Working Paper
ENGLISH ONLY

UNITED NATIONS ECONOMIC COMMISSION FOR EUROPE (UNECE)

CONFERENCE OF EUROPEAN STATISTICIANS

EUROPEAN COMMISSION

STATISTICAL OFFICE OF THE EUROPEAN UNION (EUROSTAT)

Joint UNECE/Eurostat work session on statistical data confidentiality
(Ottawa, Canada, 28-30 October 2013)

Topic (v): Confidentiality issues and case studies

Intruder Testing: Demonstrating practical evidence of disclosure protection in 2011 UK Census

Prepared by Keith Spicer, Caroline Tudor and George Cornish, Office for National Statistics, United Kingdom

Intruder Testing: Demonstrating practical evidence of disclosure protection in 2011 UK Census

Keith Spicer, Caroline Tudor and George Cornish

Office for National Statistics, Titchfield, Fareham, UK. PO15 5RR. Email: keith.spicer@ons.gov.uk.

Abstract: Legislation for the UK 2011 Census requires that no statistics are to be produced that allow the identification of an individual (or discovery of further information about an individual) by an intruder with a high degree of confidence. At the same time there is a need to facilitate sufficient detail to inform public policy and research for the public good. This balance of disclosure risk versus data utility is one of the biggest challenges facing disclosure control practitioners, with theoretical evidence for disclosure protection often not enough to convince a court of law. There has thus been a drive in recent years toward providing empirical evidence of protection, though this poses fresh challenges in its application and interpretation. This paper will describe an "intruder test" where ONS staff were supplied with actual disclosure-controlled 2011 Census tables for their local area and under secure conditions, attempted to identify people or households and consequently discover attributes. We will summarise the results and show how they provide real support for the SDC policy for 2011 UK Census data as well as highlighting and addressing the challenges of the approach for this and other data collections.

1. Introduction

With the recent push towards 'open data' and the sharing of greater amounts of information, the pressure is on National Statistical Institutes (NSIs) to publish more detailed datasets to broader audiences. It is of parallel importance for any such organisation to respect and protect the confidentiality of respondents' data. Historically disclosure risk has almost always been assessed theoretically. However, there is little known as to how easy it is in practice to identify real people from real databases, especially when they have been subjected to some form of disclosure limitation such as data swapping.

Willenborg (2012) has argued for an empirical approach in order to achieve a better understanding of 'intruders' and how they disclose information, which in turn would allow for better protection of data without straying into overprotection. This paper details the work by Office for National Statistics (ONS) in undertaking such an approach, presenting a case study of 'intruder testing' on UK Census data.

A formal intruder test carried out by data providers before release of data may be a powerful way of truly revealing whether a dataset can in fact be deanonymised. The idea of intruder testing in order to learn more about intruder behaviour has been considered in some form within NSIs – at least anecdotally. In the literature, Paass' 1988 paper demonstrates an early simulation of intruder testing using statistical models based on the German microcensus to create synthetic target persons. He studied various disclosure scenarios such as the tax investigator or the journalist considering the variables they might have access to. However there has been little reported on about formal attempts to use real intruders and to put this into practice in any systematic way. Recent examples were those of Simpson (2011) and O'Hara et al. (2011). The O'Hara work involved individual level reoffending and sentencing data from the Ministry of Justice and students acted as intruders, testing anonymised data prior to

online release. There was disclosure as a result of matching the profile of an offender named on a local news website. Although the offender was not unique, because there was a match certain deductions could be made about his/her reoffending data. This led to the data being further aggregated before final release. Simpson's (2011) work also involved students who were tasked with examining public data on *data.gov.uk* as part of an assignment on data security. They were given free rein to explore any privacy concerns in the datasets provided. They discovered that various datasets were poorly protected and the paper presents an actual example of linking statistical and personal information.

This paper provides insight on how to proceed with the task of intruder testing using real intruders and on a large-scale real dataset, in a structured way. Our case study is an exercise which we believe no other NSI has carried out and taken through to a conclusion, on real census data and prior to publication. This exercise carried some risk for the Census publication schedule and for the reputation of both ONS and disclosure control more generally - it could have generated a very different outcome if intruders were able to correctly identify all persons and then infer correct additional information about them (though one should argue that the impact would have been much less than a real disclosure claim post-publication that was shown to be correct). Despite the fact that our case study is based on a UK example, it is intended that this paper exemplify use of intruder testing as practical evidence for justifying means of disclosure control.

