HM Government

Simplifying the transition to Individual Electoral Registration

An evaluation of pilots testing data matching for the purposes of confirming electors on the Electoral Register- *Full Evaluation*

April 2013

Cabinet Office
25 Great Smith Street
London SW1P 3BQ

Publication date: April 2013

© Crown copyright 2013

You may re-use this information (not including logos) free of charge in any format or medium, under the terms of the Open Government Licence.

To view this licence, visit www.nationalarchives.gov.uk/doc/open-government-licence/
or write to the Information Policy Team,
The National Archives, Kew, London TW9 4DU,
or e-mail: psi@nationalarchives.gsi.gov.uk.

Any enquiries regarding this document/publication should be sent to us at
electoralregistration@cabinet-office.gsi.gov.uk

Contents

Chapter 1 **Introduction**..... 3
1.1 *Context*.....3
1.2 *Background*..... 3
1.3 *Research aims*..... 4

Chapter 2 **Methodology**.....5
2.1 *Participating areas*.....5
2.2 *Pilot approach*..... 6
2.3 *The matching process*.....7
2.4 *Testing the accuracy of the matching*.....8
2.5 *Local data matching*..... 8
2.6 *Assessing the process of confirmation*.....8

Chapter 3 **Results**..... 10
3.1 *Match rates*.....10
3.2 *Understanding the variation in match rates*.....14
3.3 *Testing the accuracy of the matching*.....17
3.4 *Local data matching*..... 19
3.5 *Other process issues*.....19

Chapter 4 **Summary & Conclusions**.....21

Annexes
A: Data matching methodology..... 25
B: Match results..... 29
C: Analyses of variation in match rates.....31
D: Pre and post canvass register comparisons.....32

Chapter 1
Introduction

1.1. Context

Under the current system of electoral registration an annual household canvass form is sent to each address, which is completed by one individual on behalf of everyone living at the property. From 2014 this system of registration will be replaced by one of Individual Electoral Registration (IER), with individuals making an application to register individually and providing personal identifiers (such as date of birth and National Insurance Number).

Ensuring that the registers are as complete and accurate as possible and that levels of completeness and accuracy do not decline under IER is a key aim of the Government. Data matching, whereby records on the electoral register are matched against other sources of public data, is one tool which could assist in ensuring that the registers remain as complete and accurate as possible, both during the transition to IER in 2014/15 and on an ongoing basis.

This paper presents the results of an evaluation of data matching specifically for the purposes of confirming electors who are currently on the electoral register, which has the potential to simplify the transition to IER for the majority of electors. A separate evaluation, looking at the potential for data matching to help find new electors, will be published in July 2013.

1.2 Background

In 2011 the Cabinet Office ran a set of pilots exploring whether matching entries on the Electoral Register to other trusted public data sources could help to identify individuals who are not currently registered to vote but who may be eligible to do so. By providing the information to enable Electoral Registration Officers (EROs) to contact these individuals and invite them to register the overall aim of the pilots was to help improve the completeness of the register, as well as improving the accuracy of the register through the identification of potentially inaccurate registrations.

During the course of these pilots, an additional potential use for data matching was identified. The 2011 pilots demonstrated that a large proportion of individuals on the

electoral register (around two-thirds) could be positively matched within data held by the Department for Work and Pensions (DWP). By using data matching to 'confirm' these electors, an opportunity to automatically transfer individuals to the new IER register, without the need to provide personal identifiers, was identified. This process of 'Confirmation' has the potential to simplify the transition to IER for the majority of citizens and to improve the likely completeness of the electoral register across the transition to IER[1].

However, as the 2011 pilots were not designed to test data matching for the purposes of confirmation it was recommended that further piloting be undertaken, across a variety of area types, to allow differences in the confirmation rates to be explored and work to assess the accuracy of the data and match rates to be undertaken.

These pilots took place between September 2012 and March this year, involving 14 pilot areas from across England, Wales and Scotland. Preliminary findings from the pilots were published in December 2012 by the Cabinet Office and by the Electoral Commission (EC), which has a statutory responsibility to evaluate the pilots. These findings indicated that confirmation could be effectively and reliably used to simplify the transition to IER for the majority of existing electors and it was recommended that the policy should be pursued as part of the transition to IER.

This paper builds upon those preliminary results and presents the complete findings of the Cabinet Office evaluation. The complete evaluation (and interim findings) from the EC are available at: http://www.electoralcommission.org.uk/publications-and-research/our-research.

1.3 Research aims

The key aims of the Cabinet Office evaluation were to:

- assess the potential match rate that could be achieved by comparing DWP data with Electoral Register data, to include analyses of the variation in match rates between geographical areas and population groups (where groups are identifiable within the data);
- assess the accuracy of the matching process for the purposes of confirmation; and
- examine the process of confirmation and related implications for IER.

[1] The Cabinet Office evaluation of the 2011 data matching pilots is available at https://www.gov.uk/government/publications/cabinet-office-evaluation-of-data-matching-pilots-2011

Chapter 2
Methodology

This chapter describes the methodology for the pilots, which was developed in close collaboration with researchers from the Electoral Commission which has a statutory duty to evaluate the pilots.

2.1 Participating areas

Electoral Registration Officers from across England, Scotland and Wales were invited to participate in the 2012 data matching pilots. In total, fourteen local areas volunteered to pilot data matching for the purposes of confirmation, including 12 Local Authorities in England & Wales and two Scottish Valuation Joint Boards (VJBs). However, it is important to note that these areas were not purposively sampled and cannot be assumed to be representative of all areas.

Table 2a: Overview of pilot areas

Pilot area	Country/region	Population (16+ yrs old)
Ceredigion	Wales	64,128
Conwy	Wales	96,263
Greenwich	London (Outer)	199,927
Harrow	London (Outer)	192,025
Lothian VJB	Scotland	705,824
East Lothian		*79,302*
Edinburgh, City of		*421,700*
Midlothian		*66,776*
West Lothian		138,046
Manchester	North West England	405,174
Peterborough UA	East England	145,207
Powys	Wales	110,310
Renfrewshire VJB	Scotland	278,209
East Renfrewshire		*72,205*
Inverclyde		*65,566*
Renfrewshire		*140,435*
Southwark	London (inner)	235,351
Sunderland	North East England	227,315
Tower Hamlets	London (Inner)	205,645
Wigan	North West England	258,205
Wolverhampton	West Midlands	200,314

Source: Office for National Statistics, National Records of Scotland 2011 mid-year population estimates

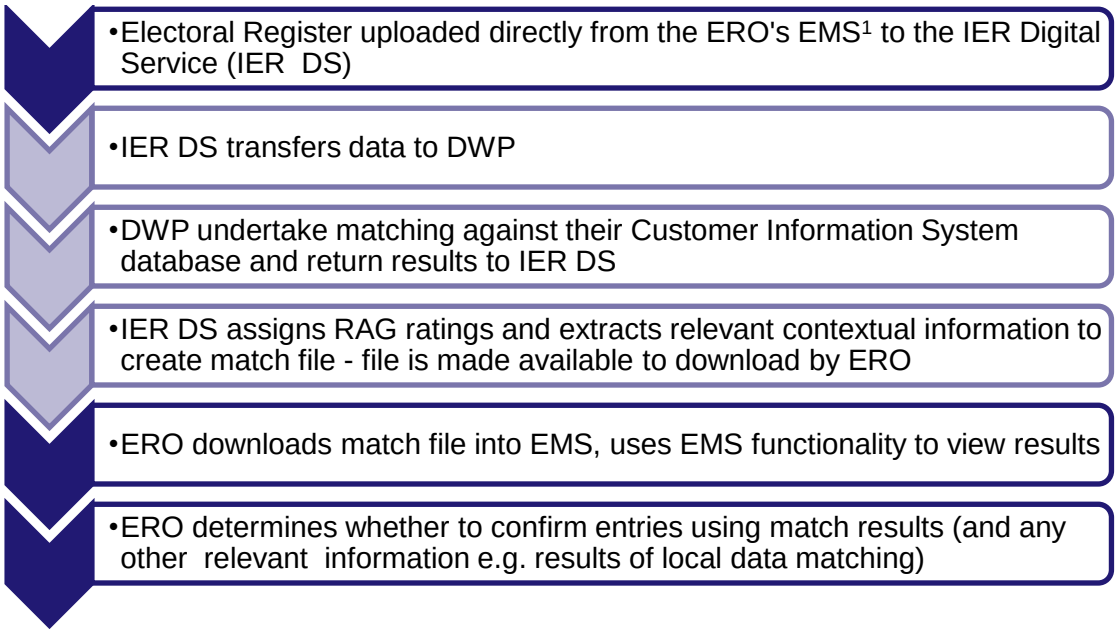
2.2. Pilot approach

The pilots sought to simulate the proposed process of confirmation, outlined in Figure 2a, as far as possible. However, it is important to note that the digital solutions that will help deliver confirmation were under development during the pilot and therefore it was not possible to replicate the process exactly. The key implications of this were that:

a) The data was transferred via secure courier as opposed to directly through the 'IER digital service[2]' – meaning that the process was more resource intensive and gave rise to the possibility that differences in the way the data was processed could have impacted on the results of the data matching. (This is because some of the data formatting was done manually, however subsequent testing indicates that this had no significant impact on the match rates)[3].

b) The match files were provided to Electoral Registration Officers (EROs) in basic CSV format as opposed to reports within their Electoral Management Software (EMS)[4] – meaning that EROs were required to analyse the data independently as opposed to using report functionality that will be available within their EMS.

