Assessing Coverage and Quality of the 2007 Prototype Census Kidlink Database

Adela Luque
U.S. Census Bureau

&

Deborah Wagner
U.S. Census Bureau

Center for Administrative Records Research and Applications
U.S. Census Bureau
Washington, D.C. 20233

Assessing Coverage and Quality of the 2007 Prototype Census Kidlink Database

Adela Luque, U.S. Census Bureau
Deborah Wagner, U.S. Census Bureau

Abstract

The Census Bureau is conducting research to expand the use of administrative records data in censuses and surveys to decrease respondent burden and reduce costs while improving data quality. Much of this research (e.g., Rastogi and O'Hara (2012), Luque and Bhaskar (2014)) hinges on the ability to integrate multiple data sources by linking individuals across files.

One of the Census Bureau's record linkage methodologies for data integration is the Person Identification Validation System or PVS. PVS assigns anonymous and unique IDs (Protected Identification Keys or PIKs) that serve as linkage keys across files. Prior research showed that integrating "known associates" information into PVS' reference files could potentially enhance PVS' PIK assignment rates. The term "known associates" refers to people that are likely to be associated with each other because of a known common link (such as family relationships or people sharing a common address), and thus, to be observed together in different files. One of the results from this prior research was the creation of the 2007 Census Kidlink file, a child-level file linking a child's Social Security Number (SSN) record to the SSN of those identified as the child's parents.

In this paper, we examine to what extent the 2007 Census Kidlink methodology was able to link parents SSNs to children SSN records, and also evaluate the quality of those links. We find that in approximately 80 percent of cases, at least one parent was linked to the child's record. Younger children and noncitizens have a higher percentage of cases where neither parent could be linked to the child. Using 2007 tax data as a benchmark, our quality evaluation results indicate that in at least 90 percent of the cases, the parent-child link agreed with those found in the tax data. Based on our findings, we propose improvements to the 2007 Kidlink methodology to increase child-parent links, and discuss how the creation of the file could be operationalized moving forward.

I. Introduction

The Census Bureau is researching ways to expand the use of administrative records (AR) data in censuses and surveys to decrease respondent burden, reduce costs and improve data quality. This research requires linking AR data across sources and to surveys and censuses (e.g., Rastogi and O'Hara (2012), Luque and Bhaskar (2014)). The process of linking individual data from different sources requires common identifiers unique to each person record. The Person Identification Validation System (PVS), one of the Census Bureau's record linkage applications, creates anonymous unique person identifiers (called Protected Identification Keys or PIKs) that serve as linkage keys to facilitate record linkage and unduplication within and across files. Expanding PVS' capability to achieve higher PIK assignment rates is, thus, an important objective.

The PVS uses a probabilistic matching algorithm to assign PIKs to persons in AR, survey or census files by comparing personally identifiable information (PII) such as name, date of birth (DOB), and address contained in those files to PII on person records in reference files. Reference files contain information from the Social Security Administration (SSA) enhanced with address data obtained from federal AR files (including Internal Revenue Service (IRS) and Medicare data)). Reference files contain all variants of a person's name, DOB, and sex, as well as current and recent addresses. For more information on the PVS, see Wagner and Layne (2014).

While PVS has high PIK assignment rates in many federal files, there are records that are not assigned a PIK. Prior research examined the possible use of "known associates" to help PVS obtain higher PIK coverage rates.[1] Their research showed that integrating known associates information into PVS' reference file could potentially enhance PVS' PIK assignment rates. The term "known associates" refers to people that are likely to be associated with each other because of a known common link, and thus, to be observed together in different files. Examples of known links include people residing within the same housing unit and family relationships. In particular, one of the end products of this prior research was a file comprised of parent-child relationships - the 2007 Census Kidlink file. Specifically, the 2007 Census Kidlink is a child-level file where the child's SSA Numerical Identification System or Numident[2] record has been linked to the SSN of those identified as the child's parents.[3]

Traditionally, PVS has not relied on child-parent links and has utilized reference files with address information that is current or recent (i.e., time dependent).[4] In contrast, the Census Kidlink provides information on "time-independent" parent-child associations that can be integrated in PVS' reference files to enhance PVS' ability to assign PIKs.

---

[1] NORC and GUNNISON contractors conducted PVS research in 2012 and 2013. The reports produced were for internal use only.

[2] The Numident is the SSA's database of the information contained in an application for a United States SSN.

[3] While SSA periodically provides the Census Bureau with a Numident file containing the SSN of the child and the name of his/her parents, this file does not provide the parents' SSNs.

[4] PVS has recently implemented algorithms based on the "known associates" methodologies developed by NORC and Gunnison (provided the input file has household information). However, these algorithms are time-dependent – as are the other algorithms traditionally employed in PVS.

In addition, by providing child-parent or intergenerational links, the Census Kidlink file will support research that relies on such linkages. For instance, Census Kidlink may help link parent historical data (such as characteristics contained in past decennial censuses) to a child's SSN record. This is particularly important because the household or child-level data available to researchers at a given point in time from a particular data source may not provide information about the parent(s) of the child. This could be due to changes in household or family composition since the birth of the child or simply because the data source at hand only provides information on the child.

In this paper, we describe the construction of the 2007 Census Kidlink file, evaluate the coverage and quality of child-parent SSN links, and discuss potential ways to increase such coverage. We start by describing the data and methodology used in the creation of the file (sections II and III respectively). In section IV, we examine the extent and pattern of child-parent links in the Kidlink file, and explore underlying child characteristics associated with them. In the same section, we also investigate how the methodology for creating the file could be improved to increase child-parent links, and how the creation of the file could be operationalized moving forward. In section V, we evaluate the quality of the child-parent links in the 2007 Kidlink by comparing/matching child-parent links in the Kidlink file to the child-parent links observed in the 2007 IRS and 2010 decennial Census files, independently. The quality evaluation also explores child demographic characteristics associated with the matched and unmatched cases. Understanding the quality of the child-parent links in the Kidlink file is important for any future implementation of this file in record linkage. If the child-parent link information is used in a reference file, then the quality of the matches is dependent on the quality of the reference data. Lastly, in section VI, we discuss next steps in light of our findings.

II. Data

This section describes the data used in the creation of the 2007 Census Kidlink file (Kidlink henceforth) as well as the data employed in the coverage and quality evaluations. The creation of the Kidlink file required two sets of data: the 2007 Numident and the 2007 PVS Geobase reference file. We also use two additional data sources in the quality evaluation of Kidlink: the 2007 IRS file and the 2010 Decennial Census file. These files are used as benchmarks against which to compare the child-parent links in the 2007 Kidlink file.

Numident

The SSA Numident file contains all transactions ever recorded against any single SSN - with each entry representing an addition or change (such as name changes) to the SSN record. This file is edited to create the Census Numident, which contains one record for each SSN.[5] Each SSN record in the Census Numident contains name, DOB, sex, race, place of birth, parents' name, citizenship status and date of

---

[5] Name edits, DOB reconciliation, and race identifiers are some of the edits conducted to produce this Numident file. This work is conducted by the Census Bureau's Center for Administrative Records Research & Applications (CARRA). The resulting Numident file contains the most recent name and DOB data. All alternate name and DOB data are kept on separate files.

death information.[6]

We use this information to examine the extent of child-parent links found, the quality of those links, and how they vary by the age, sex and citizenship status of the child. The citizenship status of the child is captured in a categorical variable that takes three separate values: 'U.S. citizen/unknown', 'noncitizen-authorized' and 'noncitizen-other'.[7] A child is classified as 'noncitizen-authorized' if the child ever was a noncitizen authorized to be in the U.S.. The 'noncitizen other' category includes all other noncitizens.

PVS Geobase Reference File

As mentioned earlier, the PVS is an application of a probabilistic matching algorithm that assigns PIKs to person records in an incoming file (e.g., AR, survey or census files) by comparing PII such as name, DOB, and address contained in that file to PII on person records in reference files.

Reference files contain data from the Numident file enhanced with address data obtained from federal AR files.[8] The reference files, thus, contain all variants of a person's name, DOB, and sex, as well as current and recent addresses. These reference files are referred to as the (PVS) Geobase reference file since addresses (a geographic component) are appended to each person record.[9] It is important to note that there are multiple Geobase reference files that are created depending on the vintage of the incoming file to be processed through PVS.

2007 IRS Data

The IRS 1040 data for tax year 2006 (or 2007 IRS) provides data for individuals who file a Form 1040 tax return. These data include all returns received by the IRS and include the mailing address on the return (generally as of around April 15, 2007), the name and SSN for the primary filer, and the name and SSN for any spouse and/or up to four dependents on the form.

A given tax return groups or associates children with adults' SSNs that are likely to be their parent(s), thus providing child-parent links against which the 2007 Kidlink file links will be evaluated.

2010 Decennial Census Data

The 2010 Decennial Census file contains the decennial census unedited data for the 50 United States, the District of Columbia, and Puerto Rico. The file contains name, DOB, gender, relationship, race, Hispanic origin, Master Address File Identifier (MAFID), and address. The MAFID serves as the household unit identifier for the Census file. Since the decennial census does not ask for respondent

---

[6] The Numident's data on race is limited. The SSA halted collection of race data when it transitioned to the Enumeration at Birth system in 1987. Prior to that, race categories did not remain consistent over time.

[7] The Numident refers to these groups as 'U.S. citizen/unknown', 'alien' (including legal alien, alien student and conditional alien), and 'other alien' respectively.

[8] Namely, data from the IRS, Department of Housing and Urban Development, several files from the Department of Health and Human Services, and Selective Service.

[9] In particular, the address data is cleaned and standardized and used to construct a variable called GEOKEY. The GEOKEY variable is constructed as a subset of the full address, and then is appended to the Numident data to create the PVS Geobase Reference file.