1.1 Support for Intruder Testing

In recent years, various initiatives relating to data transparency have drawn attention to the nature of data protection. In the UK, the Freedom of Information Act 2000 created a public 'right of access' to information held by public authorities[1]. This preceded the concept of 'open data', a UK Government initiative which held the principle that certain data should be freely available for everyone to use as they wish, embodied by *data.gov.uk* bringing together thousands of public datasets in one searchable website. The underlying theme was to promote transparency in government, supporting more informed choices in policy making and improving trust.

Until now, the field of disclosure control has been dominated by theoretical arguments and statistical measures of risk. The above initiatives have led to data providers coming under greater scrutiny; users want all the data if possible and public organisations have to provide appropriate justification for restricting levels of access. The argument is often made by users, verbally at least, that since there have never been any disclosures [with a particular dataset], the data must be safe and that more detail could be provided. There is always thus an impression that data providers are over-cautious. The argument is a difficult one for data providers to respond to, though the counter-argument is that the disclosure control methods have been successful and are doing what they are supposed to. In reality, there may well have been identifications and disclosures from data releases that have not been reported and there are good reasons why data users might not report them, namely that there may be some sort of legal sanction, or that there would be some negative effect on future data releases.

Theoretical arguments may not be persuasive in a court of law, however. The 2005 abortion statistics case involved a request under the Freedom of Information Act (FOIA) by the ProLife Alliance. In response, the UK Department of Health refused to disclose the suppressed figures on the grounds that the true counts were less than a pre-determined

[1] http://www.legislation.gov.uk/ukpga/2000/36/section/1

threshold of ten (zeros could be published) and subsequent concern about potential 'attribute disclosure'. The case was eventually concluded in 2011 at the High Court, the judgement being that the disclosure controls set out in the guidance were overly cautious in some circumstances[2] and they concluded that in this case the detailed abortion statistics were not personal data. The decision was based very much on the point that the Department of Health could give no *practical* evidence to demonstrate that individuals might be re-identified.

In 2012, the UK Information Commissioner's Office, whose responsibility it is to oversee the implementation of the Data Protection and Freedom of Information Acts, published 'Anonymisation: Managing Data Protection Risk Code of Practice' (ICO, 2012). The Code sets out recommendations for how organisations can establish whether it is reasonably likely that the publication and sharing of data will result in disclosure of personal data. Consequent to the abortions case, the Code specifically cited this example in recommending that it is good practice to adopt a 'motivated intruder test' as part of a risk assessment. The intruder test would evaluate whether it is possible to identify an individual from the anonymised information, either by itself or in combination with '*other information'*.

While results of intruder testing can be used as strong practical evidence to justify NSI SDC policy safeguards in support of theoretical evidence already gathered, their usefulness is not limited to this. Along the same theme as O'Hara et al. (2011), the results can go a long way to refining existing disclosure control measures. They can improve knowledge of the data, helping to inform decisions regarding which parts of the data are most vulnerable and under which circumstances. The case study presented in this paper will illustrate how intruder testing results informed output checking and review of table design before final release of the census data.

2. The UK Census – the concept of "sufficient uncertainty"

The disclosure control policy for 2011 UK Census is driven by the various legislation in place. Over-riding all is the Statistics and Registration Service Act (UKSA, 2007) section 39 states that personal information must not be disclosed. For the purposes of subsection (2) information identifies a particular person if the identity of that person is

(a) specified in the information

(b) can be deduced from the information or

(c) can be deduced from the information taken together with any other published information.

The legislation can serve to confuse, yet also to offer a wide range of interpretations. The Census White Paper section 6.7 (ONS, 2008) states that '*no statistics will be produced that allow the identification of an individual (or information about an individual) with a high degree of confidence*'. Section 6.8 of the Census White Paper continues to suggest that this degree of confidence may be interpreted in terms of the uncertainty '*about the true value of small cells*' and '*as to whether or not the small cell is a true value*'. In addition there is the

[2] https://www.wp.dh.gov.uk/transparency/files/2012/05/Revised-OCT-12-Data-required-for-research-purposes-bona-fide-research.pdf

Code of Practice (UKSA, 2009) accompanied by the National Statistician's Guidance: Confidentiality on Official Statistics which prescribes that it should not be possible for either identity or attribute disclosure to occur. In particular for 2011 the Registrars General determined that the focus of effort should be on addressing attribute disclosure and whether anything new can be learned about an individual or a particular group of individuals.

Unlike for the 2001 UK Census, this time the UK Registrars General agreed that small cell counts i.e. cell counts of one, could be permitted in published outputs as long there is '*sufficient uncertainty*' in those cell counts and in attributed disclosures. While it is not possible to eliminate disclosure risk entirely, the aim of the policy was to protect the data whilst retaining data utility.