Fig 2a: Outline confirmation process (Steps in light blue require no action from the ERO)



Notes: Steps in light blue require no action from the ERO. In live-running, once the steps outlined above are complete local areas will carry-out follow-up action, including writing to individuals who have been confirmed to notify them their details have been transferred to the IER register and issuing invitations to register for those individuals who have not been confirmed.

[2] The IER digital service is the mechanism through which data will be transferred between EROs and DWP in the confirmation exercise, enabling EROs to securely upload/download files from/to their existing software.

[3] These tests involved passing the data for three areas through the IER service and comparing the match results of these files to those of the comparable files from the pilots (which had been transferred via courier and then manually formatted). The results showed no significant difference in the results (ranging from 0 – 0.5 per cent increase in match rates passed directly through the IER DS).

[4] The EMS is the software local areas use to process and store their Electoral Registers

2.3. The matching process

One of the main aims of the pilot was to assess the potential match rate that could be achieved by comparing DWP data with Electoral Register data. The matching itself was conducted within DWP using a matching algorithm specifically created for the pilots. The electoral registers from each area were matched against a snapshot of data from the DWP Customer Information System (CIS) which had been extracted at a similar time, ensuring that comparable data was used.

Prior to returning the results to Local Authorities the results of the DWP matching were converted into a simpler format which included applying a scoring algorithm which assigned a basic 'Red, Amber, Green' (RAG) rating to each record. A green match indicated a positive match, an amber match indicated a possible match and red match indicated that no match could be found.

In addition to an overall RAG rating, further contextual information was included in the match file detailing, for example, whether a record had failed on the address match or the identity match.

This process was carried out by the Government Digital Service (GDS) for the purposes of the pilot, using criteria developed by the Cabinet Office in conjunction with the EC and DWP.

The pilot matching and scoring algorithms, as well as the format of the match files, were informed by feedback collated as part of the pilots, as well as the learning from the 2011 data matching pilots. In addition, following completion of the original pilot activity further refinements have been made to the matching algorithm, using the learning from the pilots.

The results of this evaluation are primarily based on 'pilot algorithm' however the results of the matching undertaken with the 'refined algorithm' are also presented. Full details of the matching process can be found at Annex A , including details of the changes made to the refined algorithm.

DWP Customer Information System (CIS) data

The DWP data used for the matching was a snapshot of the CIS database which includes details of individuals appearing in databases kept by the Secretary of State for Work and Pensions for the purposes of social security.

The source CIS database is updated daily and includes a broad coverage of the population who are eligible to vote, including anyone who has been issued with a National Insurance Number (NINO).

However, it is important to note that whilst the database has a broad coverage, it remains reliant on individuals having a recent interaction with and/or informing DWP (or other departments which such as HMRC which feed into DWP CIS) of changes in their circumstances e.g. moving home.

Therefore, whilst an individual may appear in the database they may not appear at their current address. This is particularly relevant for data matching for the purposes of confirmation because of the limited personal identifiers in the data which mean that identity matching is reliant on accurate address information.

2.4. Assessing the accuracy of confirmation

Another key aim of the pilot was to test the accuracy of data matching for the purposes of confirmation. In order to do this each participating area was asked to provide two versions of their electoral register to be matched against DWP data - their pre-canvass register (taken just prior to the annual canvass[5]) and their post-canvass register (taken just after the completion of the annual canvass), the results of which were compared against each other.

The reason for undertaking this comparison is because the completeness of the register is known to decline during the electoral cycle[6]. Therefore, by comparing the results from the pre-canvass register (which is taken at the point when it is expected that the register will be at its least complete) and the post canvass register (when it is expected that the register will be at its most complete) it is possible to assess the proportion of individuals positively matched against DWP who were subsequently confirmed as resident at the same address during the annual canvass.

This provides an indication of the accuracy of the matching including the potential level of inaccuracies that might occur in any confirmation exercise as a result of population churn (i.e. where a positive match is made but becomes inaccurate because an individual has moved home in the period between the original matching being undertaken and the electoral register being published).

Furthermore, by undertaking the matching on both the pre-canvass and post-canvass register it is possible to undertake comparisons of the confirmation rate at differing points in the electoral cycle.

2.5. Local data matching

Where they had the capacity to do so, a number of pilot areas also opted to use locally held data sets (for example Council Tax data or Housing Benefit data) to conduct supplementary data matching. Comparing the match results against locally held data aimed to provide further insight into both the accuracy of the data matching and the potential of local data matching to add to the confirmation rate (by matching individuals who could not be found within the DWP data set).

This matching was conducted separately, within the individual pilot areas, and therefore the exact processes, including the matching criteria[7] and the data sources used, varied between areas.

2.6. Evaluating the process of confirmation

[5] During the annual canvass a voter registration form is provided to every household to complete to update their information on the Register of Electors.
[6] The Electoral Commission estimate that the completeness of the registers decline, on average, by around one percentage point a month from completion of the annual canvass ('Great Britain's Electoral Registers 2011', Electoral Commission, 2011).
[7] I.e. the data fields matched and the level at which a match was accepted.

In addition to assessing the potential match rate and accuracy of the matching, the pilots were designed to provide information on the process of confirmation and related implications for policy and practice. Throughout the pilots feedback from the participating areas was sought which was used to inform the development of the pilots and for evaluation purposes. This included:

- Work with five 'beacon' pilot sites on initial development of the matching algorithm and match file. These sites were provided with the initial match results returned from DWP (i.e. before the scoring algorithm was applied) and a sample of DWP records for comparison against their electoral registers and asked for feedback on the accuracy of the matching and the presentation of results. This feedback was used to help develop both the matching and scoring algorithm as well as the format of the match return and guidance, with the aim of ensuring the results were presented in as clear and accessible a format as possible.
- Cabinet-office led workshops which provided an opportunity for the Cabinet Office to: update pilot areas and other relevant parties (e.g. EMS suppliers) on progress; to gain feedback from attendees; and to provide a forum for pilot areas to share their experiences with other participating areas.
- Qualitative interviews conducted face to face with each of the pilot areas towards the end of the pilot to examine in more depth the views and experiences of pilot areas and to gain further insight into lessons that can be learnt for the future. These interviews were recorded, professionally transcribed and then analysed using a thematic matrix.
- Testing of the refined algorithm with six pilot sites. Based on the learning from the pilots DWP proposed some additional refinements to the matching algorithm which aimed to overcome some of the common issues which had been identified as causing records to fail to match. In order to test the validity of these changes a purposively sampled set of DWP data was returned to six EROs who then manually compared these records to their register entries to assess whether they considered them to be accurate matches. These results were collated and used to determine the final matching and scoring algorithms. (See Annex A for further details).

Chapter 3
Results

This chapter presents the results of the pilots. The first sections examine the match rates achieved in the pilot including a discussion of the variations in match rates observed and the factors that may be influencing that variation. The results of analyses assessing the accuracy of using data matching for the purposes of confirmation are then discussed. The final sections explore findings in relation the process of matching including the potential for supplementary local matching to be used to add to the confirmation rate.