SSNs, PIKS were converted to SSNs for this analysis. Not every record on the 2010 Census file received a PIK - only 90.8 percent of person records were assigned one.[10]

Through this process, the 2010 Census provides child-parent links as defined by child and adult(s)[11] living in the same MAFID or household unit. For the quality evaluation analysis, it is important to remember, that not every person in the 2010 Census file has a PIK/SSN, and that the Census child-parent links may not be as accurate as those in the IRS file since PIK assignment is based on a probabilistic matching algorithm (PVS). In addition, there is a time frame misalignment between the 2007 vintage of the Kidlink file and the 2010 Census data. As a result, the child may not be observed with his/her father and/or mother in the same MAFID due to changes in family/household composition. For instance, the father of the child may have relocated due to divorce after 2007, so the Kidlink child may be observed in the 2010 Census file with a male adult that is not his/her father. These factors affect the quality of the Census child-parent links, and will have to be taken into consideration when using them as a benchmark against which Kidlink child-parent links are tested.

III.   Kidlink Construction

As mentioned earlier, the Numident file provided to the Census Bureau by the SSA contains parent's name information on a child's SSN record, but does not provide the parent's SSN. The 2007 Census Kidlink file was, then, created using the 2007 Numident and the 2007 Geobase to tie or link child SSNs to that of their parent(s). The creation of the 2007 Census Kidlink involved a number of steps that we now describe.

The first step in the construction of Kidlink was to create the universe of eligible records from the 2007 Numident. All persons were considered "children" and in scope for the Kidlink build process as long as their year of birth was not missing and there was no date of death information. Of the 441,829,508 records on the 2007 Numident, 385,079,988 records were considered eligible for the Kidlink file.

In order for a given child's SSN to be linked to the SSN of his/her parent(s), the child SSN was first located in the Geobase reference file, containing addresses.[12] All addresses for the located child, and all other person records associated with those addresses (including names) were extracted. A modified version of one of PVS' modules[13] was then used to (probabilistically) match mothers and fathers' names

---

[10] It is important to emphasize that any research conducted at the Census Bureau involving individuals uses PIKs (not SSNs) in order to protect the anonymity and confidentiality of person records. As an exception, the current internal analysis uses SSNs since it is precisely child-parent SSN links that are being evaluated.

[11] Only the householder and the spouse (if any) of the householder are considered.

[12] In a prior attempt, the parent's SSN was searched just using the parents name data from the Numident child's record. However, this search using just name and gender information did not provide a great degree of certainty for the links. In particular, this process resulted in too many ties. That is, when trying to find a match for certain, more common names, the matching process gave rise to too many individuals with that name being matched, but did not provide a way to identify the 'true' match to the actual mother/father.

[13] Specifically, a modified version of PVS' Geosearch module was used in the name match. See Layne & Wagner (2014) for more information about PVS' methodology and its modules.

in the Numident children records to the names extracted from the Geobase reference files. The Appendix contains a step-by-step explanation of this process.

If the names of both parents in the child Numident record were successfully matched to (or found in) the Geobase, then the SSNs of both parents were linked to the child's. Analogously, if only the name of the mother/father was found in the Geobase, then the Kidlink record of that child contains only the mother/father's SSN. If the name of either parent was not successfully found in the Geobase, then the Kidlink child record does not contain either parent. We will refer to these groups of Kidlink children as the 'Both Parents' category, the 'Only Mother' category, the 'Only Father' category, and the 'Neither Parent' category respectively. We will refer to these groups collectively as the "parent categories".

The analytic Kidlink extract file used in this paper includes Kidlink children that are 19 or younger as of 2007. This includes a total of 82,955,825 child records out of the 385,079,988 person records considered eligible for the Kidlink file. The decision to restrict the sample to those 19 or younger is motivated by the fact that the chances of finding a child living at the same (2007) Geobase address as his/her parents diminishes greatly as the child gets older and becomes an adult.

IV.   Coverage of Parent Links in Kidlink

This section discusses the number and percentage of Kidlink children records for whom a parent link was found, and examines what child characteristics are associated with having a parent link. It also identifies potential ways to increase child-parent links.

A. Findings & Recommendations

Table 1 shows the number and percentage of Kidlink child records in the 'Both Parents', 'Only Mother', 'Only Father' and 'Neither Parent' groups, overall and by child age and citizenship status. Unfortunately, the SSA halted collection of race data when it transitioned to the Enumeration at Birth system in 1987. Applications for a child's SSN after then would not include the race of the child in the Numident. For this reason, we cannot examine the relationship between child race and the presence of parent links in Kidlink child records using Numident data.

Results indicate that overall, links were found for both parents in almost 43 percent of Kidlink child records. In 29.5 percent of cases, only a link for the mother was found and in approximately 8 percent of cases, only the father's was found. This implies that in approximately 80 percent of cases, at least one parent was successfully linked to the child record. Possible causes for the inability to find a parent for a given child include the child or the parent not being in the federal files used to construct the Geobase, moving activity, and/or changes in the child's family/household composition - which may have prevented a match to the appropriate record in the Geobase.

Notice that the presence of parent links varies by the age of the child. In general, the percent of missing parent links is higher for younger children. In particular, the percent of children less than 1 year old with no parent link is almost 43 percent – a 19.7 percentage point difference from the 1-year olds. This is

likely due to the "tax code effect". This refers to the fact that most records in a given Geobase come from tax data and that there is a lag for the youngest children in tax data; that is, children born in the current year are not claimed in the previous year tax returns. For instance, children born in 2007 would not be claimed in 2006 taxes, and therefore, many of these children would not be found in the 2007 Geobase. In addition, IRS 1040 data used in this project only includes up to four dependents. If a family with more than four dependents enumerates the older dependents first, the younger children would not be in the tax data, and thus, the Geobase.

Finding parent links also varies considerably by the citizenship status of the child.[14] The percent of children for whom no parent link is found is higher for the noncitizen-other (66.2 percent) and noncitizen-authorized (26.0 percent) groups than for the Citizen/Unknown group (19.5 percent). Parents of children who are non-U.S. citizens may have Individual Taxpayer Identification Numbers (ITINs)[15] or may be less well-represented in Geobases. The Geobase used in the creation of the 2007 Kidlink does not contain ITINs, so any parent with an ITIN would not have been found in the Geobase.[16] Using Geobases with ITINs would help improve the linkability between children and parents.

Similarly, the percentage of cases in the 'Only Father' group is higher for children in the noncitizen groups. It is possible that fathers (rather than mothers) are better represented in Geobases. For instance, fathers in these groups may tend to participate in the labor force more than mothers, and thus, be better represented in tax data.

We find that out of the 22,998,800 Kidlink child records where the mother cannot be found, the mother's name is missing in the Numident for only 57,628 children or 0.3 percent (data not shown).[17] That is, in a very small percentage of cases does a missing mother's name in the Numident child record account for a missing mother link in the Kidlink file. In the case of fathers, out of the 40,901,854 children without a father in the Kidlink file, 7,149,313 (or 17.5 percent) had a missing father's name in their Numident record (data not shown). While this percent is higher than for mothers, a missing father name in the Numident does not account for the majority of missing father links in Kidlink.

A.1. Drill-down on Kidlink Children in the 'Neither Parent' Category

Table 2 looks more closely into Kidlink children records in the 'Neither Parent' category. For this group, it shows the percentage of children that are found in the 2007 Geobase, the percent that is not found in the 2007 Geobase but is present in the 2013 Geobase, and the percent that cannot be found in either the 2007 or the 2013 Geobase. This categorization allows, to some extent, for the quantification of the

---

[14] The Numident only contains information about the citizenship status of the record holder, not the parents. Therefore, it is beyond the scope of this paper to examine cases where the U.S. citizenship status of the child differs from that of the parents. This issue will be explored in future research.

[15] An ITIN is a tax processing number only available for certain nonresident and resident aliens, their spouses, and dependents who cannot get a Social Security Number.

[16] Children who are U.S. citizens may also have parents with ITINs.

[17] We consider a parent name to be missing if either the first name or the last name or both are missing.

tax code effect (i.e., those in the 2013 Geobase but not in the 2007 Geobase). The percentages are also shown by child age and citizenship status.[18]

A total of 8,542,312 children constitute the category of Kidlink children found in the 2007 Geobase, but without a parent link (column 2, Table 2). They represent almost 52 percent of the 'Neither Parent' group (see column 5). One of the main reasons why children records found in the 2007 Geobase have no parent links in the Kidlink file is that the child and the parents do not share the same physical or tax-return address as of 2007.[19] This is more likely to be the case as the child gets older and there is a higher likelihood of changes in the household composition. This is consistent with the percentages in column 5. In general, the percent of Kidlink children in this category increases with age. Specifically, the percent of 18, and particularly 19 year olds in the 2007 Geobase 'Neither Parent' group is higher than for younger groups – in fact, many children in this age group may have already formed their own households.

Employing Geobases from multiple years in the construction of the Kidlink file would help ameliorate this issue. Using a series of years would enable us to observe a given child over time as changes in his/her household composition and moving activity occur, and would increase the probability of finding the child with his/her parent(s) at a given address in the Geobase. The use of Geobases from years prior to the Numident year used in the construction of Kidlink would be of particular importance for older children. That is, Geobases from prior years would allow us to observe the older children when they were younger, at a time when they would have been more likely to live in the same household as one or both parents.

Another reason why children records end up with no parent links is that even when the child is in the Geobase, the parent may not be because he/she does not have an SSN. This would include, for instance, children whose parents have ITINs. For this reason, including ITINs in the Geobases used in the construction of Kidlink would help increase the chances of finding child-parent links for this group of children. Furthermore, making use of decennial censuses in the construction of Kidlink files at the turn of the decade might also increase child-parent links. This might prove particularly useful for populations that are less represented in tax data, and thus, Geobases.