Terms such as '*high degree of confidence*' and '*sufficient uncertainty*' - are vague, perhaps necessarily so, since the risk appetite will differ between those who use the data and those who may have to be subject to possible legal sanction. How uncertain do disclosure claims have to be before they are sufficiently uncertain? The interpretation is unclear but, as a minimum, one would expect it to mean that the majority of claims should prove to be incorrect.

2.1 SDC Policy for 2011 UK Census

A detailed evaluation of a number of possible methods was undertaken and record swapping was selected as the primary method for protecting tabular outputs. Swapping would be targeted towards those records most 'risky' at low level geographies, and the swap rate kept sufficiently low to maintain utility within the data. The details provided to the public on the exact nature of the methodology were limited (ONS, 2011). In addition, the level of detail was limited in tables to avoid sparsity at lower level geographies in particular, so as to reduce the risk of correct spontaneous recognition. The targeted swapping methodology used for the 2011 UK Census was an adaptation that is described in more detail in Shlomo et al. (2010). Households are targeted such that those that represent a greater disclosure risk across high risk/impact variables have a greater chance of being swapped. Households are paired so that they are similar across some of their census attributes but are in different geographical locations. Households are swapped at the geography small enough to make them non-risky: this new variation introduced to increase utility of the data by limiting change at the higher levels of geography. The percentage of records actually swapped is small and, as the public do not know which records these are, a level of uncertainty can be attached to *every* cell count. Thus a cell count of one, for example, may represent a swapped person and an intruder cannot be sure whether any claim of disclosure is correct. This is the basis for our intruder testing in seeking to measure the real level of uncertainty.

There is not zero risk in this approach. With the resultant tabular outputs, there are many real disclosures on display, but also some fake disclosures. A fake disclosure might be a cell count of one, for example, that ostensibly reveals information about an individual but actually represents a person with matched characteristics swapped into the area. Justification for ONS SDC policy is centred on the fact that although an intruder may claim disclosure, even with high confidence, he or she does not know for certain whether it is a fake or true disclosure. Thus when thinking about an intruder test, disclosure must be thought of in terms of probabilities. Sufficient uncertainty must be thought in terms of the overall percentage of correct claims of disclosure rather than a single case of correct disclosure. This percentage will relate to how effective the targeting of high risk records, the nature of intrusion as well as various other processes introducing uncertainty such as non-response or capture error.

3. Intruder Testing on the 2011 UK Census tabular outputs

This section describes the structure for the intruder test based on assessing protected census tables as appropriate for publication. Microdata samples and other products from the 2011 UK Census such as origin-destination tables were not considered as part of this test, though there is work in progress on a similar intruder testing exercise for the Census microdata.

3.1 Intruder Scenarios

In carrying out an intruder exercise, it is likely to be expensive, impractical and unethical to study real intruders and their intrusion methods (Willenborg, 2012). He suggests '*friendly intruders'* could be employed, that is, employees working at the statistical agency acting as much as possible as real intruders might do. These friendly intruders' task is to see if disclosure is possible, how much effort and knowledge is needed for this, and what other information may be available, etc.

The approach taken in this case study was to simulate the intrusion scenario of a member of the public with reasonable competence to be able to assess census tables and allowing them free rein to see what they could discover, if necessary matching with other publicly available information. Legally, under the SRSA (2007), we only have to consider publicly available information rather than privately obtained information e.g. that a business might hold. It would be reasonable, moreover, to assume that the intruder could have some information on their target, similar to that described in the spontaneous recognition scenario (Hundepool et al, 2012) where personal knowledge of basic demographics such as age, sex and marital status are likely to be known. On that basis the intruders were asked to think about the following disclosure scenarios:

(i) Can they identify themselves or their household
(ii) Can they identify someone they know, either individually or within a group, and their characteristics
(iii) Starting with public information, can they then identify someone, or a group of people, in a table (and learn more than the public information)
(iv) Starting with the census tables, can they identify a person or a group of people, and link this to some public information

The National Statistician's Guidance: Confidentiality on Official Statistics (UKSA, 2009) declares that for the 2011 Census it should not be possible for identity or attribute disclosure to occur, with attribute disclosure seen to be the key disclosure risk. The intruders were therefore asked, in each of the above scenarios, to think about what new, additional information they could glean on any Census respondent or household. It was stressed however that intruders should be free to use their own initiative and that ultimately they should try to disclose information as they wish.

3.2 Recruiting Intruders

We were restricted in terms of whom we could recruit as 'friendly intruders' since the census data were pre-publication with very strict rules on access. Anyone accessing the data was required to be security cleared to handle material classified as 'Secret'. They were also required to have signed a Census Confidentiality Undertaking which included training in

proper use of the data. This meant intruders had to be recruited from within ONS. Though this was a restriction, it did have the advantage that many of the intruders had good knowledge of census processes and census data, which could possibly lead to better judged claims.