3.1 Match rates

Pilot algorithm match rates

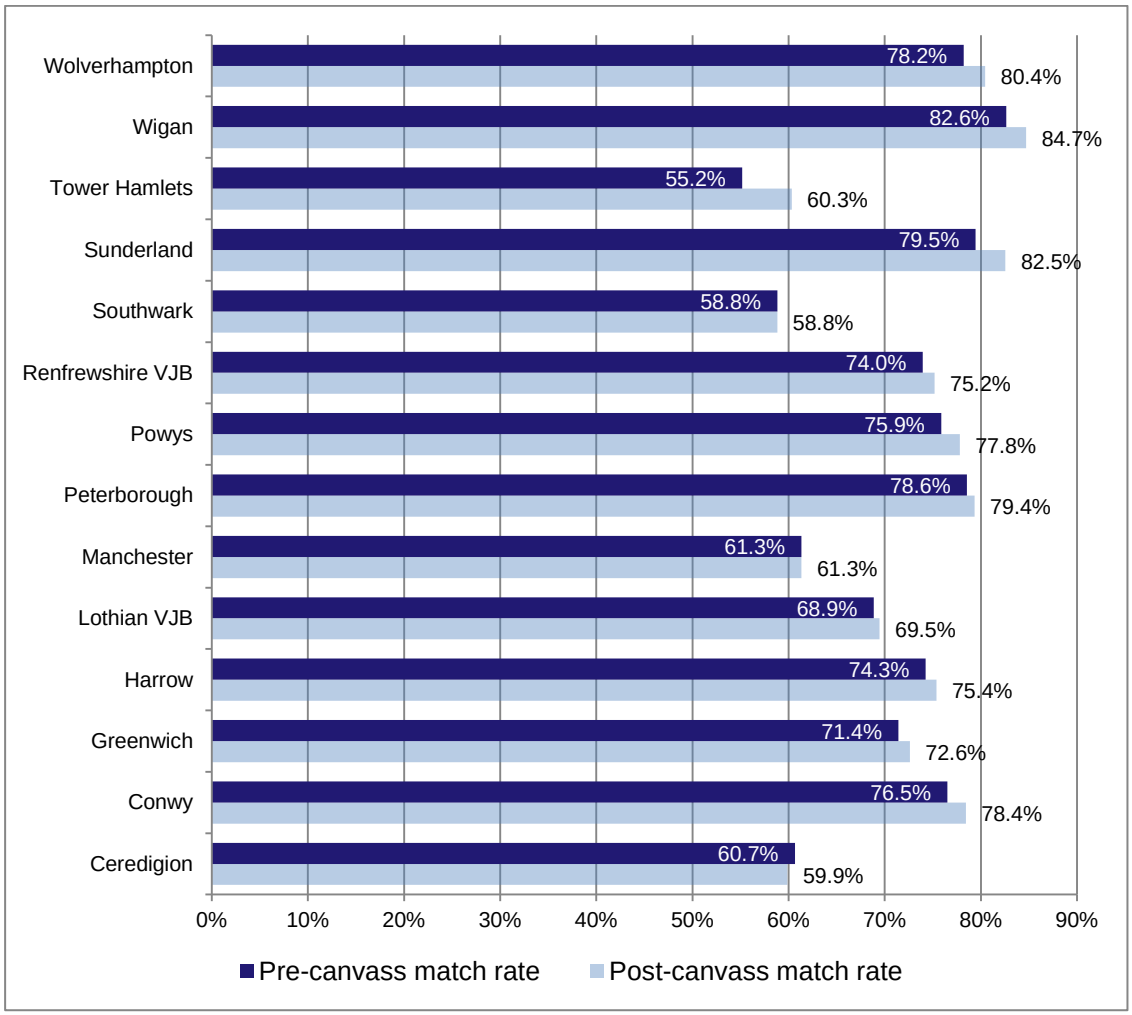
The match rate/confirmation rate is the proportion of individuals currently on the electoral register whose details could be positively confirmed within DWP data i.e the 'green' matches. This is the proportion of currently registered electors that have the potential to be transferred onto the new IER register without the need to re-apply with their personal identifiers[8].

Table 3a overleaf shows the match rate achieved for each pilot site based on the *pilot algorithm*. Results for both the pre-canvass and post-canvass registers are presented. It shows that overall 71 per cent of entries in the pre-canvass register could be positively matched (i.e. confirmed). Importantly, the results also show that the match rate varied considerably between areas, from 55 per cent in Tower Hamlets to 83 per cent in Wigan in the pre-canvass results. This is discussed more fully in the next section of the report.

The total match rate for the post-canvass register was slightly higher at 72 per cent (ranging from 59 per cent in Southwark to 85 per cent in Wigan). It may be expected that the confirmation rate will be highest for the electoral register that is published immediately following the annual canvass, where the electoral register is at its most complete and this can be seen in the results of the matching. However, overall the increase in match rates across the pilots were small (on average one per cent) and some areas (Manchester and Southwark) saw no change in their overall match rates, and one area (Ceridigion) experienced a drop in match rates (although by less than one per cent).

[8] Whilst these records can be automatically transferred onto the IER register it should be noted that it will ultimately remain the decision of the ERO to determine which entries are Confirmed.

Table 3a: Match rates by Local Authority using 'Pilot Algorithm'



The complete match results are presented in Annex B, including the proportion of records that could not be matched to DWP split by 'possible matches' (amber category) and no matches (red category). In practice these records will be required to go through the same checks and processes by EROs. Nevertheless, some areas have fed back that being able to identify entries where there was at least some indication of a match may help them to prioritise resources, therefore it is intended that the match reports will continue to make this distinction in live roll-out. Around three per cent of register entries matched using the pilot algorithm were identified as possible (but not certain) matches.

Refined algorithm match rates

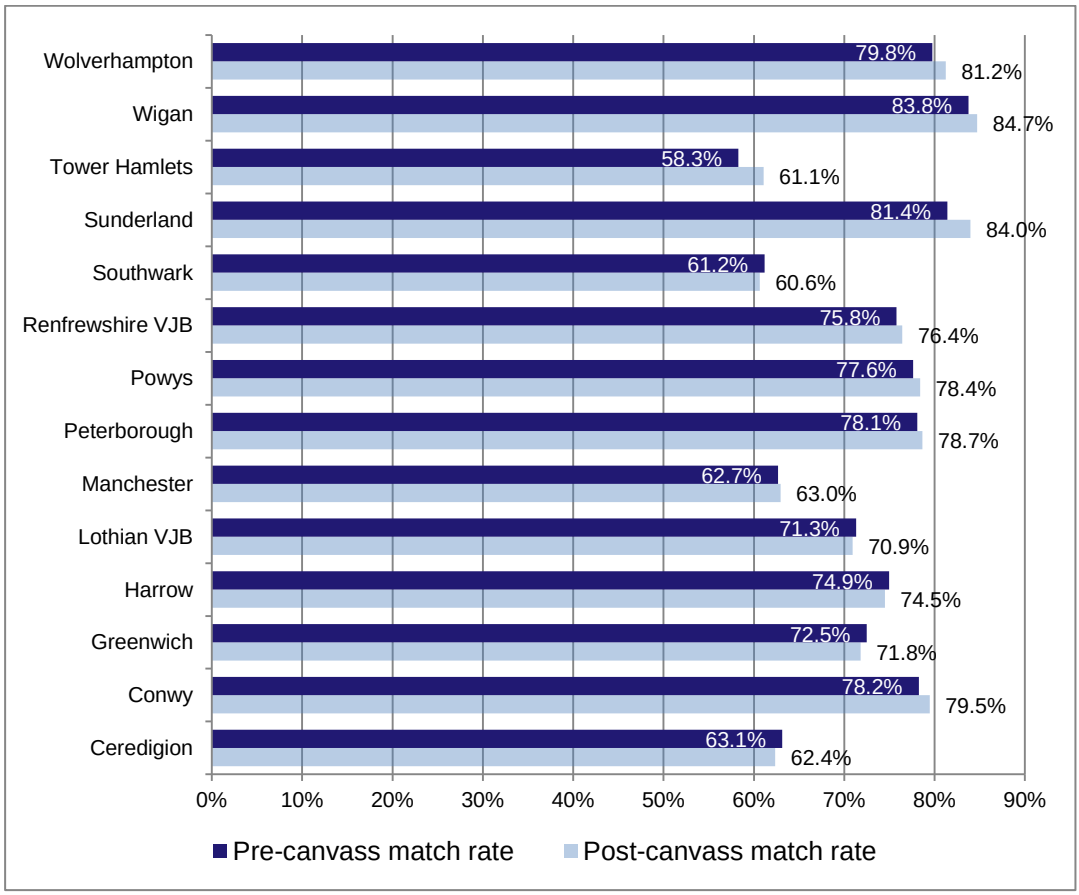
As previously noted, following the initial pilot activities and based on the learning from the pilots some further refinements were made to both the matching and scoring algorithm. These aimed to further improve both the quality and accuracy of the matching[9]. Table 3b shows the match rate achieved for each pilot site based on this refined algorithm, however it should be noted that the additional analyses of the data

[9] See Annex A for further details of the matching and scoring algorithms including details of the refinements made as a result of the pilots.

presented in this report are based on the match results achieved using the pilot algorithm.

In total 72 per cent of entries in the pre-canvass register could be positively matched (i.e. confirmed) using the refined algorithm, varying from 58 per cent in Tower Hamlets to 84 per cent in Wigan in the pre-canvass results. The overall match rate for the post-canvass register was 73 per cent (ranging from 61 per cent in Southwark to 85 per cent in Wigan). This represents an increase in the overall match rate from the pilot algorithm of just under two per cent for the pre-canvass registers and just around one per cent for the post canvass registers, although the match rate did not increase in all areas (range -0.5 per cent to 3%). In addition, the refined algorithm resulted in an increase in the numbers of records that were identified as possible matches, from around three per cent of records in pilot algorithm match files to around five per cent of records in the refined algorithm match files. Full details of the changes to the algorithm are detailed in Annex A and the full results of the matches using the refined algorithm are presented in Annex B.

Table 3b: Match rates by Local Authority using 'Refined Algorithm'



‘Carry-forward’ electors

Under the proposals for IER electors that have been ‘carried forward’ from the previous register have been included in the overall match rates presented above[10]. However, whilst carry-forward electors will be eligible for confirmation, unlike other confirmed electors they will not automatically be transferred onto the IER register unless their name has been included on a household enquiry form as part of the IER canvass.

This is important as it may impact on the overall confirmation rates because it cannot be assumed that the details of all of these electors will be returned on a household enquiry form. Research by the Electoral Commission on the 2010 registers showed that on average across Great Britain, between 3% and 4% of the electors have been carried forward but that this varies significantly from area to area[11].

