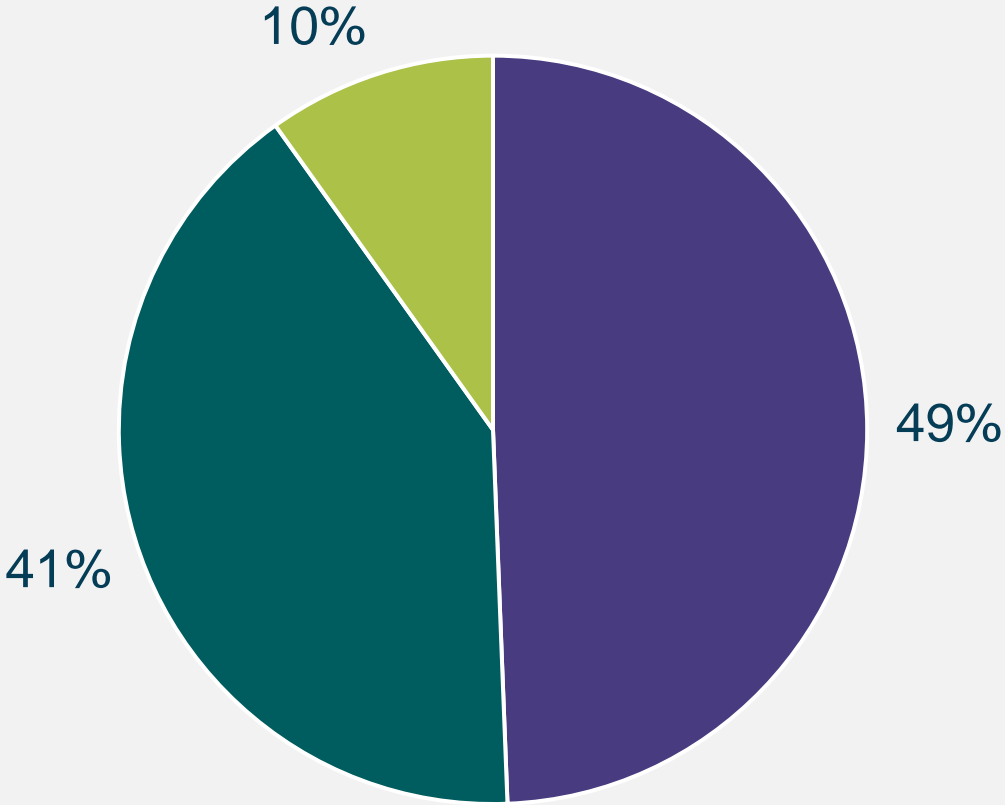
# Method

- Recruit intruders – ONS people only
- Consent intruders
- Train them and advise of the disclosure control methods
- Get the data on a secure pre-release system
- Intruders try to identify individuals in the data
- Collate results including feedback
- Analyse in Excel

# Results

- 51 Intruders recruited
- 30 confirmed as working on the project
- 24 intruders made claims
- 81 Claims made (excluding duplicates)