The ICO's Code (2012) also provides some guidance on who an intruder might be in their recommended approach to an intruder test:

- Any person who starts without any prior knowledge but who wishes to identify the individual or individuals in the information and who will take all reasonable steps to do so.
- Reasonable competence: investigative techniques may be employed and additional knowledge of the identity of the data subjects, no specialist hacking skills or equipment.
- Consideration should be given to the risks associated with the disclosure of 'ordinary' or 'innocuous' information.
- Consideration should be given to the consequences and impact

Eighteen intruders were recruited in total. A range of intruders from administrative officers right up to director level staff were invited to take part, among these was one external contractor. Initially there was some reticence in the intruders' willingness to take part with many citing existing work pressures. This was resolved with greater endorsement from senior management and further explanation of what was involved. Perhaps paradoxically most intruders did not seem bothered about the privacy aspects of the study. In fact all intruders eventually agreed to take part apart from two where there were issues with Security Clearance. As the exercise took shape, other people were actually volunteering to participate. Intruders reported they found it fascinating to look at data for their own areas potentially containing either their own household or that of friends and neighbours. It was believed that this personal factor provided strong motivation. Only two intruders had any significant experience in disclosure control however they were all picked because they could be expected to rigorously test the data, had excellent IT skills and were adept with external databases. Of the eighteen intruders recruited from a range of local areas in England and Wales including urban, suburban and semi-rural. The results started to become clear at this point and given time/resource availability it was not necessary to recruit any further testers.

3.3 Data and level of detail

The census tables used for this test were those that were likely to be the most risky, i.e. those at the lowest levels of geography and also those that were the most detailed. These were provided for each intruder for their chosen area of interest. This provision was measured against the standard of '*disproportionate amount of time, effort and expertise for an intruder to identify a statistical unit*' taken from the ONS Protocol on Data Access and Confidentiality (ONS, 2004).

A complicating factor was that this exercise was carried out before final publication of tables (pending final table design and evidence from this work). Thus a mock-up of the final tables to be published was created which had the SDC methodology described in section 2.1 applied (targeted record swapping and table design): in total 89 carefully selected tables were compiled which were mostly a combination of Local Characteristics, Theme tables and Detailed Characteristics as described in the ONS 2011 Census Prospectus (ONS, 2013).

Theme tables contain many different variables (often ten or more) on a particular theme such as religion. Local Characteristics contain around three or four variables only but are at the lowest output geography available: OA level (Output Area, around 125 households). Detailed Characteristics have a detailed categorisation of variables at the MSOA level (Middle Super Output Area, around 25 OAs) but only on a small cross-classification of variables. There have been only very minor amendments to the tables since creation so they should be an accurate representation of the published data. Where there were similar tables, only one or two from the set were created (due to resource issues) so the chance of intrusion via differencing between tables was reduced (though not eliminated), as was also the chance of attribute disclosure via comparison of tables where a person has already been identified in one of the tables.

It is recognised that the incidence of identification of individuals might have been higher if there had been unlimited access to the census tables. However this was controlled for as much as possible by selecting the most risky tables with the tables provided being either OA or MSOA level tables (with only a few univariates at higher geographies). Furthermore, to get an accurate reflection of disclosure across the country, a range of areas were included: urban, semi-rural and suburban (influenced by the initial recruitment of intruders).

3.4 Externally available information

Elliot et al. (2009) detailed barriers to obtaining public information which include cost and level of access which, as in our intruder test, determined that intruders could only do a case-by-case search as opposed to being able to use an entire record file. Intruders were provided access to two separate laptops; one with census data and one with unrestricted internet access (the two kept separate for security reasons to avoid emailing out of census data). This may have hindered the use of the internet for the risk scenario of linking census tables to publicly available datasets containing the population. However obtaining entire datasets publicly is difficult and believed more likely in the scenarios of Paass (1988) where privately available information is used, that that a company might hold – a business register for example.

Intruders were provided with a guide to websites that might be useful in finding or matching information, though it was stressed that they were not restricted to these. Intruders were told they would be reimbursed up to the sum of £50 if they felt payment was needed to assist in making a disclosure; for example to register on paid websites which provide name and address details and basic socio-demographic information on age and sex. This amount seemed appropriate given a review of the websites and what extra information might be obtained.

3.5 Conditions of reasonable time and effort

Each intruder was invited to a session and was encouraged to spend around three and a half hours examining the data; an amount of time considered ‘likely’ and ‘reasonable’ in line with National Statistician Guidance. The intruders worked in a meeting room within a secure census area which required pass access. Each intruder was given a 20-minute briefing before starting their session. They were given a summary of the SDC Census 2011 policy (including a written Q&A), talked through possible theoretical examples of disclosure, supplied with a

guide on ‘how to find people online’, a list of the selected Census tables and their title/definition, and maps of the OA and MSOA to which the tables related.