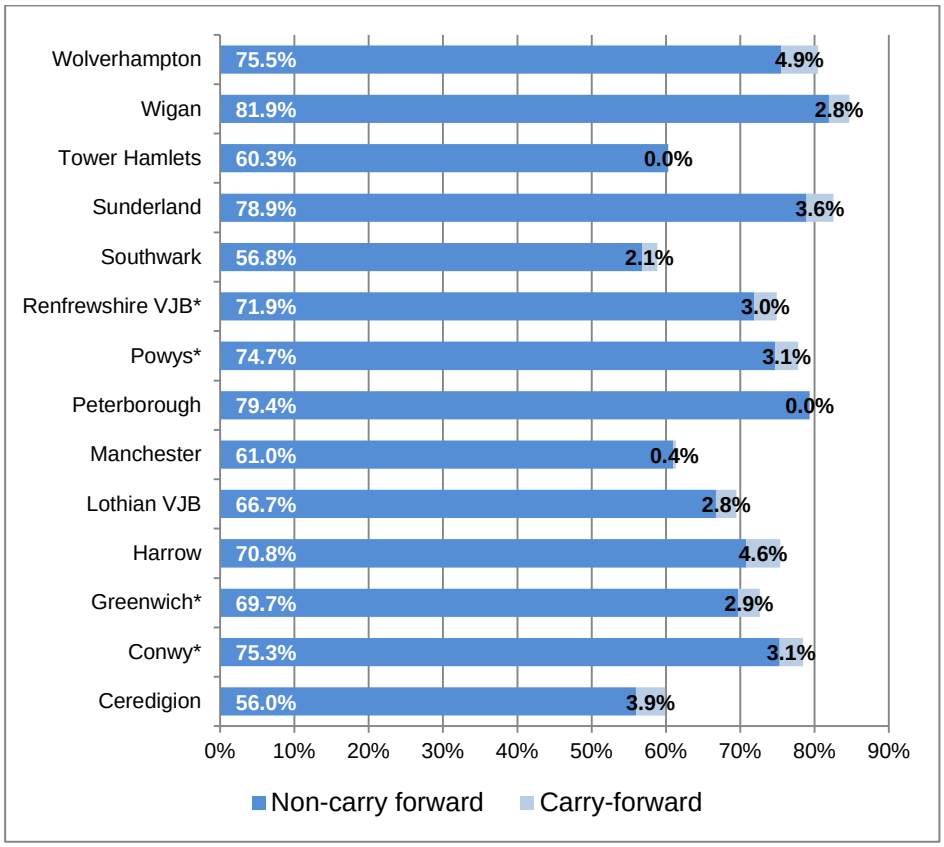
Where data was available for the pilot it showed that the rate of carry-forwards varied across the pilot areas from 0% to 14% across the whole register and from 0% to 9% of records that could be positively matched within DWP[12]. Figure 3c illustrates the proportion of the post-canvass match rate in each pilot area that is made up of carry-forwards (where data was unavailable average pilot figures have been applied for illustrative purposes). It illustrates that even under the worst case scenario (i.e. where none of the carry-forward electors responded to the household enquiry form and therefore could not be confirmed) it is estimated that on average 70% of the post-canvass register in the pilot areas would be confirmed (based on the pilot algorithm).

[10] Currently, where an ERO receives no response to the canvass from a household, and has not been able to confirm their details using other data, they may retain an electors’ details on the new register for one year through the process of carry forward. The carry forward was designed to give EROs the option of avoiding disenfranchising some residents as a result of their non-response to the canvass, however there will inevitably less certainty over the accuracy of register entries that have been carried forward in this way.

[11] In 2010, for example, it ranged from 0% to 17%. It also varies by authority type, with the median figures for 2007–10 ranging from 2.6% in English two-tier districts to 6.1% in Scotland. (‘Great Britain’s Electoral Registers 2011’, Electoral Commission, 2011)

[12] Not all files included accurate flags for carry-forwards records so data was not available for all pilot areas

Fig 3c: Post-canvass pilot match rates split by carry-forward and non-carry-forward electors



*Note: * = pilot average proportion of carry-forward figures applied as unable to accurately isolate carry-forward electors in the data files.*

3.3 Understanding the variation in match results

As described above there is significant variation in the match rates between areas. This is important as it has implications in terms of the potential completeness of the register and the levels of resource required in the transition to IER. Areas with lower match rates are likely to require greater resources, as the work involved in following-up electors who could not be confirmed will be greater than those who can be confirmed (e.g. issuing additional letters and using door to door canvassers). In summer 2013 every Local Authority in England, Wales and the Scottish Valuation Joint Boards will participate in a Confirmation Dry Run. This will provide the opportunity for each ERO to complete a fully IT enabled dry-run of the confirmation process and obtain indicative match results for their area. This information can then be used to assist in the planning and allocation of resources for the transition[13].

The results of the dry-run will provide a more complete assessment of the proportion of electors across England, Wales & Scotland that have the potential to be confirmed including the levels of variation in match rates. Nevertheless the pilots provided a valuable opportunity to explore some of the potential drivers of that variation, as discussed below.

[13] For example, these match rates will be used as part of the funding formula for determining funding allocations for Local Authorities.

Annex C provides a full breakdown of ward level match results and shows that ward level match rates varied significantly within pilot areas as well as across pilot areas. This suggests that factors other than differences in the way that local registers are maintained are influencing match rates. The variation in match rates within areas was highlighted in the 2011 data matching pilots, where it was suggested that it is likely to be explained by differences in population types. Feedback from the 2012 pilots further supports this. For example, participants in the pilot reported that areas that are known to have higher-population turnovers were also observed to have lower match rates, whilst areas with higher match rates tended to be those with more settled populations. Student areas and communal residencies were also observed as having lower match rates.

As discussed in previous sections, owing to the limited personal information currently held on the register confirmation matching is heavily reliant on matching address information. This is because the name is not unique enough on its own to enable accurate matching. Where an individual has recently moved it can be expected that the likelihood of either the register or DWP CIS containing out of date or inaccurate (and therefore conflicting) information may be greater, meaning that more mobile populations may be less likely to have their details matched.

Relatedly, for some segments of the DWP customer data base their interaction with DWP (or other Departments whose data feeds into CIS) is not reliant on having the correct address. This means that in some instances the likelihood of the DWP CIS database holding up to date address information will be linked to how proactive individuals in the population are at providing notification of their change of address. In the case of students, as well as being a relatively mobile population, many students have different home and term-time addresses and may not have cause to notify DWP or others of their term-time address (e.g. where they choose to use their home address for official correspondence).

Overall this suggests that the currency/accuracy of the address data on either database is likely to cause records to fail to match, and as a result certain groups are less likely to be matched under-confirmation. Comparisons of ward level match rates and data from the 2011 England and Wales Census provide further support for this. This analysis showed that areas with higher proportions of students (aged 16-74), higher proportions of people living in private rented accommodation and/or higher proportions of people living in communal establishments had relatively lower match rates[14]

In addition to the findings above it was also possible to isolate within the match files certain elector types, namely attainers[15] and postal voters. Comparing the match rates for these groups against other electors showed that the average pilot match

[14] Regression analysis showed that these three variables accounted for the majority of variation in match rates. The proportion of people living in private rented accommodation was the strongest predictor, however all variables were significantly related to the match rate. (Regression is a statistical analysis technique which takes a key variable of interest, in this case match rates, and measures the ability to predict the key variable by knowing the other variables). Due to the lack of availability of recent data on population mobility this was not specifically tested, however it is known that people living in private rented accommodation are more likely to move home more frequently than those living in owner occupiers/those renting from a council or housing association ('Great Britain's Electoral Registers 2011', EC, 2011) These analyses were only undertaken for pilot areas in England and Wales as recent comparable data for Scotland is not currently available

[15] An attainer is someone who will turn 18 during the life of the register

rate for attainers was higher than for non-attainers, on average eight per cent higher in the pre-canvass register (79% compared to 71%) and the pilot match rate for postal voters was on average eight per cent higher (77% compared to 71%). This provides a further indication that population characteristics influence match rates [16].

Whilst these findings suggest that variation in match rates is largely driven by population factors, other factors are also likely to have an influence. For example, as the data sets are not standardised to a single address reference list mismatches can occur because of differences in address structures. Feedback obtained during the process of developing the matching algorithm suggests this may have a particular impact on some address types – notably flats, sub-divided properties and Scottish tenement housing[17].

Overall however, the proportion of addresses that could be successfully matched across the registers was high[18] and some address cleansing/standardisation of records was undertaken as part of the matching process. In addition, as a result of the learning from the 2011 pilots the DWP data was populated with Unique Property Reference Numbers (UPRNs). This aims to minimise the overall address match failures by reducing the reliance on matching against address line text. Nevertheless, it will not be possible to eradicate all such issues as not all registers will necessarily contain up to date UPRNs, nor will DWP have accurate UPRNs for all addresses.[19]

Other factors that might influence match rates include differences in the format of names which may lead to failures on the identity matches. For example, some pilot areas raised concerns that mismatches may occur where individuals prefer to be known by their middle name as opposed to first name and therefore their names may appear differently on the register and DWP CIS. However, it was not possible to isolate these cases in the data in order to ascertain whether this was in practice having any impact. In addition, the results of the comparison of ward level match rates against data from the 2011 England and Wales Census indicated that there was a slightly lower match rates in areas with higher proportions of Welsh language speakers. This could indicate that the matching algorithm could be less effective at matching Welsh names and/or addresses, and therefore it will be important to monitor this carefully as part of the Confirmation Dry-Run.