As expected, the percent of children in the 2007 or 2013 Geobases is highest for U.S. citizens, followed by the 'noncitizen-authorized' group, and then followed at a distance by the 'noncitizen-other' group. Children who are in the U.S. citizens and noncitizen-authorized groups are expected to be better represented in Geobases than children in the noncitizen-other group. Note that over one half of children in the 'noncitizen-other' group cannot be found in either Geobase (see column 7).

A total of 5,374,378 children (almost 33 percent of those in the 'Neither Parent' category) constitute the category of Kidlink children without a parent link that can be found in the 2013 Geobase, but not in the

---

[18] As already mentioned, the current analysis is limited by the fact that the Numident does not contain other demographic characteristics.
[19] A child may not live with his/her parent, but still be part of the same tax unit because, for instance, the mother or father may still claim the child as a dependent.

2007 Geobase[20] (column 3, Table 2). They represent children that could potentially be linked to a parent if an additional Geobase from a more recent year is used in the construction of Kidlink.[21] Also note how the percent of children in this category is higher for younger children, and in particular the youngest ones (column 6). Specifically, the percent of 'Neither Parent' 0 year olds who are not in the 2007 Geobase but are in the 2013 Geobase (column 6, Table 2) is over 30 percentage points higher than the next age category. Meanwhile, the percent of 'Neither Parent' 0 year olds in the 2007 Geobase (column 5, Table 2) is much lower than those children at least 1 year old. These results can be largely attributed to the tax code effect. That is, many Kidlink children less than 1 year old that are not in the 2007 Geobase because of the lag in tax data can be found in the 2013 Geobase. Just as before, these findings suggest that the use Geobases from multiple years in the construction of the Kidlink file would increase the number of children records for whom a parent link could potentially be found.

Overall, most Kidlink children records in the 'Neither Parent' category (almost 85 percent) can be found in either the 2007 or 2013 Geobase – only 15.4 percent of them cannot be found in either Geobase. There is a much higher potential for finding at least one parent for children found in a Geobase than for those not found in either Geobase. Therefore, it is important to examine the potential reasons for why a child might not be found in a Geobase.

A total of 2,531,132 children (15.4 percent of those in the 'Neither Parent' category) constitute the category of Kidlink children that cannot be found in either the 2007 or 2013 Geobase. A likely reason for why the child cannot be found in the Geobase is that the child him/herself may simply not be in the AR files that make up the Geobase.   Another potential reason is that the child may come from a household with a large number of dependents. This is because, as already mentioned, tax data are an important source of Geobase records and the tax records data available contain up to just four dependents per household. A preliminary match of the children not found in either Geobase to the 2009 IRS electronic file[22] reveals that for the children found in both the Kidlink file and the 2009 IRS electronic return file,[23] approximately 25 percent of the households where these children reside have more than four dependents. Finally, it is also possible that the use of Geobases from multiple years (beyond just 2) might help finding parents for the group of Kidlink children in the 'Neither Parent' category.

---

[20] Recall that the Kidlink file used in this study makes use of just the 2007 Geobase.

[21] This number is slightly lower if we discount mothers and fathers with missing names in the Numident. Out of the 5,374,378 children, 4,843 lack a mother's name in the Numident, 567,641 lack a father's name, and 580,480 lack both parents' names.

[22] The 2009 IRS electronic file contains primary SSN, secondary SSN for up to 20 dependents. The full names of the dependents are also available.

[23] 3.5 percent of children not found in the 2007 and 2013 Geobases were successfully matched to the 2009 ELF IRS file.

B. Characteristics of Kidlink Children by Parent Category

This subsection presents information about the age and citizenship status of Kidlink children by parent category.[24] Kidlink children in the 'Neither Parent' category are further divided into two groups: those found in either the 2007 or 2013 Geobase, and those not found in either Geobase.

Table 3 shows the age and citizenship status distributions of Kidlink children by parent category. Column 9 reveals that compared to the 'Both Parents' group, the 'Only Mother', 'Only Father' and 'Neither Parent – no Geobase' groups tend to have lower proportions of younger children. Also, the 'Neither Parent – in Geobase' group has a higher proportion of 0 year olds relative to the 'Both Parents' group. Column 9 shows that compared to Kidlink children in the 'Both Parents' group, children not found in either Geobase tend to have a higher proportion of children who are not U.S. citizens. Also note that Kidlink children in the 'Only Father' category tend to have a higher proportion of children who are authorized noncitizens compared to 'Both Parents' children (see column 6).

C. Coverage Analysis Conclusions

The analysis revealed a number of ways to potentially increase child-parent links in future Kidlink files. One, using Geobases of multiple vintages will likely increase child-parent links for the youngest as well as the older children. Using Geobases from vintages *prior* to the Numident year will help improve child-parent links for older children by finding child and parent living at the same address - as moving activity and changes in household composition are more likely as the child gets older. Using Geobases from vintages subsequent to the Numident year (whenever available) will ameliorate the tax code effect and will help increase child-parent links for younger children (over 5 million children that were not in the 2007 Geobase were found in the 2013 Geobase). In addition, it would also be helpful to include the previous Decennial Census in the list of reference files used to construct Kidlink– particularly when creating a Kidlink file at the turn of the decade.

In some cases, multiple SSNs were found in a Geobase for a given parent. Father's names were more likely to result in such cases, possibly due to first names being passed down from one generation to the next. Also, the age range for possible candidate matches was wider for fathers than for mothers, leading to more such cases. The researchers that constructed the 2007 Kidlink dropped these cases. However, using Geobases from vintages prior to the Numident year would help identify the actual father/parent. In those earlier Geobases, the child would be too young to be confused for a parent, and the chances of finding the very young child with his/her actual parent(s) would be higher.

---

[24] The analysis is limited to the reliable demographic characteristics available in the Numident. The 2007 Kidlink file was also matched to the 2010 decennial census in an attempt to obtain other demographic characteristics, such as race and Hispanic origin. Unfortunately, due to the time frame misalignment between the vintage of the Kidlink file (2007) and the 2010 census, the matched data produces a biased sample. The youngest Kidlink children found in the 2010 Census are at least 3 years old. Compared to those that cannot be found in the 2010 Census, Kidlink-Census children have a much higher percentage of 'Both Parent ' (46.3 vs. 28.5 percent), and also a much lower percentage of 'Neither Parent' (15.8 vs. 36.2 percent).

Adding other data sources containing child-parent pairs with address information may also increase child-parent links. This would include sources such as Supplemental Nutrition Assistance Program (SNAP) AR data. As already mentioned, employing decennial censuses in the construction of Kidlink files at the turn of the decade might also increase child-parent links.

In addition, further research would also benefit from using a Numident file of the same vintage as the 2010 Decennial Census year. This would make possible an analytic sample that would include the youngest children - and thus, less sample bias. Having a richer set of child characteristics, such as race and Hispanic origin, would shed more light onto the factors underlying the presence (or lack thereof) of parent links in Kidlink child records and the likelihood of finding the parent for a given child.

Another potential improvement to the construction of Kidlink files would be the use of PVS modules that make use of the moving history of a family and of changes in the child's household.[25] The findings indicate that just using the 2007 Geobase, approximately 80 percent of all Kidlink children were linked to at least one parent. Having a parent already linked to a child record in conjunction with a child-parent link search methodology that takes into account the moving history of the family or changes in the child's household could increase the probability of finding a second parent.

Using tax records that include more than four dependents would also prove useful. A preliminary match of the children not found in either Geobase to the 2009 IRS 1040 electronic file revealed that for the children found in both the Kidlink file and the IRS file, approximately 25 percent of the household where these children reside have more than four dependents. In addition, including ITINs in the Geobase file(s) employed in the construction of the Kidlink file is important as many parents of U.S. citizen as well as noncitizen children would not otherwise be part of the Geobase.

Increasing the number of child-parent links will enable and enhance academic research that relies on intergenerational data linkage. However, an important step towards making available this type of data for research is the assessment of the quality of child-parent links. We turn to this analysis in the next section.

V.   Quality Evaluation of Kidlink Child-parent Links

Understanding the quality of the child-parent links in the Kidlink file is important for any future implementation of this type of file in record linkage. If the child-parent link information is used in a reference file, then the quality of the matches is dependent on the quality of the reference data.

A. Analytic Sample & Methodology

We evaluate child-parent(s) links in the Kidlink file by independently matching child-parent links in the 2007 Kidlink file to the child-parent links observed in two separate files that are used as benchmarks: the 2007 IRS and 2010 Census files.

---

[25] Remember that the methodology employed in the creation of the 2007 Kidlink file only used PVS' Geosearch module. See Layne & Wagner (2013) for more information.

In the cases where only one parent has been linked to the child record (i.e., the 'Only mother' or 'Only father' groups), we consider the Kidlink child-parent link to be "correct" if the child-parent link in the Kidlink file is the same as (or matches) the child-parent link observed in the IRS/Census. We use the term 'correct' with caution, though, since child-adult links observed in IRS or Census data do not necessarily correspond to child-parent relationships.[26]

When the Kidlink child record contains two parents (the 'Both parents' group), we create two alternative quality outcomes. The first is more restrictive and requires that both the child-mother and child-father links agree with the child-mother and child-father links observed in the IRS/Census. Only when this is the case, is the Kidlink child-parent link considered a 'correct' link for children in the 'Both parents' group. The second measure is less restrictive and requires that just one of the Kidlink child-parent links (either the mother's or the father's) matches the one observed in the IRS/Census.[27]

The quality of the Kidlink child-parent links is quantitatively measured by a simple match rate defined as the percentage of Kidlink child-parent links that successfully match to child-parent links in the 2007 IRS/2010 Census. Note that the denominator or analytic sample is the number of Kidlink children with a parent link that are found in the 2007 IRS/2010 Census. Table 4, column (1) shows the number and percentage of Kidlink children in the 2007 IRS and the 2010 Census. Out of the almost 83 million Kidlink children, 66,508,003 (or 80.2 percent) have at least one parent link. Out of those with a parent link, 65,129,570 are found in the 2007 IRS file, and 56,104,606 are found in the 2010 Census (column 2). These children represent 78.5 percent and 67.6 percent respectively of all Kidlink children,[28] and comprise the analytic samples used in the quality analysis.[29]

To explore the child characteristics associated with correct child-parent links, we examine descriptive statistics as well as logistic regressions. The dependent variable of the logit equals 1 if the Kidlink child-parent link matches the child-parent link in the 2007 IRS/2010 Census, and takes the value of 0 otherwise.