Intruders were directed to write down names and addresses and the reference of the cell(s) and census table(s) that indicated any disclosure. Intruders were encouraged to think about the exercise beforehand and to bring in any extra material to help them make any disclosures. Although a few did come up with a strategy beforehand, no-one brought in any extra material.

3.6 Security and Ethical Considerations

Security of the census data in this exercise was taken very seriously. Intruders were asked to sign a confidentiality declaration which, although it had no legal basis, reinforced the message not to misuse the data. Intruders wrote all notes in notebooks which were taken and locked away on completion of the sessions. Laptops were chained to the desk during the exercises and all data were securely wiped after each session. In addition, intruders were made aware that any searches they made could be stored by website providers, particularly those that required registration or payment.

This intruder exercise, by definition, involved searching for personal information about people on the web. This raised ethical questions since on some websites (e.g. 192people.com) it is possible to pay to see who has ever web-searched you. Intruders were made aware of this at the start so they could avoid these websites if they felt it was necessary. Intruders were also asked to supply names and addresses along with each claim of disclosure necessary for validation purposes. Thus the exercise required a certain level of trust on the intruders’ part for the information they provided to be dealt with confidentially. Assurances were provided to intruders in the confidentiality declaration and in briefing and speaking with them. Personal details were stored on secure census computers and very limited summary details provided outside of the team directly involved.

4. Analysis of Intruder Reports

This section discusses validation and subsequent results from the intruder test, and summarises opinions of the intruders. Exact figures on number of claims crossed with the various breakdowns have been omitted to avoid compromising confidential information.

As discussed in section 2, the concept of disclosure risk is more challenging with tables protected by record swapping since many disclosure claims will in fact be correct. This is because record swapping does not eliminate disclosure but aims to provide uncertainty around which cases (usually small cell counts) are genuine. In keeping with empirical evidence already gathered on proportions of ones and attribute disclosures that are real, the intruder testing validation (described in section 4.1) operated on the basis of assessing the proportions of claims of disclosure that were true. Note that disclosures introduced by processes of imputation would be considered *not real* as well as those introduced by swapping.

Intruders were asked to note down a, preferably numerical, level of confidence they had that any claim was correct. Claims with high levels of confidence are of particular interest since these are the ones that the public are most likely to use in challenging SDC policy. The high confidence claims are also those that the courts are most likely to take seriously given that it would be 'reasonable' to expect the intruder to provide significant justification rather than making entirely random claims. Some confidence levels were based on what the intruder had written down in words or said, for example, 'fairly sure' was taken to imply a confidence interval of 60-79%. Table 1 defines the intruder confidence levels: this aid evolved from the words used by the intruders and their numerical association of confidence.

Confidence Level	**Meaning in words**
0-19%	Not at all confident, complete guess
20-39%	Not very confident, bit of a guess
40-59%	Not quite sure, uncertain
60-79%	Fairly sure, reasonably confident
80-100%	Very confident, absolutely sure

Table 1: Interpreting intruder confidence levels (based on intruders' words and numbers)

Overall there were more than 50 claims from all 18 intruders with between 0 and 8 claims per intruder during their 3.5 hour session. There were two types of claim – a claim of identity disclosure and a claim of attribute disclosure. This generally translated to "I can recognise somebody I already know well, such as a close friend or family member, in these data" in the case of identity disclosure and "I can find out something I didn't already know about someone I know well, such as a close friend or family member" in the case of attribute disclosure.

There were several known reasons for certain claims being wrong such as non-response, imputation, capture processing (especially coding from either free text or multiple ticks), respondent error as well as record swapping. Interestingly though, several claims were incorrect due to intruder error. There were a variety of reasons for this, for example intruders typically forgot what they or their household wrote on their census form and thought themselves to be in a different category to what was originally recorded. On other occasions, intruders were incorrect when trying to work out what a family member would have put when answering (e.g. on questions about occupation or health), sometimes making incorrect assumptions. Also there were instances where the table was misread, or the variables in the table were not understood fully. The test also illustrates the difficulties of matching against public data, issues with timeliness, collection modes and variable categories all being different.

4.1 Context - Validation

The majority of intruders made some claims of disclosure although two did not make any. The validation process involved five steps:

(i) The cell reference and the table relating to the disclosure claim were noted.
(ii) Using that information, the record(s) in that cell was identified in the Census unit record database (post-swapping, post-imputation).
(iii) Each record in the database contains a form identifier which was noted.
(iv) The form identifier(s) was used to select the relevant image(s) of Census form(s) in the Census database – (obviously pre swapping and pre imputation now as we were looking at the original scanned images).
(v) The image form was checked to see if it matched the name and address supplied by the intruder.