---

[16] Whilst it is not possible to determine the exact causes for this, potential contributory factors include: attainers may be less mobile than other groups e.g. where still living with their families and dates-of-birth are stored on the register for attainers which may improve their likelihood of data matching. Postal voters tend to be older and move less frequently than the general population.

17 These address structures may be more complex giving greater room for discrepencies e.g. Flat A, Block 2, Smith House, 23 John Street vs Smith House Flat A2, John Street.

18 In the post-canvass match file overall 96% of records could be matched at the address level (range 92% to 98%)

[19] UPRNs are standardised unique identifiers for each land and property unit and are heavily used by EROs to conduct their current activity. 13/14 pilot areas had registers that were populates with UPRNs. UPRNs are assigned to every unit of land and property recorded by local government who have a statutory obligation to record all changes in property details, including all new builds.. Overall DWP were able to allocate UPRNs to 88% of the data extracted for the pilots (range 44%-97% per pilot area), this work has been ongoing which is expected to increase the overall UPRN allocation (comparable figures for the pilot areas are expected to be 94% overall, range 82%-98%), however, it is not possible to successfully match all addresses to the centrally held data on UPRNs and differences in the frequency within which UPRN data in DWP and the local registers are updated will also lead to some discrepencies between data sets.

Importantly, during the course of the pilot some additional issues specific to individual pilot areas were identified which had the potential to impact on their match rates[20]. . These issues proved relatively simple to address but demonstrate an additional advantage of the Confirmation Dry-Run in enabling such issues to be tested and wherever possible resolved in advance of the live run.

3.4 Testing the accuracy of the matching

The completeness of the register is known to decline during the electoral cycle and is expected to be at its most complete and accurate just after completion of the annual canvass[21]. Therefore, by tracking the match results of records from the pre-canvass register in the post-canvass match file it is possible to make an assessment of the accuracy of the matching process.

The full results of this analysis are provided at Annex D. Overall they showed that on average ten per cent of individuals on the pre-canvass register were no longer registered at the same address in the post-canvass register (range 4 to 26 per cent)[22]. Looking at just those individuals who were positively matched against DWP CIS data (see Table 3d overleaf), on average four per cent were no longer registered at the same address post-canvass (range two to fourteen per cent[23]). A further one per cent were still registered at the same address but could no longer be matched within DWP data.

This suggests that the risk of inaccuracies resulting from confirmation (as measured by post-canvass responses) is small. Overall around 95 per cent of individuals in the pilot areas who were confirmed through data matching against DWP CIS in the pre-canvass register were confirmed at the same address during the annual canvass.

[20] For example, one pilot area's original EMS extract did not pick up UPRNs because they were held in a different field within their EMS to other areas

[21] The EC estimate that the completeness of the registers declines, on average, by around one percentage point a month from the completion of the annual canvass 'Great Britain's Electoral Registers 2011', Electoral Commission, 2011

[22] Figures exclude Manchester and Ceredigion as it was not possible to accurately match the pre-and post canvass registers for these areas because of inconsistencies in the data.

[23] Notably the rates of inaccuracy, as measured by the post-canvass match results, were highest in London: Greenwich 7%; Harrow 6%; Southwark 5%; and Tower Hamlets 17%. However it should be noted that Tower Hamlets experienced a relatively large (6%) drop in the overall size of their register, which will have influenced this result (The percentage change in size of the registers for other pilot areas ranged from +/- 2%).

Table 3d: Post-canvass match results of positively matched entries on the pre-canvass register

	Proportion (%) confirmed in canvass and still match to DWP data	Proportion (%) confirmed in canvass but no longer match to DWP data	Proportion (%) removed from the register at that address
Conwy	96.8%	0.2%	3.0%
Greenwich	92.9%	1.1%	6.0%
Harrow	94.0%	0.8%	5.2%
Lothian VJB	95.0%	1.8%	3.2%
East Lothian	95.0%	1.6%	3.4%
Edinburgh	94.4%	2.2%	3.4%
Midlothian	96.0%	1.2%	2.7%
West Lothian	96.0%	1.3%	2.7%
Peterborough	97.4%	0.1%	2.5%
Powys	97.6%	0.1%	2.4%
Renfrewshire VJB	96.4%	1.0%	2.6%
East Renfrewshire	96.6%	0.8%	2.6%
Inverclyde	96.1%	1.5%	2.4%
Renfrewshire	96.3%	0.9%	2.7%
Southwark	95.1%	1.8%	3.2%
Sunderland	97.1%	0.1%	2.8%
Tower Hamlets	84.2%	1.6%	14.2%
Wigan	96.7%	0.0%	3.3%
Wolverhampton	96.7%	0.1%	3.2%
Total	95.2%	0.9%	3.9%
Average	95.0%	0.7%	4.3%

Notes: 1) Manchester and Ceredigion are excluded from this analysis as difficulties in matching the pre and post canvass files together mean that we cannot be confident in the accuracy of the results 2) Carry-forward records were excluded from this analysis where possible because these electors have not responded to the canvass. The pilot areas for whom it was not possible to isolate (and therefore exclude) carry-forward figures were Greenwich, Lothian, Powys & Renfrewshire. 3) Averages calculated using combined figures Scottish Valuation Joint Boards.

In addition to enabling an assessment of the accuracy of the matching for those individuals that could be matched using DWP data, comparisons between the pre and post canvass match results also provide useful information about the results where an individual failed to be matched in the DWP data. This analysis shows that around three in four individuals on the pre-canvass register who could not matched within DWP data were subsequently confirmed as being resident at the same address during the annual canvass.

As discussed previously, whilst the DWP data set has a broad coverage it remains reliant on individuals providing updates of any changes in their circumstances (e.g. moving home) or having a recent interaction which leads to an update on DWP CIS. Therefore, whilst an individual may appear in the DWP CIS data, they may not appear at their current address which is required for effective data matching for the purposes of confirmation. This means that whilst data matching against DWP can confirm the majority of electors, there will remain a proportion of electors who are

accurately included on the register but who will not be positively matched in the confirmation exercise.

3.5. Local Matching

Under the current system EROs are able to access other locally held information, such as council tax records, in order to assist in compiling the registers. This secondary checking is used to determine whether a resident is still living at an address in order to keep them on the register. During the transition to IER EROs will be able to use the results of this local matching to confirm individuals who could not be matched within DWP, adding to the overall confirmation rate in their area.

Whilst it was not a compulsory part of the pilot, where they had the capacity to do so, pilot areas were invited to use locally held data sets (for example Council Tax data or Housing Benefit data) to conduct similar secondary data checking. Six pilot areas carried out such matching, the results of which suggest that local matching (primarily against Council Tax records) added between seven and fifteen per cent to their overall match rates. As these results are taken from a small number of areas (who will have adopted different approaches to the matching) they cannot be considered representative, nevertheless they show that local matching has clear potential to add to the confirmation rate.

Research by the EC suggests that eighty-nine per cent of EROs indicated that they made some use of secondary checking in 2010[24] and many of the pilot areas fed back that they intend to use local matching for this purpose during the transition to IER. However, many of these will carry out such checking manually and it cannot be assumed that all areas currently have the technical resource available within their electoral services teams to undertake this work.

Whilst pilot areas generally saw the opportunity to complement the confirmation exercise with local matching as positive, a number areas also highlighted what they perceived to be a delicate balance between ensuring that there was consistency in the approach to local matching during transition, whilst maintaining the autonomy of individual EROs to determine the most effective way to use this data (based on their assessment of the quality and ease of access to different data sets in their area).

3.6 Other process issues

Guidance

Data collected during the qualitative interviews with pilot areas showed that overall pilot areas were confident in matching but emphasised the importance of having transparency in the matching process, to ensure that EROs could feel confident in accepting the matches. This was seen as particularly important given that EROs do not have access to the DWP data and therefore cannot see the CIS record which DWP matched against in order to make their own judgement on the match.

---

[24] 'Great Britain's Electoral Registers 2011', Electoral Commission, 2011

For some areas this raised the question of whether the 'amber' category of match added much value, as without being able to see the DWP records they had no easy way of verifying the match. (Whilst local data matching can be used this applies equally to red and amber matches). However, other areas felt that retaining the amber category would be beneficial in enabling them to prioritise their resources. This suggests that the amber category remains useful, but further emphasises the need for clarity in the guidance for EROs on interpreting the match files.

Resource

The levels of Local Authority/VJB resource required for the pilots varied significantly between pilots. However, this was largely a reflection of the differing levels of involvement in the pilot. This included the amount of additional testing and analyses undertaken to assist the development of the matching and scoring algorithm, as well as whether any local matching was undertaken. It is therefore difficult to isolate the costs that were specific to the pilots and those which would reflect the resources required in live running.