B. Findings on Kidlink Quality Evaluation

In this sub-section we first present results on the number and percentage of correct child-parent links in the 2007 Kidlink file, and then discuss child characteristics associated with them – including results from a series of logistic regressions involving a variety of model specifications.

---

[26] This will give rise to false negatives (i.e., true child-parent links in Kidlink may be deemed incorrect if they do not agree with those found in IRS/Census data).

[27] Our choice to create these alternate outcome measures is supported by results from an exploratory logit analysis. Results revealed that there are differences between Kidlink children in the Both parents' group where both parents have the correct link and Kidlink children in the 'Both parent' group' where only one of the parent's link is correct. Children who are older (except the oldest), those obtaining their SSN in the Midwest and noncitizens are more likely to have the link of both parents correctly identified.

[28] That is, Kidlink children with or without a parent link.

[29] Exploratory analysis shows that children in the analytic sample tend to have (relative to the rest of the Kidlink children) lower proportions of the youngest and oldest children, and higher proportions of U.S. citizens, children obtaining their SSNs when they are less than a year old and children receiving their SSNs in the Midwest or New England.

B.1. Quantifying the Quality of Child-Parent Links in Kidlink

Table 5 shows, overall and by parent category, the percentage of Kidlink child-parent links that match child-parent links in the 2007 IRS, and the 2010 Census independently (i.e., 'correct' Kidlink child-parent links).

The percentage of Kidlink child-parent links that match those found in the 2007 IRS varies from nearly 90 percent (using the restrictive measure for 'Both-parents' children) to almost 95 percent (using the non-restrictive measure for 'Both-parents' children) (see columns 9 and 10 respectively).[30] For the 2010 Census, those figures are 81.4 percent (restrictive) to 88.4 percent (non-restrictive) (see columns 9 and 10).[31] The lower Kidlink-Census rate can be attributed to several factors. First, as pointed out in the Data section, only 90.8 percent of person records in the 2010 Census received a PIK, and are thus, "matchable". Second, PIKs are assigned to persons in the Census file via a probabilistic matching algorithm (i.e., PVS) that may introduce some PIK assignment error. Third, there is a time misalignment between the vintage of the Kidlink file (2007) and the 2010 Census – while the IRS file has the same vintage as the Kidlink file.[32] That is, Kidlink children found in the 2010 Census are 3 to 22 years old in 2010. As time passes and children grow older, there is an increasing chance that they may not live with their parents due to changes in household composition. For instance, three years later, parents could have divorced or many children that are 18 to 23 years old in 2010 may have moved out of their parents' home. The result is that child-parent links that are actually accurate in the 2007 Kidlink file may be classified as incorrect when benchmarked against the child-parent links in the 2010 Census – producing a "false negative" effect. To partly control for this effect, we conduct robustness checks that included restricting the analytic sample to only children that are biological or adopted sons/daughters of the householder according to the 2010 Census. When using this sample of children, the non-restricted match rate for the 2010 Census increases to 96.3 percent.[33]

The percentage of Kidlink child-parent links that match those found in the 2007 IRS is high across parent categories. By contrast, the rate is not as consistent among parent categories in the Kidlink-Census comparison. The percent of children with correct child-parent links in the 'Only Father' category is 68.4 (column 7), which is over 10 percentage points lower than the next group ('Both parents'). This is in sharp contrast to the percent of Kidlink children in the 'Only Father' category with correct child-parent links (93 percent) when the benchmark is the 2007 IRS (column 7). These results could be attributed to the nature of the IRS versus Census data as well as the time frame misalignment between the vintage of the Kidlink file and the 2010 Census. As already pointed out, there is an increasing chance of changes in the household composition as time passes. Changes in household composition may often reflect

---

[30] That is, the numerator of the 90 percent figure is the sum of single-parent matches and the restrictive 2-parent matches (columns 5, 7 and 8). Meanwhile, single-parent and non-restrictive 2-parent matches make up the numerator of the 95 percent figure (columns 6, 7 and 8).

[31] As a percentage of all Kidlink kids, the child-parent link match rate is 70.6 percent for the 2007 IRS and 55.1 percent for the 2010 Census for the restrictive outcome measure, and 74.5 and 59.8 percent respectively for the non-restrictive measure.

[32] The 2007 IRS file is part of the reference file used in the construction of Kidlink.

[33] There are 45,527,342 Kidlink children with a parent link found in the 2010 Census who are biological or adopted children of the householder.

changing household structures and new household formation among older children. Thus, for the Census, the Kidlink child may become the householder him/herself (in the case of older children) or be living with a householder that is not his/her father – however, the child and the householder will be "connected" because both reside at the same address in the Census.

Because of the reasons discussed above, the 2007 IRS file represents a better benchmark for evaluating the quality of the child-parent links in the Kidlink file. With this in mind, we place a higher emphasis on results from the Kidlink-IRS analysis. In the remainder of this section, we then focus on the Kidlink-IRS match, and discuss Kidlink-Census results relative to those obtained using IRS data. Tables from the Kidlink-Census analysis are included in the Appendix.

B.2. Characteristics of Correct Kidlink Child-Parent Links

Table 6 presents the distributions of demographic characteristics of Kidlink children whose child-parent links match those found in the 2007 IRS (i.e., 'correct' child-parent links). Characteristics available in the data include the age, gender, and citizenship status of the child as well as the age and state where the child received his/her SSN.[34] Because the distributions when using the non-restrictive outcome measure are qualitatively the same as those when using the restrictive measure, we present results for the non-restrictive measure here, and show results from the restrictive measure in Table A1 in the Appendix.[35]

Columns 4-6 in Table 6 show that, relative to correct Kidlink child-parent links, incorrect links tend to have higher proportions of Kidlink children in the oldest and the youngest age categories, as well as children that received their SSN when they were at least 1 year old, and in the South. Differences in terms of child citizenship status and gender are very small - below 1 percentage points.

When the 2010 Census is used as a benchmark, incorrect Kidlink child-parent links also have higher proportions of older children and children that received their SSN when they were at least 1 year old (Appendix, Tables A2 and A3).[36] The 2010 Census has additional demographic characteristics such as race, Hispanic origin and the relation of the child with respect to the householder. Examination of these characteristics reveals that the Hispanic origin distribution is very similar for correct and incorrect Kidlink child-parent links (within 1 percentage point). Incorrect child-parent links have somewhat lower proportions of whites and Asians,[37] and much higher proportions of Kidlink children who are (according to the 2010 Census) either householders themselves, grandchildren of the householder or not related to the householder. This last finding is expected given the time frame misalignment between the 2007 Kidlink file and the 2010 Census, and resulting false-negatives effect discussed earlier.

---

[34] These characteristics are obtained from the Numident file. Race is not available because the Social Security Administration halted collection of race data when it transitioned to the Enumeration at Birth system in 1987. Prior to that, race categories were inconsistent over time.

[35] The only difference worth noting is that mismatches with the restrictive measure tend to have a higher proportion of the oldest children (17.8 percent versus 12.1 percent).

[36] Just as with the IRS, the distributions when using the non-restrictive outcome measure are qualitatively the same as those when using the restrictive measure.

[37] When the restrictive outcome measure is used, differences in the race distribution between correct matches and mismatches are reduced (see Table A3 in Appendix).

When only Kidlink children that are biological or adopted sons/daughters of the householder are included in the sample, the age distribution differences between correct and incorrect Kidlink child-parent links are diminished (see Table A4 in Appendix).[38] Likewise, differences in the distribution of the age when the SSN is received disappear. On the other hand, race distribution differences between correct and incorrect child-parent links are exacerbated, and Hispanic origin differences emerge – with incorrect child-parent links having a higher proportion of Hispanics children.

B.3. Logit Results

As mentioned earlier, we run a series of logistic regressions where the dependent variable equals 1 if the Kidlink child-parent link successfully matches the child-parent link in the 2007 IRS, and takes the value of 0 otherwise. The base model specification includes the following covariates: parent categories (i.e., 'Both parents', 'Only father', 'Only mother'), child age categories, child gender, child citizenship status, the region where the child received his/her SSN, and a dummy variable indicating whether the length of the parent's name is short.[39] As part of the sensitivity analysis, we run a second model specification that includes how old the child was when receiving his/her SSN. This variable is not included in the base specification because it is likely to be correlated with citizenship status. These two model specifications are also run separately by parent categories to explore any potential interactions between the parent groups and other variables of interest.[40] Tables 7 and 8 report odds ratios[41] - table 7 presents results based on the non-restrictive outcome measure and Table 8 on the restrictive one.[42]

Results from the non-restrictive outcome base model (Table 7 column 1) indicate that, compared to the 'Both parents' group (the reference category), Kidlink children in the 'Mother only' and 'Father only' groups are less likely to have child-parent links that match those found in the 2007 IRS. This is not the case when we employ the restrictive outcome measure as our dependent variable (Table 8 column 1). This is the case simply because in the non-restrictive outcome measure, only one of the parents (not the two parents) has to be linked to the child in order for our dependent variable to equal one.

Continuing with the full sample non-restrictive Kidlink-IRS logit (column 1 Table 7), results show that compared to the youngest children (the reference category), older children are more likely to have correct child-parent links – with the exception of the oldest category. This result holds in the restrictive-outcome logit results as well (Table 8 column 1). Note, however, that the lower likelihood for the oldest group is driven by the 'Both parents' and 'Only father' children in both the restrictive and non-restrictive outcome logits (Tables 7 and 8, columns 3 and 7).