Note that this was a relatively rudimentary approach to validation, working on the principle that if the person/household referenced in the cell correctly related to the original Census form then that counted as a correct disclosure. However validation of these claims had the potential to be more complex if one were to give further expert thought on how to go about it. For example if an intruder made a claim about a cell count of one stating that it was their neighbour – say person A, it could be the case that although the data correctly recorded them in the original pre-swapped, pre-imputed database, they had been swapped with person B. Therefore in the post-swapping post-imputation cell it shows person B. This claim would be incorrect under our principle when one could argue it should be a correct claim because the particular characteristics in this table are the same. This approach was considered to be the best and most straightforward because any further inferences on the person (in this case a neighbour) *are likely* to be incorrect since the person has been swapped and will not necessarily match on any other characteristics in any other table. This fits with the policy focus on attribute disclosure where new information learned is of main concern rather than merely recognising a person. With more thought and time the results could have been broken down to highlight the nuances of the impact of the swapping and imputation methodology (not discussed either are cell counts which do not equate to the same value in the pre-swapping pre-imputation database). It was considered that a court of law is probably more likely to work on the straightforward principle stated at the beginning of this paragraph.

It was sometimes difficult to discern whether intruders were claiming identification or attribute disclosure. An attribute disclosure occurs when a table enables an intruder to infer further, previously unknown, knowledge about a person or household. However many intruders may not have realised that further information could be inferred, or did not actively look for this type of disclosure once they had identified someone (although this was encouraged in the briefing beforehand). There were also several claims of group identification disclosure involving identification within small cells, for example claiming a relative was represented in a cell which had a count of two or three.

4.2 Summary of Results

Table 2 shows an overall pattern where claims made with greater confidence were more likely to be correct although unexpectedly there is a slight dip towards the claims made with certainty. Note the dip may be due to small samples in each confidence level rather than anything statistically significant.

Confidence Level	Percentage of claims correct
0-19%	0%
20-39%	11%
40-59%	38%
60-79%	67%
80-100%	47%

Table 2: Percentage of correct claims by confidence level

The results were then filtered based on whom the intruder had identified as in table 3. 'Neighbour' refers to any neighbour in the immediate area. Importantly, aside from one single case, all claims made involved persons for whom the intruder claimed to have personal knowledge. Moreover the results indicated a far greater chance of disclosing correct information about their own household or a family member emphasising the difficulty in identifying someone you know less about, which is usually the aim of an intruder. Another interesting result was that intruders who claimed to have found themselves were not always correct.

Who is being claimed	Percentage of claims correct
Neighbour	36%
Self/Family	61%

Table 3: Subject of the disclosure claim

When the claims were filtered by geography (table 4), it was clear that being able to identify someone at MSOA level proved very difficult with only a very small number of claims made in comparison with at OA level. Intruders often commented on how the Local Characteristics tables, which are all at OA level, were far more useful than any other tables despite having less detail in the constituent variables. Intruders who thought they might have identified someone in the table in an MSOA would often then look for greater confirmation in the OA tables.

Geography Level	% claims correct	% claims made
OA	45%	94%
MSOA	50%	6%

Table 4: Level of geography at which a claim was made

The results were also broken down by the type of claim made (table 5). The types of claim were ‘Person’ or ‘Household’, in which a person or household has been identified, or ‘Swapping’ which were fascinating cases referring to where an intruder believes somebody has been swapped in or out of the data. Again the majority of claims were incorrect within each category of type of claim. Whilst claims of disclosure were made at the household level, in the main they were concerned with a particular person.

Claim Type	Percentage of claims correct	Percentage of claims made
Household	50%	16%
Person	46%	74%
Swapping	33%	10%

Table 5: Correctness of claim by type (household/person/swapping)

Finally which Census tables were used to make the most claims was analysed and the associated success rates of the claims – the former shown in table 6. A select number of tables were much more commonly used than others (giving rise to five claims or more) making up over half of all claims. These popular tables all included sex and age bar one which only included sex. It is worth pointing out that whilst most intruders made a claim in just one table, some were able to identify a person/household across several tables. Interestingly, the three most used tables for making claims show an above average percentage of claims correct in comparison with the other tables (figures not disclosed).