During the qualitative interviews pilot areas were asked for their feedback on the likely resources required for the conducting the process of confirmation. The responses showed that these similarly varied. In most cases this was related to the processes employed for analysing the data and for local matching. For example, whether areas had prior experience of local matching and access to staff with expertise in matching or whether this resource had to be bought-in or undertaken manually.

Chapter 4
Summary and conclusions

As a result of pilot studies carried out by the Cabinet Office in 2011, the potential for matching the electoral register against data held by DWP in order to confirm the majority of existing entries on the electoral register was identified. By removing the requirement for these individuals to provide the additional personal identifiers that will be required under IER the process of confirmation has the potential to simplify the transition for the majority of existing electors, enabling resources to be focussed on the minority of electors who cannot be confirmed in this way, as well as those eligible electors who are not currently registered to vote. However, as earlier pilots had not set out to test this specifically further testing was recommended. This report presents the results of pilots undertaken across 14 areas to test the effectiveness and reliability of the policy of Confirmation.

Key findings & implications for policy and practice:

Overall, the results of the final evaluation support the findings presented in our preliminary evaluation report, demonstrating that the policy of Confirmation provides a reliable and effective method of transferring existing electors onto the new IER register, offering the opportunity to simplify the process for the majority of electors.

Match rates & accuracy

- *Overall over 70 per cent of electors in the pilot areas could be matched within DWP CIS data and findings showed that we can be confident in the accuracy of this matching - the vast majority of electors matched from the pre-canvass register (95%) were subsequently confirmed as resident at the same address in the annual canvass.*
- *There is however significant variation in Confirmation rates across areas, which has subsequent impacts in terms of the levels of resource required by different EROs during the transition to IER.*
- *The Confirmation Dry Run will provide an essential opportunity for the match rate for all EROs to be tested for the purposes of resource planning and will be used as part of the Cabinet Office funding formula for Local Authorities.*

Following initial pilot activities, and based on the learning from the pilots, some additional refinements were made to the pilot matching and scoring algorithm. Based on this refined matching algorithm, results show that overall 72 per cent of existing electors in the pilot areas could be positively matched against DWP data in the pre-canvass register and 73 per cent could be matched in the post-canvass register. Match rates varied between pilot areas, ranging from 58 per cent to 84 per cent in the pre-canvass match files and 61 per cent to 85 per cent in the post canvass match files[25] . This has subsequent impacts in terms of the levels of resource required by different EROs during the transition to IER - areas with lower match rates are likely to require greater resources.

Whilst these results cannot be considered representative of all areas they provide a strong indication of the likely confirmation rate that can be achieved by matching against DWP data. In summer 2013 every Local Authority will be required to participate in a Confirmation Dry Run (CDR) which will enable each ERO to complete a fully IT enabled dry-run of the confirmation process and to obtain indicative match results for their area. The results of this exercise will be used to help determine the CO funding allocations and enable EROs to effectively plan for the transition. [26]

Assessments of the accuracy of the data matching, undertaken by comparing the results of data matching undertaken prior to the annual canvass with results immediately following the canvass, also demonstrate that we can be confident in the accuracy of the matching. Of the 71 per cent of electors that were confirmed in the pre-canvass register (using the pilot algorithm), 95 per cent were subsequently confirmed as resident at the same address during the annual canvass.

Understanding the variation in match rates

- *Variations in match rates are likely to be driven in large part by population characteristics.*
- *With the notable exception of attainers who have higher match rates than other electors, a number of the groups that have traditionally been less likely to register appear less likely to be confirmed (i.e. students, private renters and people living in communal establishments).*
- *Whilst these individuals will still receive invitations to register under IER, ensuring that resources are effectively targeted on the groups that have lower confirmation rates will be an important part of planning for the transition. Wider activities to maximise registration amongst these groups may also complement this activity.*
- *It is not possible to determine the exact causes for these groups having lower confirmation rates, however the accuracy/currency of address information held in the data sets is likely to be a key factor*.

[25] This compares to 71 per cent in the pre-canvass register (range 55% to 83% across pilot areas) and 72 per cent in the post canvass register (range 59%-85%) based on the original pilot algorithm.

[26] The pilot areas included areas with relatively high populations of groups that analyses shows are less like to confirm therefore it is possible that the national match rate will be higher than that of the pilots.

Participants in the pilot reported that variations in match rates appear to be linked to population factors, for example areas that are known to have higher-population turnovers were also observed to have lower match rates, whilst areas with higher match rates tended to be those with more settled populations. Student areas and communal residencies were also observed as having lower match rates.

Owing to the limited personal information currently held on the register confirmation matching is reliant on matching address information, which may lead to failed matches even where an individual's details are held on DWP CIS. Where an individual has recently moved it can be expected that the likelihood of either the register or DWP CIS containing out of date or inaccurate (and therefore conflicting) information may be greater, meaning that more mobile populations may be less likely to have their details matched.

In addition, for some segments of the DWP customer data base their interaction with DWP (or other Departments whose data feeds into CIS) is not reliant on having the correct address, meaning that the accuracy of DWP CIS address information will be reliant on individuals in the population providing notification of their change of address. In the case of students, as well as being a relatively mobile population, many students have different home and term-time addresses and may not have cause to notify DWP or others of their term-time address (e.g. where they choose to use their home address for official correspondence).

Overall this suggests that the currency/accuracy of the address data on either database is likely to cause records to fail to match, and as a result certain groups are less likely to be matched under-confirmation, including more mobile populations. Comparisons of ward level match rates and data from the 2011 England and Wales Census provide further support for this, showing that areas with either higher proportions of students (aged 16-74), higher proportions of people living in private rented accommodation and/or higher proportions of people living in communal establishments had relatively lower match rates These groups overlap with those that have been identified in previous research as less likely to be registered, although not all traditionally under-registered groups are less likely to confirm. For example, attainers were found to have higher match rates than other electors.

Whilst the currency/accuracy of address information is important other factors will also impact on match rates. For example, whilst the DWP CIS database has broad coverage some individuals will not appear in the data set and issues with data quality and the lack of standardisation between data sets may impact on match rates[27].

[27] There is also some evidence that Welsh language may have a small impact on match results but it is not possible to be conclusive based on the pilot results and therefore this will be monitored carefully in the Confirmation Dry-Run.

Local data matching

- *Whilst data matching against DWP CIS data can confirm a majority of electors, some electors who are accurately included in the register will not be able to be matched within DWP data. Supplementary data matching against local data sets may be a useful tool for confirming additional electors who could not be matched within DWP data.*

- *Whilst many EROs already conduct similar local matching as part of their usual canvass activities, it cannot be assumed that all currently have the capability to do so. Developing guidance and/or sharing best practice across EROs is therefore likely to be important.*

As detailed above, whilst matching against DWP data is able to confirm a majority of existing electors, there remains a significant minority of people whose details are accurately included on the register but will not be able to be confirmed in this way. For example, comparisons of the pre and post canvass match results highlighted that around three in four individuals on the pre-canvass register who could not matched within DWP data were subsequently confirmed as being resident at the same address in the canvass.

Results from the pilot suggest that supplementary data matching against local data sets may be a useful tool for confirming additional electors who could not be matched within DWP data[28]. However, whilst many areas already use local data as part of their usual canvass activities, not all currently have the capability to do so giving rise to the potential for inconsistencies in approaches across areas[29].

As the availability and quality of local data sets is likely to vary, as is the additional resource required to undertake such an exercise, the ability to determine the extent to which local data matching is used remains at the discretion of the ERO. However, learning from the pilots emphasises the need for ERO guidance to incorporate detail on how local matching could be used, including for example how the quality of data sets can be assessed. The Confirmation Dry Run also provides a valuable opportunity to further develop our understanding of local matching capability and identify best practice (where areas chose to do so)[30]..

---

[28] Whilst it was not a compulsory part of the pilot, where they had the capacity to do so, pilot areas were invited to use locally held data sets (for example Council Tax data or Housing Benefit data) to conduct similar secondary data checking. Six pilot areas carried out such matching, the results of which suggest that local matching (primarily against Council Tax records) added between seven and fifteen per cent to their overall match rates.

[29] In addition, many areas who currently conduct local matching may do so manually.

[30] There is no requirement for EROs to conduct local matching as part of the CDR or beyond, however the CO aims to use this as an opportunity to ask EROs about what local matching is currently undertaken/planned and where possible to identify examples of best practice.

Annex A – Data Matching methodology

The initial matching of the data was undertaken by DWP using a matching algorithm created by their Information, Governance and Security team. The database against which the electoral register is matched is the DWP Customer Information System database. CIS is an amalgamated data source, consisting of information received from internal DWP heritage systems, as well as other government sources, such as HMRC. As a result CIS is seen within DWP as being the master of customer information.

The source CIS database is updated daily, however for the purposes of the pilot matching was undertaken on snapshots of the data extracted at a similar time to the electoral registers to ensure comparability.

The DWP matching algorithm works like a filter, the stages of which can be broadly summarised as:

1. Some data standardisation of both the electoral register and the DWP is undertaken to make the records more consistent (e.g. Str. or St. to Street).
2. The algorithm searches for a matching address in the DWP data. In the first instance through Unique Property Reference Number (UPRN), then, if a match cannot be found through UPRN, by comparing the lines of the address.
3. If an address can be located the algorithm then compares the name fields to those of the individuals held in the DWP CIS database at that address.