---

[38] This could be due to the fact that the remaining households have undergone less changes in their composition.
[39] This variable equals 1 if the length of the parent name is below the 95th percentile in the length distribution of all parent names. It is included since the methodology for constructing the Kidlink file involved probabilistically matching parent names in the Numident to a reference file.
[40] The sensitivity analysis also involved additional robustness checks including testing a variety of model specifications, and citizenship status-race interactions (results not shown).
[41] In this analysis, the odds ratios represent how much more likely a given group is a correct match relative to the base group.
[42] As indicated earlier, results from an exploratory logit indicated the need to run separate logits for the restrictive and non-restrictive outcome measures.

The Kidlink-2010 Census logit results (both for the non-restrictive and restrictive outcome models) show the same age pattern, but here, the oldest children are a lot less likely to have correct child-parent links and this lower likelihood also applies to the 'Mother only' group (see Tables A5 & A6). Again, this may be attributed to some extent to the time frame misalignment between the 2007 Kidlink file and the 2010 Census. Older children moving out of their parents' home may still be claimed as dependents in IRS tax forms, but the child-parent link in the 2010 Census is more likely to "get lost" three years later. This consideration is consistent with the logit results obtained when only biological and adopted sons/daughters of the householder are included in the sample (see Table A7 in Appendix). Results from this subsample of children indicate that the oldest children are now more likely than the youngest to have a correct child-parent link – with the exception of the 'Father only' children.

Gender does not seem to play a role in the probability of Kidlink children having a correct child-parent link – although Kidlink-Census results indicate that female children have a lower probability than male children to have correct child-parent links.

The citizenship status variable has a somewhat unexpected effect.[43] Kidlink-IRS results from the restrictive-outcome logit (Table 8) indicate that non-U.S. citizen children in Kidlink are generally more likely to have correct child-parent links relative to U.S. citizens. With the exception of the noncitizen–authorized category in the 'Only Father' group, this finding is robust when running the model by parent group (Table 8, columns 3- 8), as well as in the Kidlink-2010 Census sample (Tables A5 and A6 in Appendix). It is also robust to a number of specifications including race-citizenship status interactions in the Kidlink-Census sample (results not shown).[44] Results from the Kidlink-IRS non-restrictive outcome logit (Table 7) are slightly different indicating that the noncitizen-authorized group is less likely than the U.S. citizen group to have correct child-parent links. This result, however, is mostly driven by the 'Both parents' group (column 3 Table 7).

These citizenship status findings may be partly attributable to sample bias. Remember from section V.A. that the sample of Kidlink children used in the evaluation of child-parent links is composed (by definition) of Kidlink children records with at least one parent link and that can be found in the 2007 IRS/2010 Census. That is, Kidlink children records in the 'Neither parent' category are, by definition, excluded from the sample. Our coverage evaluation (section IV) showed that Kidlink children who are not U.S. citizens, especially the noncitizen-other group, have a much higher proportion of children with no parent link than the citizen group (66.2 percent and 26.0 percent for the noncitizen-other and noncitizen-authorized groups compared to 19.5 percent for citizens). Additional supporting evidence is provided by results from an exploratory logit model designed to assess a Kidlink child's likelihood of being in the 2007 IRS/2010 Census. Results indicate that, overall, non-U.S. citizen children in Kidlink are less likely to be found in the IRS/Census relative to U.S. citizens. However, when Kidlink children records without a parent link are excluded from the sample, the remaining non-U.S. citizens are now more likely

---

[43] Results hold even when the noncitizen category is defined as a child *currently* having noncitizen status (versus ever having noncitizen status).

[44] The 2010 Census file contains additional variable (namely, child race, Hispanic origin and relationship to the householder) that serve as additional controls in the Kidlink-Census logit. Even controlling for additional characteristics and including race-citizenship status interactions, the citizenship status results remain robust.

or as equally likely to be found in the IRS/ Census file - compared to U.S. citizens (results not shown). This evidence suggests that the Kidlink children in the noncitizen groups remaining in our analytic sample are not representative.

As part of the sensitivity analysis, we also include the age at which the child receives his/her SSN in a separate model specification. Generally, children receiving their SSNs when they are older and not right after birth are less likely to have correct child-parent links (Tables 7 and 8, columns 2, 4, 6, 8).[45] As already mentioned, this variable is likely to be correlated with citizenship status, as noncitizens are more likely to receive their SSN after birth. Note that, as predicted, the noncitizen categories are sensitive to the addition of this variable in the model (columns 2, 4, 6, 8 in Tables 7 and 8). Specifically, its inclusion increases the odds ratios of the noncitizen categories.

C. Quality Analysis Conclusions

Results indicate that, using 2007 IRS data as a benchmark, the percentage of correct Kidlink child-parent links varied from 90 (restrictive outcome measure) to 95 percent (non-restrictive outcome measure). When the 2010 Census is the benchmark, this percentage varied from 81.4 percent to 88.4 percent.

Logit results indicate that the oldest children, children receiving their SSNs while living in the South and West-Mountain region[46] (compared to the Midwest-East region), and those receiving their SSNs after 1 year of age are more likely to have incorrect child-parent link in Kidlink. Employing Geobases from years prior to the Numident file used in the construction of Kidlink may help improve the quality of child-parent links for the oldest children as they will be more likely to live with at least one of their parents earlier in life.

The child's citizenship status categories offer some unexpected results. Compared to U.S. citizens, the noncitizen-other category is more likely to have correct child-parent links. The noncitizen-authorized category exhibits mixed results depending on whether the benchmark was the 2007 IRS file or the 2010 Census, and also on whether the model specification included the age at which the child received his/her SSN. We provide evidence, though, that suggest that these findings on citizenship status can be attributed to sample bias.

VI. Next Steps

The creation of an updated Kidlink file following the recommendations on methodology that we outlined earlier in the paper is an important first step. The updates would include adding Geobases from different vintages, including ITINs in the Geobase reference files whenever possible, using the decennial census as part of the reference files to obtain parent links, adding to the Geobase sources (such as SNAP AR data) that contain child-parent units at addresses, and employing PVS modules that

---

[45] In the Kidlink-IRS restrictive outcome logit, children receiving their SSNs between 1 and 2 years of age tends to be slightly more likely to have a correct child-parent link compared to children receiving their SSN at less than 1 year of age. This result is primarily driven by the children in the 'Both parents' group.

[46] The West-Mountain region includes Arizona, Colorado, Idaho, New Mexico, Montana, Utah, Nevada and Wyoming. The Midwest-East region includes Indiana, Illinois, Michigan, Ohio and Wisconsin. See Table A8 in Appendix for a list of all regions and corresponding states.

take into account the moving history of a family as well as changes in the child's household composition. These changes will likely increase the number of child-parent links that can be obtained from a given Numident. After the update is complete, the coverage and quality of the child-parent links will be evaluated.

The creation and improvement of a Census Kidlink file has important and diverse implications. By integrating child-parent links into PVS' reference files, Census Kidlink will help enhance PVS' capability to assign PIKs, and hence, person record "linkability" across and within files. This, in turn, will help the Census Bureau's research aimed to decrease respondent burden and reduce costs while improving data quality.

In addition, the type of intergenerational linkage provided by Kidlink has the potential of advancing social science research on topics that rely on such type of linkage. Linking children to parents enables studies on income or social mobility, poverty persistence, labor market outcomes and child development, among others. Findings from this type of research will likely reveal new insights with important policy implications.

References

Farber, James and Charlene Leggieri. 2002. "Building and Validating a National Administrative Records Database for the United States," New Zealand Conference on Database Integration.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez. 2014. "Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States." National Bureau of Economic Research Working Paper 19843.

Luque, Adela and Renuka Bhaskar. 2014. "2010 American Community Survey Match Study." CARRA WP No. 2014-03.

Massey, Catherine G., Amy O'Hara, and David S. Johnson. 2014. "The Opportunities and Challenges of Using Administrative Data Linkages to Evaluate Mobility." *The ANNALS of the American Academy of Political and Social Science*, 657 (1): 247-264.

NORC. 2012. "2010 Census Match Study Related Research." Center for Administrative Records Research & Applications (CARRA) Internal Report.

NORC. 2013. "Further PVS Research Final Research Report." Center for Administrative Records Research & Applications (CARRA) Internal Report.

Rastogi, Sonya and Amy O'Hara. 2012. "Census Match Study," 2010 Census Program for Evaluations and Experiments, 2010 Planning Memoranda Series No. 247, U.S. Census Bureau.

Wagner, Deborah and Mary Layne. 2013. "The Person Identification Validation System (PVS): Applying the Center for Administrative Records Research & Applications' (CARRA) Record Linkage Software," CARRA Research Report Series, U.S. Census Bureau. CARRA-05-13.