Table Used	Summarised Topic	Claims of Disclosure
CAS036	sex/industry/age	10
CAS002	age/sex/marital status	7
CAS028	sex/age/economic activity	6
CAS038	sex/industry/employment	5
CAS016	sex/age/health/disability	5
CAS119	sex/age/travel to work	4
CAS056	tenure/central heating	4

Table 6: Census tables used to make the claims

As mentioned earlier, the Registrars General had agreed that attribute disclosure, not identity disclosure, was to be the primary concern for the 2011 UK Census. Some attribute disclosure claims were made, but whether the figures for these are truly representative of how many could or should have been made is unclear. Of the instances where an intruder said they could identify someone in the tables, only in a tiny number of these instances were the intruders able to infer further new knowledge with not all being correct. However, this is not totally reliable, since some may not have realised further information could be inferred, or did not actively look for this type of disclosure, as mentioned earlier. The number of cases is very small and not really adequate to make sweeping judgements.

Referring back to the scenarios the intruders were asked to consider in section 3.1, it was very difficult for intruders to obtain data from the internet and then link it to anyone in the census tables and only one claim of this type, which proved to be incorrect, was made. Of the others who tried to do this, they said they were unable to match up the information they had found to a low count in the tables. Some intruders did use the internet to help with identifying people, in particular to verify basic information such as names, addresses and ages. On one occasion, the use of internet information to support a claim led to incorrect identification of a person, whereby an age obtained from an internet site proved incorrect. A possible alternative explanation is that the intruder did not take into account the difference in time between the census and the date of the test.

4.3 Intruder Comments

All intruders were asked to write any comments they had at the end of the exercise. There were many comments stating how difficult it was to try to identify someone, with these intruders generally saying how their area had little variation demographically, making it difficult for people they knew, or had found, to stand out in the tables.

There were concerns expressed with small counts of certain ethnic groups and religions within the OA tables. They suggested that these low counts, although not identifiable to

themselves, may well be to others and they perceived this to be unsafe. Coincidentally in all of these cases the counts were found to be incorrect due to either misreading the categorisation in the table or due to swapping.

A final point is that intruders were asked whether they felt the data generally reflected the characteristics of their area and were prompted to ascertain the swap rate. The general consensus was that the tables looked correct for their area and that the swap rate looked 'low'.

5. Conclusions

Intruder testing provided empirical evidence, within the legislative framework for the 2011 UK Census, of whether the SDC policy was satisfactory. It backed up theoretical evidence already gathered and indicated whether specific strands to the methodology were working, i.e. the targeting of risk variables for swapping. It was helpful to demonstrate where further work needed to be done, highlighting vulnerabilities in the data. For example variables such as age and sex were used as common identifiers. Part of the ONS SDC strategy has required refining the format of the tables prior to final publication; results of the intruder testing informed which tables might need special attention and these were prioritised for review. Intruder testing gave indications on how easy it was to match data with public sources and (mostly) unguided by pre-defined scenarios, the creative ways intruders might try to identify individuals or groups of individuals. Intruder testing is finally very helpful to get a handle on perception of disclosure. It gave a unique insight into how the public might perceive disclosure control, if any has been applied and to what extent.

To be specific, this novel example of intruder testing on disclosure-controlled (anonymised) census data has shown that it is very difficult to re-identify respondents correctly in the 2011 UK Census and moreover, it is virtually impossible in this case to identify anyone correctly without any personal knowledge about them. The ICO guidance recommends assessing disclosure based on a person without any prior knowledge. Even given that the claims that were made about people the intruder would expect to know reasonably well, i.e. family, themselves or a neighbour, the percentage of correct claims was still surprisingly low. Indeed, the majority of all claims even made with 80-100% confidence were still incorrect; these were cases where the intruder was almost or absolutely certain that they knew the identity of the person in the table cell. This work has also demonstrated that a great deal of external information about a person would be needed to be able to match them to a table cell. There was a very low incidence of claims of attribute disclosure, where an intruder found out 'new' information about a person and again not all of these were correct. These attribute disclosures were all conditional on a prior identification disclosure, of which the majority were not correct. This was deemed to be the primary disclosure risk for 2011 UK Census so in that context the result was very pleasing.

In reference to ICO's Code of Practice (2012) on a motivated intruder test and the impact of the claims, identification of people in tables with sensitive variables such as health or disability was shown to have an even lower chance of being correct, likely to be a direct result of the ONS' SDC policy of targeting high risk/impact variables.

Leaving aside the nature of the disclosure claims, it is important to remember that some claims made *were* correct. It is generally impossible for NSIs to publish data that retain some

usefulness with zero disclosure risk, so it has always been about getting the right balance. The precise nature of the claims as discussed indicate that the level of disclosure control applied provides sufficient uncertainty where needed in the tabular outputs. This work has since been endorsed by the UK ICO "*proving re-identification risk (as we call it) is manageable*".