The end result is a series of match statements that describe the levels at which a record has ‘passed’ or ‘failed’ the different matching criteria.

It is on the basis of these statements that the scoring algorithm was then applied, assigning a RAG category to the address, identity and then applying an overall RAG. This part of the process will be part of the IER DS in live running, and was carried out by GDS for the purposes of the pilot.

The RAG status was assigned according to the following criteria:

Address RAG

Any DWP address match was assigned a ‘green’ RAG. The minimum match criteria for an address is complete postcode + the numeric from address lines one/two. Details were also provided to the ERO as to whether the address match had been made using the UPRN or through a straight address match.

Identity RAG

A positive ‘green’ identity match was assigned to records that matched at any one of the following levels:

- Full first name **plus** full last name
- Full last name **plus** first three initials of first name

- As above but including middle name/initial/DOB where available and matched
- Date of birth (DOB) **plus** full lastname
- Full last name **plus** soundex[31] match on first name **plus** either middle name **or** middle initial **or** DOB
- Soundex match on last name plus first name **plus** middle name **or** middle initial **or** DOB
- Full last name and full first name but reversed

A possible 'amber' identity match was assigned to records that did not match at the levels above but did match at one of the following levels as minimum:

- Soundex match on last name **plus** DOB **or** first three initials of first name
- Full last name **plus** soundex match of first name

The experience of the pilots was that the vast majority of records where a positive identity match could be found were matched on at least the full last name and full first name (approx 95%).

Overall RAG

The final match rate was calculated using the overall RAG, which in most cases was the same as the identity match, apart from the following exceptions:

a) Where DWP had a record of the individual being deceased the overall rating defaulted to 'red'

b) Where DWP had a record of the individual being older than 100 the overall rating defaulted to 'red'[32]

c) Where the DWP match has returned more than one 'best match' for a record with a property the overall rating defaulted to 'amber.'

As detailed in the methodology section, feedback from the pilot areas was used to inform the development of both the matching and scoring algorithm, including working with five Beacon pilots at the start of the pilot to refine the algorithm and the presentation of the match files.

Refined algorithm – overview of changes

For the purposes of the pilot the matching/scoring algorithms were frozen at the point of the pre-canvass matching to ensure that the pre and post canvass registers could be accurately compared. However, based on the feedback from areas obtained during the pilots some further refinements and testing of some of the elements of the algorithms was undertaken. The key elements are summarised below:

a) Address matching - refinements to the matching algorithm to try and widen the address matching criteria. Whilst, in the pilot areas the vast majority of records

[31] A soundex match is a match made using a phonetic algorithm for indexing names by sound, as pronounced in English, so that they can be matched despite minor differences in spelling. The SOUNDEX algorithm is seen to English biased and is less useful for languages other than English.

[32] This age was set as a proxy for the pilots, in go-live this will be aligned with information known about the oldest living person.

(96%) did pass the initial address matching stage[33] feedback also suggested that some specific address types e.g. sub-divided properties were less likely to match. Therefore DWP explored ways to loosen the criteria for the address matching in an attempt to enable more addresses to be matched. This was done by allowing records that could not be matched at other stages to pass the address match stage where a combination of the postcode and full last name matched the electoral register record. Testing undertaken across 6 pilot areas, which included individually reviewing 823 records matched on this criteria found that 78 per cent of these matches were deemed accurate. Whilst this suggests that the majority of these records can be correctly matched it remains a significantly lower accuracy level than other matches[34] and therefore, in the refined algorithm, anything matching on this address category is defaulted to an 'amber' overall RAG.[35]

b) Multiple matches – as detailed above the pilot algorithm gave any records that had been matched with more than one record on the DWP database a default rating of 'amber', however a number of pilot areas queried whether this was the correct approach as they did not feel that the fact that more than one DWP record could be matched against the register entry weakened the validity of the match. Testing undertaken across 6 pilot areas, which included individually reviewing 446 records matched on this criteria, found no significant difference in the accuracy of the matching of these records compared to other positive matches. Therefore for the refined algorithm the overall RAG for these records will no longer default to an overall RAG of amber.

c) Last name plus first three initial matches – these records were given a 'green' in the pilot algorithm, but some areas expressed uncertainty about this classification as there is the potential for inaccuracies, for example "Stephen" and "Stephanie" could be matched under this category despite being different people. Testing undertaken across 6 pilot areas, which included individually reviewing 675 records matched on this criteria found that 94.4 per cent of these matches were deemed accurate. Whilst this was a slightly lower result than other 'green' match categories in the pilot algorithm (98.4% of which were deemed accurate based on individually reviewing 1,062 records) the results suggest that including this category (on which approximately 3% of register entries matched in the pilot) would not weaken the accuracy of the matching significantly and therefore these records continue to be assigned a 'green' RAG in the refined algorithm.

In addition, DWP have made some additional changes to the pilot matching algorithm for the 'refined algorithm' including:

- The standardisation aspects of both name and address fields have been extended to incorporate additional strings such as *'Apartments'*, *'Buildings'* as

[33] Range 92-98%.

[35] Two additional stages of address matching which further weakened the address match (for example by using post code stub as opposed to full postcode) were also tested but the results showed that the majority of these matches were inaccurate and therefore they were discounted.

well as taking account of the variation in non-english names such as potentially 10 variations of Mohammed

- The address detail matching has been extended to incorporate the fact that either within the register or within the DWP CIS data there is a potential for relevant information to be present within the Address line 3 data. This is particularly relevant for places such as care homes, where the first line of the address may contain the name of the care home, which then pushes the content of the physical address further across the address fields.
- The extract of DWP CIS data matched against excludes individuals under 16 years of age who would not be eligible to register.

Annex B – Match results

Table B1: Pilot algorithm match results by Local Authority

	Pre-canvass register match results				Post-canvass register match results			
	Green (Positive match)	Amber (Possible Match)	Red (No match)	Total records matched	Green (Positive match)	Amber (Possible Match)	Red (No match)	Total records matched
Ceredigion	60.7%	2.3%	37.1%	58,985	59.9%	2.3%	37.8%	59,145
Conwy	76.5%	2.6%	20.9%	91,966	78.4%	2.6%	19.0%	93,144
Greenwich	71.4%	2.6%	25.9%	171,905	72.6%	2.7%	24.7%	173,529
Harrow	74.3%	3.5%	22.3%	179,173	75.4%	3.5%	21.1%	178,937
Lothian VJB	68.9%	2.8%	28.3%	600,192	69.5%	2.5%	28.1%	613,500
East Lothian	*75.7%*	*2.9%*	*21.4%*	*76,214*	*77.5%*	*2.6%*	*19.9%*	*77,290*
Edinburgh	*61.8%*	*2.3%*	*35.9%*	*333,606*	*61.9%*	*1.9%*	*36.2%*	*342,342*
Midlothian	*78.7%*	*3.6%*	*17.6%*	*63,403*	*80.2%*	*3.2%*	*16.5%*	*64,396*
West Lothian	*78.4%*	*3.8%*	*17.8%*	*126,969*	*79.4%*	*3.4%*	*17.2%*	*129,472*
Manchester	61.3%	3.0%	35.6%	369,996	61.3%	3.0%	35.7%	370,959
Peterborough	78.6%	2.9%	18.6%	138,464	79.4%	2.9%	17.8%	135,487
Powys	75.9%	1.8%	22.4%	103,072	77.8%	1.8%	20.4%	103,381
Renfrewshire VJB	74.1%	3.8%	22.1%	263,598	75.2%	4.0%	20.8%	262,001
East Renfrewshire	*79.6%*	*3.7%*	*16.7%*	*70,019*	–	–	–	–
Inverclyde	*69.8%*	*4.2%*	*26.0%*	*62,002*	–	–	–	–
Renfrewshire	*73.2%*	*3.7%*	*23.2%*	*131,577*	–	–	–	–
Southwark	58.8%	4.2%	36.9%	202,918	58.8%	4.2%	36.9%	204,574
Sunderland	79.5%	3.6%	16.9%	218,445	82.5%	3.8%	13.7%	216,295
Tower Hamlets	55.2%	4.5%	40.4%	171,055	60.3%	4.9%	34.8%	161,472
Wigan	82.6%	2.4%	15.0%	242,973	84.7%	2.5%	12.9%	245,077
Wolverhampton	78.2%	3.0%	18.8%	177,306	80.4%	3.2%	16.4%	178,230
Total	70.6%	3.1%	26.3%	2,990,695	71.9%	3.1%	24.9%	2,995,731
Average	71.1%	3.1%	25.8%		72.6%	3.1%	24.3%	

Notes: 1)Average figures based on total match rate for Scottish VJBs 2)Renfrewshire VJB data unavailable at LA level for post-canvass match and LA level results for pre-canvass based on data use for pre-post comparisons resulting in a small discrepancy in total data although this will not impact overall results. 3)Tower Hamlets pre-canvass register extract excluded UPRNS (included in post-canvass)