Tables

Table 1: Parent-Child Links in 2007 Kidlink, by Child Age & Citizenship Status

	Number					Row Percent			
	Total	Both Parents	Only Father	Only Mother	Neither Parent		Both Parents	Only Father	Only Mother	Neither Parent
Total	82,955,825	35,502,993	6,550,978	24,454,032	16,447,822		42.8%	7.9%	29.5%	19.8%
Child Age										
0	1,691,514	625,862	90,396	251,522	723,734		37.0%	5.3%	14.9%	42.8%
1	4,271,424	2,122,589	296,576	864,683	987,576		49.7%	6.9%	20.3%	23.1%
2	4,243,547	2,154,755	286,409	901,544	900,839		50.8%	6.7%	21.3%	21.2%
3	4,239,675	2,150,103	285,481	941,242	862,849		50.7%	6.7%	22.2%	20.4%
4	4,177,606	2,092,082	289,023	962,499	834,002		50.1%	6.9%	23.0%	20.0%
5	4,134,904	2,032,045	295,909	987,438	819,512		49.1%	7.2%	23.9%	19.8%
6	4,208,427	2,030,231	309,617	1,046,810	821,769		48.2%	7.4%	24.9%	19.5%
7	4,209,448	1,930,150	308,980	1,162,206	808,112		45.9%	7.3%	27.6%	19.2%
8	4,159,758	1,745,094	299,626	1,321,058	793,980		42.0%	7.2%	31.8%	19.1%
9	4,154,360	1,670,550	303,677	1,392,042	788,091		40.2%	7.3%	33.5%	19.0%
10	4,134,246	1,643,109	316,636	1,392,097	782,404		39.7%	7.7%	33.7%	18.9%
11	4,138,745	1,633,140	330,623	1,394,565	780,417		39.5%	8.0%	33.7%	18.9%
12	4,195,836	1,630,811	348,610	1,427,236	789,179		38.9%	8.3%	34.0%	18.8%
13	4,283,851	1,637,900	366,394	1,468,969	810,588		38.2%	8.6%	34.3%	18.9%
14	4,377,144	1,647,425	388,429	1,504,810	836,480		37.6%	8.9%	34.4%	19.1%
15	4,438,704	1,658,812	403,912	1,528,382	847,598		37.4%	9.1%	34.4%	19.1%
16	4,525,207	1,688,073	412,853	1,571,663	852,618		37.3%	9.2%	34.7%	18.8%
17	4,554,949	1,738,857	411,927	1,576,333	827,832		38.2%	9.0%	34.6%	18.2%
18	4,465,544	1,811,478	406,952	1,463,996	783,118		40.6%	9.1%	32.8%	17.5%
19	4,350,936	1,859,927	398,948	1,294,937	797,124		42.7%	9.2%	29.8%	18.3%
Citizenship Status										
Citizen/ Unknown	80,490,975	34,713,094	6,033,797	24,061,607	15,682,477		43.1%	7.5%	29.9%	19.5%
Noncitizen-authorized	2,157,158	774,574	437,753	383,102	561,729		35.9%	20.3%	17.8%	26.0%
Noncitizen-other	307,692	15,325	79,428	9,323	203,616		5.0%	25.8%	3.0%	66.2%

Sources: 2007 Numident & 2007 Geobase.

Table 2: 2007 Kidlink Children with No Parent Link

	Number				Row Percent		
	Total	Child in 2007 Geo	Child in 2013 Geo, but not in 2007 Geo	Child not in either Geobase	Child in 2007 Geo	Child in 2013 Geo,but not in 2007 Geo	Child not in either Geobase
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Total	16,447,822	8,542,312	5,374,378	2,531,132	51.9%	32.7%	15.4%
Child Age							
0	723,734	132,672	542,381	48,681	18.3%	74.9%	6.7%
1	987,576	447,049	430,998	109,529	45.3%	43.6%	11.1%
2	900,839	450,751	331,875	118,213	50.0%	36.8%	13.2%
3	862,849	445,394	294,947	122,508	51.6%	34.2%	14.2%
4	834,002	436,529	273,463	124,010	52.3%	32.8%	14.9%
5	819,512	429,683	261,903	127,926	52.4%	32.0%	15.6%
6	821,769	430,785	258,580	132,404	52.4%	31.5%	16.1%
7	808,112	426,244	250,652	131,216	52.7%	31.0%	16.3%
8	793,980	425,644	239,039	129,297	53.6%	30.1%	16.3%
9	788,091	423,702	234,401	129,988	53.8%	29.7%	16.5%
10	782,404	418,622	230,737	133,045	53.5%	29.5%	17.0%
11	780,417	412,961	232,053	135,403	52.9%	29.7%	17.4%
12	789,179	415,339	241,972	131,868	52.6%	30.7%	16.7%
13	810,588	426,080	258,321	126,187	52.6%	31.9%	15.6%
14	836,480	439,134	270,656	126,690	52.5%	32.4%	15.1%
15	847,598	443,123	272,084	132,391	52.3%	32.1%	15.6%
16	852,618	451,836	258,322	142,460	53.0%	30.3%	16.7%
17	827,832	455,764	221,516	150,552	55.1%	26.8%	18.2%
18	783,118	474,033	162,620	146,465	60.5%	20.8%	18.7%
19	797,124	556,967	107,858	132,299	69.9%	13.5%	16.6%
Citizenship Status							
U.S. Citizen / Unknown	15,682,477	8,223,426	5,193,501	2,265,550	52.4%	33.1%	14.4%
Noncitizen-authorized	561,729	255,156	147,696	158,877	45.4%	26.3%	28.3%
Noncitizen-other	203,616	63,730	33,181	106,705	31.3%	16.3%	52.4%

Sources: 2007 Numident & 2007, 2013 Geobases.

Table 3: Age and Citizenship Status Distributions of 2007 Kidlink Children by Parent Category

Kidlink Child Citizenship Status	Parent Categories					Percentage Point Difference from 'Both Parents'			
	Both parents (1)	Only father (2)	Only mother (3)	Neither - in Geobase (4)	Neither – not in Geobase (5)	Only father (6)	Only mother (7)	Neither - in Geobase (8)	Neither – not in Geobase (9)
No. of Children	35,502,993	6,550,978	24,454,032	13,916,690	2,531,132				
Child Age									
0	1.8%	1.4%	1.0%	4.9%	1.9%	-0.4	-0.8	3.1	0.1
1	6.0%	4.5%	3.5%	6.3%	4.3%	-1.5	-2.5	0.3	-1.7
2	6.1%	4.4%	3.7%	5.6%	4.7%	-1.7	-2.4	-0.5	-1.4
3	6.1%	4.4%	3.8%	5.3%	4.8%	-1.7	-2.3	-0.8	-1.3
4	5.9%	4.4%	3.9%	5.1%	4.9%	-1.5	-2.0	-0.8	-1.0
5	5.7%	4.5%	4.0%	5.0%	5.1%	-1.2	-1.7	-0.7	-0.6
6	5.7%	4.7%	4.3%	5.0%	5.2%	-1.0	-1.4	-0.7	-0.5
7	5.4%	4.7%	4.8%	4.9%	5.2%	-0.7	-0.6	-0.5	-0.2
8	4.9%	4.6%	5.4%	4.8%	5.1%	-0.3	0.5	-0.1	0.2
9	4.7%	4.6%	5.7%	4.7%	5.1%	-0.1	1.0	0.0	0.4
10	4.6%	4.8%	5.7%	4.7%	5.3%	0.2	1.1	0.1	0.7
11	4.6%	5.0%	5.7%	4.6%	5.3%	0.4	1.1	0.0	0.7
12	4.6%	5.3%	5.8%	4.7%	5.2%	0.7	1.2	0.1	0.6
13	4.6%	5.6%	6.0%	4.9%	5.0%	1.0	1.4	0.3	0.4
14	4.6%	5.9%	6.2%	5.1%	5.0%	1.3	1.6	0.5	0.4
15	4.7%	6.2%	6.3%	5.1%	5.2%	1.5	1.6	0.4	0.5
16	4.8%	6.3%	6.4%	5.1%	5.6%	1.5	1.6	0.3	0.8
17	4.9%	6.3%	6.4%	4.9%	5.9%	1.4	1.5	0.0	1.0
18	5.1%	6.2%	6.0%	4.6%	5.8%	1.1	0.9	-0.5	0.7
19	5.2%	6.1%	5.3%	4.8%	5.2%	0.9	0.1	-0.4	0.0
Total[1]	100.0%	100.0%	100.0%	100.0%	100.0%				
Child Citizenship Status									
US Citizen/ Unknown	97.8%	92.1%	98.4%	96.4%	89.5%	-5.7	0.6	-1.4	-8.3
Noncitizen – authorized	2.2%	6.7%	1.6%	2.9%	6.3%	4.5	-0.6	0.7	4.1
Noncitizen – other	0.0%	1.2%	0.0%	0.7%	4.2%	1.2	0.0	0.7	4.2
Total	100.0%	100.0%	100.0%	100.0%	100.0%				

Sources: 2007 Numident & 2007, 2013 Geobases.
[1] Some columns do not sum up to 100% due to rounding.

Table 4: Kidlink Children in the 2007 IRS and 2010 Census

		All Kidlink Children (1)	Kidlink Children with parent link (2)
Total		82,955,825	66,508,003
In IRS	Number	73,120,316	65,129,570
	As a % of Total	88.1%	97.9%
In Census	Number	66,664,033	56,104,606
	As a % of Total	80.4%	84.4%

Universe: Kidlink children.
Sources: 2007 Kidlink file, 2007 IRS, 2010 Census.

Table 5: Kidlink Quality Assessment: Correct Kidlink Child-Parent Links using 2007 IRS & 2010 Census as Benchmarks

Benchmark	Kidlink children with parent link in 2007 IRS/2010 Census				Correct Kidlink Child-Parent Links					
	Both Parents	Only Father	Only Mother	All	Both Parents		Only Father	Only Mother	Overall[a]	
					Father & Mother (Restrictive)	Father or Mother (Non-restrictive)			Overall – Restrictive	Overall – Non-restrictive
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
In 2007 IRS	35,131,309	6,323,074	23,675,187	65,129,570	31,430,340 89.5%	34,673,208 98.7%	5,882,072 93.0%	21,265,073 89.8%	58,577,485 89.9%	61,820,353 94.9%
In 2010 Census	30,853,222	5,280,043	19,971,341	56,104,606	25,109,495 81.4%	29,037,062 94.1%	3,613,594 68.4%	16,964,105 84.9%	45,687,194 81.4%	49,614,761 88.4%

Universe: Kidlink children with parent link.
[a] The number of overall-restrictive is the sum of columns 7, 9 and 10. The number of overall-nonrestrictive is the sum of columns 8, 9 and 10.
Sources: 2007 Kidlink file, 2007 IRS, 2010 Census.