The authors note there were limitations to this exercise which should be considered when making any generalised conclusions. To be specific; we refer to the socio-demographic characteristics of the intruders and their motivation given that they worked for ONS, the small sample of postcodes assessed, the fact that only a representation of the full set of output tables to be published was examined, the time spent by intruders at each session, and the availability of external information. Further work of this kind would be helpful in working towards a standardised approach for intruder testing in terms of methodology, time, likelihood, use of external information. For example it may be argued that such an experiment could be prolonged over a longer period for more reliable results.

To conclude, intruder testing can have all sorts of applications. It can be applied to survey as well as census data to assess levels of disclosure risk pre- or post-SDC. It can also be applied to administrative data (registration based data from government and government-funded organisations). There is a continuing drive to examine mechanisms for combining administrative data and making it available for research safely. Intruder testing can be very useful to pinpoint parts of the data or variables which are of particular concern for which greater protection is required. At ONS, collaborative work of this nature is underway with Manchester University to design innovative intruder tests for social survey microdata – data that are currently only available under licence - and being considered for more open access. More detailed disclosure scenarios will be examined and both public and private external databases will be used.

Given the push to get more data out into the public domain, NSIs should consider their thinking about justification of SDC policy. In particular, as well as considering theoretical disclosure risk, they might consider whether there is sufficient practical evidence *not* to release data rather than automatically assuming the contrary. Intruder testing exercises are difficult to set up, and do take considerable resource, but they can provide valuable empirical evidence that might be more persuasive in a court of law.

References

Elliot M. J., Mackey E., Purdam K. (2009) 'Data Environment Analysis Service (DEAS)', ONS Internal Report.

Hundepool, A. Domingo-Ferrer, J., Franconi, L. Giessing, S., Schulte Nordholt, E., Spicer, K. and de Wolf, P.P. (2012) **Statistical Disclosure Control**, Wiley Series in Survey Methodology

Information Commissioner's Office (ICO), (2012) '*Anonymisation: managing data protection risk code of practice'* Available online at: http://www.ico.gov.uk/for_organisations/data_protection/topic_guides/~/media/documents/library/Data_Protection/Practical_application/anonymisation_code.ashx [Accessed June 2013]

O'Hara, K. Whitley, E. and Whittall, P. (2011) '*Avoiding the Jigsaw Effect: Experiences with Ministry of Justice Reoffending Data'*, Southampton University Technical Report. Available online at: http://eprints.soton.ac.uk/273072/ [Accessed June 2013]

Office for National Statistics (ONS), (2008) '*2011 Census White Paper'*, Available online at: www.ons.gov.uk/ons/guide-method/census/2011/how-our-census-works/how-we-took-the-2011-census/how-we-collected-the-information/questionnaires--delivery--completion-and-return/2011-census-questions/index.html [Accessed June 2013]

Office for National Statistics (ONS), (2011) '*Evaluating a statistical disclosure control (SDC) strategy for 2011 Census Outputs'* Available online at: http://www.ons.gov.uk/ons/guide-method/census/2011/how-our-census-works/how-we-took-the-2011-census/how-we-planned-for-data-delivery/protecting-cofidentiality-with-statistical-disclosure-conctrol/index.html *[sic]* [Accessed June 2013]

Office for National Statistics (ONS) (2013) '*2011 Census Prospectus'* Available online at *http://www.ons.gov.uk/ons/guide-method/census/2011/census-data/2011-census-prospectus/2011-census-prospectus.pdf* [Accessed June 2013]

Paass, G. (1988) 'Disclosure risk and disclosure avoidance' *Journal of Business Economics and*
Statistics, **6**, pp.487-500.

Shlomo, N., Tudor, C. and Groom, P. 'Data Swapping for Protecting Census Tables.' In *PSD'2010 Privacy in Statistical Databases*, 41-51. Germany: Springer LNCS 6344, 2010. eScholarID:168603

Simpson, A.C. (2011) 'On Privacy and Public Data: A study of data.gov.uk' *Journal of Privacy and Confidentiality* **3** (1), pp. 51-65

UK Statistics Authority (UKSA), (2007) '*Statistics and Registration Service Act'* Available online at: http://www.legislation.gov.uk/ukpga/2007/18/contents [Accessed June 2013]

UK Statistics Authority (UKSA), (2009) '*Code of Practice for Official Statistics'* Available online at: www.statisticsauthority.gov.uk/assessment/code-of-practice/index.html [Accessed June 2013]

Willenborg, L. (2012) Discussion [of Skinner, C.: Statistical Disclosure Risk: Separating Potential and Harm]. *International Statistical Review*, **80** (3) pp. 375-378.