Table B2: Refined algorithm match rates by Local Authority

	Pre-canvass register match results				Post-canvass register match results			
	Green (Positive match)	Amber (Possible Match)	Red (No match)	Total records matched	Green (Positive match)	Amber (Possible Match)	Red (No match)	Total records matched
Ceredigion	63.1%	4.3%	32.5%	58,985	62.4%	4.3%	33.3%	59,145
Conwy	78.2%	3.5%	18.3%	91,966	79.5%	3.7%	16.8%	93,144
Greenwich	72.5%	3.1%	24.4%	171,905	71.8%	3.1%	25.1%	173,529
Harrow	74.9%	2.7%	22.4%	179,173	74.5%	2.7%	22.8%	178,937
Lothian VJB	71.3%	7.4%	21.3%	600,198	70.9%	7.6%	21.5%	613,500
East Lothian	*77.8%*	*3.6%*	*18.6%*	*76,214*	*78.3%*	*3.7%*	*18.0%*	*77,290*
Edinburgh	*64.4%*	*11.1%*	*24.5%*	*333,606*	*63.6%*	*11.3%*	*25.1%*	*342,342*
Midlothian	*81.3%*	*2.6%*	*16.1%*	*63,403*	*81.5%*	*2.7%*	*15.9%*	*64,396*
West Lothian	*80.6%*	*2.5%*	*16.9%*	*126,975*	*80.8%*	*2.5%*	*16.7%*	*129,472*
Manchester	62.7%	3.0%	34.3%	369,997	63.0%	2.9%	34.2%	371,295
Peterborough	78.1%	2.3%	19.6%	138,464	78.7%	2.3%	19.1%	135,487
Powys	77.6%	4.1%	18.3%	103,072	78.4%	4.1%	17.5%	103,381
Renfrewshire VJB	75.8%	7.6%	16.6%	264,245	76.4%	7.7%	15.8%	262,001
Southwark	61.2%	4.9%	33.9%	202,918	60.6%	4.8%	34.5%	204,574
Sunderland	81.4%	1.8%	16.8%	218,445	84.0%	1.9%	14.2%	216,295
Tower Hamlets	58.3%	7.3%	34.5%	171,055	61.1%	4.7%	34.3%	161,472
Wigan	83.8%	1.3%	15.0%	242,971	84.7%	1.3%	14.0%	245,083
Wolverhampton	79.8%	3.1%	17.1%	177,306	81.2%	3.2%	15.6%	178,230
Total	72.3%	4.5%	23.2%	2,990,700	72.8%	4.4%	22.8%	2,996,073
Average	72.8%	4.0%	23.2%		73.4%	3.9%	22.8%	

Notes: 1) Averages calculated using totals for Scottish VJBs. 2) Renfrewshire VJB data unavailable at Local Authority level. 3)Tower Hamlets pre-canvass register extract excluded UPRNS (included in post-canvass)

Annex C – Analyses of variation in match rates

Ward level match rates[36]

As detailed in Chapter 3, in order to explore the variation in match rates within areas analyses of the ward level match results were undertaken, the results of which are presented in Table C1 below.

Table C1: Ward level variation in match rates by Local Authority (based on pre-canvass match file using the pilot matching algorithm)

Local Authority[1]	Overall match rate for Local Authority (%)	Highest green match rate at Ward Level (%)	Lowest green match rate at Ward level (%)
Ceredigion[2]	60.7	77.7	18.1
Conwy	76.5	82.2	62.2
East Renfrewshire	78.2	66.7	52.5
Greenwich	71.4	82.1	59.3
Harrow	74.3	79.7	62
Inverclyde	68.1	74.8	69.8
Manchester[2]	61.3	78.6	24.9
Peterborough	78.6	87	65.9
Powys	75.9	83.5	69.7
Renfrewshire	71.2	81.4	76.4
Southwark	58.8	70.5	48.7
Sunderland	79.5	84.8	65.2
Tower Hamlets	55.2	63.4	46.4
Wolverhampton	78.2	85.5	59.9
Wigan	82.6	86.7	79.2

Notes: 1) Results for Lothian were excluded from this analysis as the record level file did not include accurate Ward codes (although it should be noted that this was present in the original extract provided by the Lothian VJB) 2). Only 6 wards across the pilot areas presented had match rates of less than 40%, all of these Wards also had high proportions of the population who were full-time students (aged 16-74), ranging from 49% to 69%.

I

[36] A ward is a geographical division/district into which boroughs are divided for administrative/political purposes.

Annex D – Pre and post canvass register comparisons

Table D1: Post-canvass match results of pre-canvass register entries by ERO and RAG classification

	Post-canvass match result for records that were positive (green) matches in the pre-canvass register				Post-canvass match result for records that were possible (amber) matches in the pre-canvass register				Post-canvass match results for records that did not match (red) in the pre-canvass register			
	Green (Positive match)	Amber (Possible Match)	Red (No match)	Not on the post-canvass register	Green (Positive match)	Amber (Possible Match)	Red (No match)	Not on the post-canvass register	Green (Positive match)	Amber (Possible Match)	Red (No match)	Not on the post-canvass register
Conwy	96.8%	0.0%	0.2%	3.0%	1.4%	94.5%	0.2%	4.0%	1.2%	0.1%	69.4%	29.3%
Greenwich	92.9%	0.0%	1.0%	6.0%	2.2%	90.6%	0.6%	6.6%	3.9%	0.1%	63.1%	33.0%
Harrow	94.0%	0.1%	0.7%	5.2%	2.0%	89.9%	0.4%	7.7%	2.5%	0.1%	65.1%	32.3%
Lothian VJB	95.0%	0.1%	1.7%	3.2%	15.4%	79.6%	1.1%	3.8%	4.0%	0.1%	67.8%	28.1%
East Lothian	95.0%	0.1%	1.5%	3.4%	12.8%	82.3%	1.0%	3.8%	6.2%	0.1%	69.2%	24.5%
Edinburgh	94.4%	0.1%	2.2%	3.4%	18.6%	75.9%	1.7%	3.8%	3.3%	0.1%	76.6%	20.0%
Midlothian	96.0%	0.0%	1.2%	2.7%	13.7%	81.6%	0.8%	3.9%	5.7%	0.2%	73.7%	20.4%
West Lothian	96.0%	0.1%	1.2%	2.7%	12.5%	83.3%	0.5%	3.6%	5.5%	0.1%	76.9%	17.4%
Peterborough	97.4%	0.0%	0.1%	2.5%	1.0%	94.2%	0.1%	4.7%	0.8%	0.1%	86.6%	12.5%
Powys	97.6%	0.0%	0.1%	2.4%	0.5%	96.1%	0.1%	3.4%	0.6%	0.0%	78.4%	20.9%
Renfrewshire VJB	96.4%	0.1%	0.9%	2.6%	1.3%	94.5%	0.4%	3.8%	2.3%	0.1%	78.5%	19.2%
East Renfrewshire	96.6%	0.1%	0.7%	2.6%	0.6%	96.7%	0.2%	2.5%	2.8%	0.1%	78.1%	19.0%
Inverclyde	96.1%	0.3%	1.2%	2.4%	1.3%	96.2%	0.3%	2.2%	2.0%	0.1%	79.8%	18.1%
Renfrewshire	96.3%	0.1%	0.8%	2.7%	1.4%	96.1%	0.3%	2.3%	2.4%	0.1%	78.1%	19.5%
Southwark	95.1%	0.1%	1.7%	3.2%	1.4%	93.1%	1.5%	4.0%	3.1%	0.2%	81.3%	15.3%
Sunderland	97.1%	0.0%	0.1%	2.8%	0.7%	95.3%	0.1%	3.9%	1.2%	0.1%	56.3%	42.4%
Tower Hamlets	84.2%	0.1%	1.5%	14.2%	2.6%	80.2%	1.2%	16.0%	9.7%	0.7%	45.4%	44.2%
Wigan	96.7%	0.0%	0.0%	3.3%	0.5%	94.0%	0.0%	5.5%	0.0%	0.0%	61.6%	38.4%
Wolverhampton	96.7%	0.0%	0.1%	3.2%	0.9%	95.5%	0.1%	3.5%	1.0%	0.1%	63.2%	35.7%
Total	95.2%	0.0%	0.8%	3.9%	4.1%	90.1%	0.6%	5.2%	3.4%	0.2%	67.4%	29.0%
Average	95.0%	0.0%	0.7%	4.3%	2.5%	91.5%	0.5%	5.6%	2.5%	0.1%	68.1%	29.3%

Notes: 1) Manchester and Ceredigion are excluded from this analysis as difficulties in matching the pre and post canvass files together mean that we cannot be confident in the accuracy of the results 2) Carry-forward records were excluded from this analysis where possible because these electors have not responded to the canvass. The pilot areas for whom it was not possible to isolate (and therefore exclude) carry-forward figures were Greenwich, Lothian, Powys & Renfrewshire. 3) Averages calculated using combined figures Scottish Valuation Joint Boards.4) Results based on data files created by matching records across the pre and post canvass match files.This process resulted in a small differences in total records between files, but these were not significant enough to impact on overall results