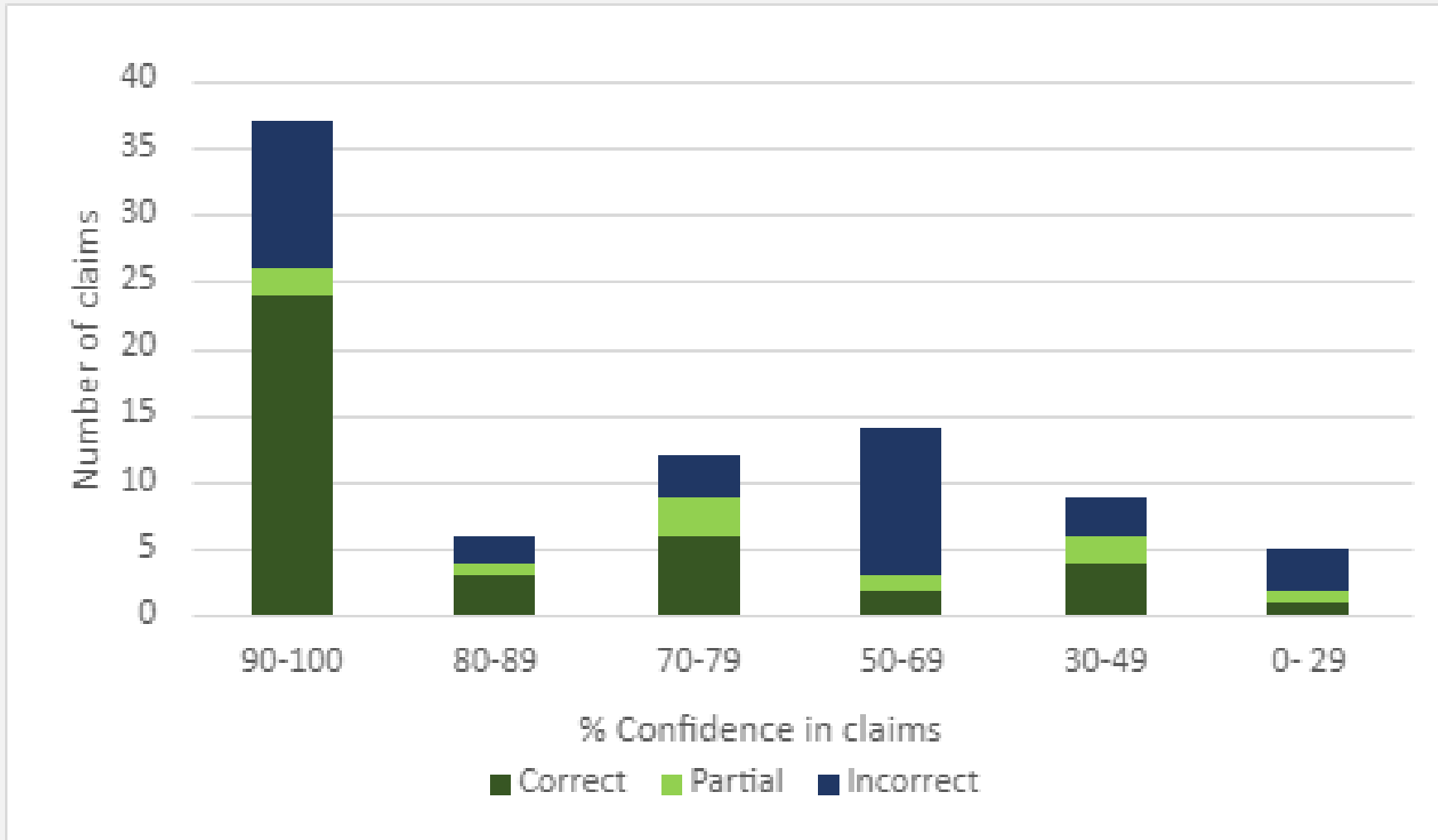
# Claims



■ Correct ■ Incorrect ■ Partial



# Confidence and Correctness



# Variables Used

	Correct	All	% Correct
Age	21	35	60%
Multi	12	29	41%
Occupation	2	8	25%
Other	5	9	56%
	40	81	49%

# Cell Counts

	Correct	All	% Correct
0-49	2	9	22%
50-99	3	9	33%
100-149	9	20	45%
150-199	6	9	67%
200+	20	34	59%
Total	40	81	

# Options

Remove detailed classifications from the Create Your Own Dataset system

- Loss of useful classifications at higher geography
- There may be other classifications not tried that also pose a risk

Limit max number of Cells

- Loss of useful functionality at higher geography

Specify Max cells specific to geography for univariates

- Would prevent the main risk

# Limit max cells by Geography?

- The majority of datasets used for claims and correct claims used would not be available
- Might need to apply to LSOA too as some OA claims could equally have been successful at LSOA
- MSOA claims were already likely to be unsuccessful

# Conclusions

- Detail available at low geography was a risk that was addressed in the live release system
- Some variables carry higher risk than others
- Changes to rules effectively blocked the main risks identified
- Automated rules in the [Create a Custom dataset](#) system worked to make claims harder to arrive at

# Actions

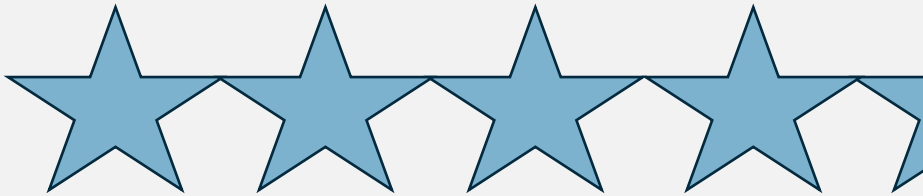
- Limit detail available at low geography
- Keep detailed topic summaries at MSA level geography
- Releases could take place as planned

# User Experience

We asked the intruders their opinions of the new system



## Ease of use



Choosing Variables

Choosing  
Classifications

Data was clear

**Questions and comments please**

**[SDC\\_queries@ons.gov.uk](mailto:SDC_queries@ons.gov.uk)**

**[Samantha.trace@ons.gov.uk](mailto:Samantha.trace@ons.gov.uk)**

# Speakers

## Samantha Trace

Methodologist

Statistical Disclosure Control

Office for National Statistics



Swyddfa  
Ystadegau Gwladol  
Office for  
National Statistics