Table 6: Characteristics of Correct Kidlink Child-Parent Links, Non-restrictive Outcome Measure - Kidlink & 2007 IRS Comparison

	Number			Percent		
	Total	Incorrect Link	Correct Link	Incorrect Link	Correct Link	Pctg. Point Difference
	(1)	(2)	(3)	(4)	(5)	(6)
Total	65,129,570	3,309,217	61,820,353	100.0%	100.0%	100.0%
Child Age (as of 2007)						
0	959,870	60,825	899,045	1.8%	1.5%	-0.4
1	3,261,762	197,760	3,064,002	6.0%	5.0%	-1.0
2	3,313,243	169,952	3,143,291	5.1%	5.1%	-0.1
3	3,343,577	147,233	3,196,344	4.4%	5.2%	0.7
4	3,307,579	133,480	3,174,099	4.0%	5.1%	1.1
5	3,276,119	126,667	3,149,452	3.8%	5.1%	1.3
6	3,342,526	127,883	3,214,643	3.9%	5.2%	1.3
7	3,352,854	129,192	3,223,662	3.9%	5.2%	1.3
8	3,313,998	128,721	3,185,277	3.9%	5.2%	1.3
9	3,310,391	128,412	3,181,979	3.9%	5.1%	1.3
10	3,291,670	123,957	3,167,713	3.7%	5.1%	1.4
11	3,294,899	119,662	3,175,237	3.6%	5.1%	1.5
12	3,338,679	119,760	3,218,919	3.6%	5.2%	1.6
13	3,399,488	123,461	3,276,027	3.7%	5.3%	1.6
14	3,459,653	127,384	3,332,269	3.8%	5.4%	1.5
15	3,500,772	132,136	3,368,636	4.0%	5.4%	1.5
16	3,568,680	145,497	3,423,183	4.4%	5.5%	1.1
17	3,603,347	186,657	3,416,690	5.6%	5.5%	-0.1
18	3,527,270	292,122	3,235,148	8.8%	5.2%	-3.6
19	3,363,193	588,456	2,774,737	17.8%	4.5%	-13.3
Gender						
Female	31,863,952	1,605,052	30,258,900	48.5%	48.9%	0.4
Male	33,265,230	1,704,115	31,561,115	51.5%	51.1%	-0.4
Missing	388	50	338	0.0%	0.0%	0.0
Citizenship Status						
Citizen/Unknown	63,457,645	3,205,613	60,252,032	96.9%	97.5%	0.6
Noncitizen – authorized	1,568,908	99,903	1,469,005	3.0%	2.4%	-0.6
Noncitizen –other	103,017	3,701	99,316	0.1%	0.2%	0.0
Age SSN Received						
<1	58,794,151	2,854,860	55,939,291	86.3%	90.5%	4.2
1-2	2,551,080	145,361	2,405,719	4.4%	3.9%	-0.5
2-3	1,635,808	131,372	1,504,436	4.0%	2.4%	-1.5
3-4	678,637	60,830	617,807	1.8%	1.0%	-0.8
4-5	321,484	27,725	293,759	0.8%	0.5%	-0.4
5+	1,148,410	89,069	1,059,341	2.7%	1.7%	-1.0
Region SSN received						
NE -New England	3,050,815	103,613	2,947,202	3.1%	4.7%	1.6
NE -Mid Atlantic	8,717,146	443,576	8,273,570	13.4%	13.4%	0.0
Midwest -East	10,455,007	460,352	9,994,655	13.9%	16.2%	2.3
Midwest -West	4,501,254	172,216	4,329,038	5.2%	7.0%	1.8
S -Atlantic	11,388,861	663,528	10,725,333	20.1%	17.3%	-2.7
S -East Central	3,751,330	260,793	3,490,537	7.9%	5.6%	-2.2
S -West Central	7,984,310	502,999	7,481,311	15.2%	12.1%	-3.1
W -Mountain	3,871,292	179,341	3,691,951	5.4%	6.0%	0.6
W -Pacific	10,956,330	475,982	10,480,348	14.4%	17.0%	2.6
Unknown	453,225	46,817	406,408	1.4%	0.7%	-0.8

Sources: 2007 Kidlink file, 2007 IRS, 2010 Census.

Table 7: Logit Results – Kidlink & 2007 IRS Comparison – Non-restrictive Outcome Measure

	All		Both Parents		Only Mother		Only Father	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Parent link								
Both								
Father	0.2***	0.2***						
Mother	0.1***	0.1***						
Child Age (as of 2007)								
<1								
1-3	1.3***	1.3***	1.3***	1.3***	1.3***	1.3***	1.2***	1.2***
4-7	1.9***	1.9***	1.9***	1.9***	2.1***	2.1***	1.4***	1.4***
8-15	2.4***	2.4***	1.7***	1.8***	2.9***	2.9***	1.5***	1.5***
16+	0.9***	1.0**	0.2***	0.2***	1.6***	1.6***	0.5***	0.6***
Gender								
Male								
Female	1.0***	1.0***	1.1***	1.1***	1.0***	1.0***	1.0***	1.0***
Citizenship Status								
Citizen/Unknown								
Noncitizen – authorized	0.8***	1.2***	0.5***	0.9***	1.1***	1.3***	0.9***	1.2***
Noncitizen –other	2.3***	2.9***	1.0***	1.3***	1.2***	1.3***	3.4***	4.2***
Region SSN received								
Midwest -East								
Midwest -West	1.1***	1.1***	1.0***	0.9***	1.1**	1.1***	1.0**	1.0***
NE -Mid Atlantic	1.3***	1.3***	0.8***	0.8***	1.4***	1.4***	0.8***	0.8***
NE -New England	1.2***	1.3***	1.0***	1.1***	1.3	1.3***	1.0	1.0
S -Atlantic	0.9***	0.9***	0.8***	0.8***	0.9***	0.9***	0.7***	0.7***
S -East Central	0.7***	0.7***	0.7***	0.7***	0.7***	0.7***	0.7***	0.7***
S -West Central	0.8***	0.8***	0.6***	0.6***	0.9***	0.9***	0.7***	0.7***
W -Mountain	0.9***	0.9***	0.6***	0.6***	1.0***	1.0***	0.8***	0.8***
W -Pacific	1.1***	1.1***	0.7***	0.7***	1.2***	1.2***	0.9***	0.9***
Unknown	0.5***	0.5***	0.2***	0.2***	0.6***	0.6***	0.5***	0.5***
Age SSN received								
<1								
1-2		0.9***		0.8***		1.0***		0.9***
2-3		0.7***		0.6***		0.9***		0.7***
3-4		0.6***		0.5***		0.8***		0.7***
4-5		0.6***		0.5***		0.8***		0.6***
5+		0.6***		0.5***		0.8***		0.7***
Short parent name	1.1***	1.1***	1.1***	1.1***	1.2***	1.2***	0.9***	0.9***

Notes: *, ** and *** represent significance at the 10, 5 and 1 percent level respectively.
Universe: Kidlink children with parent link found in 2007 IRS file.
Sources: 2007 Kidlink file, 2007 IRS, 2010 Census.

Table 8: Logit Results – Kidlink & 2007 IRS Comparison – Restrictive Outcome Measure

	All		Both Parents		Only Mother		Only Father	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Parent link								
Both								
Father	1.6***	1.6***						
Mother	1.0***	1.0***						
Child Age (as of 2007)								
<1								
1-3	1.2***	1.2***	1.2***	1.2***	1.3***	1.3***	1.2***	1.2***
4-7	1.7***	1.7***	1.6***	1.6***	2.1***	2.1***	1.4***	1.4***
8-15	2.0***	2.0***	1.6***	1.6***	2.9***	2.9***	1.5***	1.5***
16+	1.0***	1.0**	0.8***	0.8***	1.6***	1.6***	0.5***	0.6***
Gender								
Male								
Female	1.0***	1.0***	1.0***	1.0***	1.0***	1.0***	1.0***	1.0***
Citizenship Status								
Citizen/Unknown								
Noncitizen – authorized	1.0***	1.2***	1.2***	1.3***	1.1***	1.3***	0.9***	1.2***
Noncitizen –other	2.0***	2.2***	1.3***	1.3***	1.2***	1.3***	3.4***	4.2***
Region SSN received								
Midwest -East								
Midwest -West	1.1***	1.1***	1.1***	1.1***	1.1***	1.1***	1.0**	1.0***
NE -Mid Atlantic	1.1***	1.1***	0.8***	0.8***	1.4***	1.4***	0.8***	0.8***
NE -New England	1.1***	1.1***	1.0***	1.0***	1.3***	1.3***	1.0	1.0
S -Atlantic	0.8***	0.8***	0.7***	0.7***	0.9***	0.9***	0.7***	0.7***
S -East Central	0.7***	0.7***	0.7***	0.7***	0.7***	0.7***	0.7***	0.7***
S -West Central	0.7***	0.7***	0.7***	0.7***	0.9***	0.9***	0.7***	0.7***
W -Mountain	1.0***	1.0***	0.9***	0.9***	1.0***	1.0***	0.8***	0.8***
W -Pacific	0.9***	0.9***	0.8***	0.8***	1.2***	1.2***	0.9***	0.9***
Unknown	0.5***	0.5***	0.4***	0.4***	0.6***	0.6***	0.5***	0.5***
Age SSN received								
<1								
1-2		1.1***		1.2***		1.0***		0.9***
2-3		0.9***		1.0***		0.9***		0.7***
3-4		0.8***		0.9***		0.8***		0.7***
4-5		0.8***		0.8***		0.8***		0.6***
5+		0.8***		0.9***		0.8***		0.7***
Short parent name	1.1***	1.1***	1.1***	1.1***	1.2***	1.2***	0.9***	0.9***

Notes: *, ** and *** represent significance at the 10, 5 and 1 percent level respectively.
Universe: Kidlink children w/ parent link found in 2007 IRS file.
Sources: 2007 Kidlink file, 2007 IRS, 2010 Census.

Appendix

Step-by-Step Description of Kidlink Methodology

Specifically, to build the file, the following steps were performed:

1. Create input file of eligible Numident records for parent's SSN search:
    a. Extract all 2007 Numident SSN records where date of birth is not missing and there is no date of death information
        i. Of the 441,829,508 person records in the 2007 Numident, 385,079,988 were extracted and considered eligible and in scope for the Kidlink construction
    b. Extract all addresses for these SSNs from the PVS Geobase (vintage 2007 with one year of address information)
    c. Standardize[47] the parents names

2. Search for mother's SSN for each of the records in the file constructed in Step 1 (using the 2007 vintage PVS Geobase), matching on name, gender and address.

    a. Create final set of mother's SSN links from the preliminary matches as follows:
        i. Mother must be at least 14 years older than child, but not greater than 50 years older
        ii. Drop all links in cases where more than one mother SSN were found for the child[48]

3. Search for father's SSN for each of the records in the file constructed in Step 1 (using the 2007 vintage PVS Geobase), matching on name, gender and address.

    a. Create final set of father's SSN links from the preliminary matches as follows:
        i. Father must be at least 14 years older than child, but not greater than 75 years older
        ii. Drop all links in cases where more than one father's SSN found for child[49]

4. Create final Census Kidlink file by merging file from Step1 to final mother's SSN file and final father's SSN file by child's SSN.

5. The analytic Kidlink extract file includes Kidlink children that are 19 or younger as of 2007. This includes a total of 82,955,825 child records out of the 385,079,988 person records considered eligible for the Kidlink file.

---

[47] The Name Standardization process parses name data into separate name component fields and standardizes key components. The standardization process provides consistent elements to facilitate matching (e.g., JUNIOR and II are both converted to JR). In addition, it processes the parsed first name data through a name variant reference file, providing an additional first name field for matching.

[48] 52,210 were dropped.

[49] 165,603 were dropped. Note this number is higher than the one for mothers. This is likely to be the case because i) the age range for the father was greater (25 more years in range), leading to more ties, which are then dropped, and ii) names are passed on from one generation to the next more frequently for males, leading to more ties.

Table A1: Characteristics of Correct Kidlink Child-Parent Links, Restrictive Outcome Measure - Kidlink & 2007 IRS Comparison

	Number			Percent[1]		
	Total	Incorrect Link	Correct Link	Incorrect Link	Correct Link	Pctg. Point Difference
Total	65,129,570	6,552,085	58,577,485	100.0%	100.0%	
Child Age (as of 2007)						
0	959,870	138,298	821,572	2.1%	1.4%	-0.7%
1	3,261,762	447,818	2,813,944	6.8%	4.8%	-2.0%
2	3,313,243	392,340	2,920,903	6.0%	5.0%	-1.0%
3	3,343,577	347,047	2,996,530	5.3%	5.1%	-0.2%
4	3,307,579	313,991	2,993,588	4.8%	5.1%	0.3%
5	3,276,119	295,981	2,980,138	4.5%	5.1%	0.6%
6	3,342,526	290,414	3,052,112	4.4%	5.2%	0.8%
7	3,352,854	280,072	3,072,782	4.3%	5.2%	1.0%
8	3,313,998	263,804	3,050,194	4.0%	5.2%	1.2%
9	3,310,391	257,616	3,052,775	3.9%	5.2%	1.3%
10	3,291,670	251,262	3,040,408	3.8%	5.2%	1.4%
11	3,294,899	247,264	3,047,635	3.8%	5.2%	1.4%
12	3,338,679	248,439	3,090,240	3.8%	5.3%	1.5%
13	3,399,488	255,711	3,143,777	3.9%	5.4%	1.5%
14	3,459,653	263,493	3,196,160	4.0%	5.5%	1.4%
15	3,500,772	275,537	3,225,235	4.2%	5.5%	1.3%
16	3,568,680	304,388	3,264,292	4.6%	5.6%	0.9%
17	3,603,347	377,114	3,226,233	5.8%	5.5%	-0.2%
18	3,527,270	510,522	3,016,748	7.8%	5.2%	-2.6%
19	3,363,193	790,974	2,572,219	12.1%	4.4%	-7.7%
Gender						
Female	31,863,952	3,176,942	28,687,010	48.5%	49.0%	0.5%
Male	33,265,230	3,375,071	29,890,159	51.5%	51.0%	-0.5%
Missing	388	72	316	0.0%	0.0%	0.0%
Citizenship Status						
Citizen/Unknown	63,457,645	6,396,530	57,061,115	97.6%	97.4%	-0.2%
Noncitizen – authorized	1,568,908	150,684	1,418,224	2.3%	2.4%	0.1%
Noncitizen –other	103,017	4,871	98,146	0.1%	0.2%	0.1%
Age SSN received						
<1	58,794,151	5,791,151	53,003,000	88.4%	90.5%	2.1%
1-2	2,551,080	271,501	2,279,579	4.1%	3.9%	-0.3%
2-3	1,635,808	219,573	1,416,235	3.4%	2.4%	-0.9%
3-4	678,637	97,726	580,911	1.5%	1.0%	-0.5%
4-5	321,484	43,234	278,250	0.7%	0.5%	-0.2%
5+	1,148,410	128,900	1,019,510	2.0%	1.7%	-0.2%
Region SSN received						
NE -New England	3,050,815	264,561	2,786,254	4.0%	4.8%	0.7%
NE -Mid Atlantic	8,717,146	767,397	7,949,749	11.7%	13.6%	1.9%
Midwest –East	10,455,007	962,728	9,492,279	14.7%	16.2%	1.5%
Midwest –West	4,501,254	376,755	4,124,499	5.8%	7.0%	1.3%
S –Atlantic	11,388,861	1,289,665	10,099,196	19.7%	17.2%	-2.4%
S -East Central	3,751,330	464,419	3,286,911	7.1%	5.6%	-1.5%
S -West Central	7,984,310	947,608	7,036,702	14.5%	12.0%	-2.5%
W –Mountain	3,871,292	366,555	3,504,737	5.6%	6.0%	0.4%
W –Pacific	10,956,330	1,045,214	9,911,116	16.0%	16.9%	1.0%
Unknown	453,225	67,183	386,042	1.0%	0.7%	-0.4%

Sources: 2007 Kidlink file, 2007 IRS.
[1] Column percentages may not add up to 100% due to rounding.

Table A2: Characteristics of Correct Kidlink Child-Parent Links, Non-restrictive Outcome Measure - Kidlink & 2010 Census Comparison

	Number			Percent[1]		
	Incorrect Link	Correct Link	Total	Incorrect Link	Correct Link	Pctg. Point Difference
Total	6,489,845	49,614,761	56,104,606	100%	100%	
Child Age (as of 2010)						
3	60,722	779,271	839,993	0.9%	1.6%	0.6%
4	193,671	2,660,746	2,854,417	3.0%	5.4%	2.4%
5	168,328	2,754,284	2,922,612	2.6%	5.6%	3.0%
6	153,110	2,815,280	2,968,390	2.4%	5.7%	3.3%
7	145,675	2,805,691	2,951,366	2.2%	5.7%	3.4%
8	144,519	2,789,009	2,933,528	2.2%	5.6%	3.4%
9	150,309	2,856,822	3,007,131	2.3%	5.8%	3.4%
10	155,851	2,874,921	3,030,772	2.4%	5.8%	3.4%
11	159,096	2,844,596	3,003,692	2.5%	5.7%	3.3%
12	163,305	2,842,028	3,005,333	2.5%	5.7%	3.2%
13	169,458	2,825,130	2,994,588	2.6%	5.7%	3.1%
14	179,307	2,822,555	3,001,862	2.8%	5.7%	2.9%
15	195,882	2,845,444	3,041,326	3.0%	5.7%	2.7%
16	215,981	2,870,951	3,086,932	3.3%	5.8%	2.5%
17	247,536	2,879,503	3,127,039	3.8%	5.8%	2.0%
18	321,747	2,677,988	2,999,735	5.0%	5.4%	0.4%
19	512,682	2,041,810	2,554,492	7.9%	4.1%	-3.8%
20	810,679	1,719,024	2,529,703	12.5%	3.5%	-9.0%
21	1,087,516	1,531,714	2,619,230	16.8%	3.1%	-13.7%
22	1,254,471	1,377,994	2,632,465	19.3%	2.8%	-16.6%
Child Gender						
Female	3,411,158	24,066,294	27,477,452	52.6%	48.5%	-4.1%
Male	3,078,556	25,548,293	28,626,849	47.4%	51.5%	4.1%
Missing	131	174	305	0.0%	0.0%	0.0%
Citizenship Status						
Citizen/Unknown	6,329,689	48,461,903	54,791,592	97.5%	97.7%	0.1%
Noncitizen – authorized	151,443	1,123,998	1,275,441	2.3%	2.3%	-0.1%
Noncitizen –other	8,713	28,860	37,573	0.1%	0.1%	-0.1%
Age SSN received						
<1	5,182,206	45,913,889	51,096,095	79.9%	92.5%	12.7%
1-2	480,176	1,554,083	2,034,259	7.4%	3.1%	-4.3%
2-3	433,051	846,401	1,279,452	6.7%	1.7%	-5.0%
3-4	172,261	360,629	532,890	2.7%	0.7%	-1.9%
4-5	65,636	188,515	254,151	1.0%	0.4%	-0.6%
5+	156,515	751,244	907,759	2.4%	1.5%	-0.9%
Region SSN received						
NE -New England	245,291	2,390,625	2,635,916	3.8%	4.8%	1.0%
NE -Mid Atlantic	678,568	6,791,239	7,469,807	10.5%	13.7%	3.2%
Midwest -East	1,050,414	8,138,957	9,189,371	16.2%	16.4%	0.2%
Midwest -West	493,312	3,474,327	3,967,639	7.6%	7.0%	-0.6%
S -Atlantic	1,124,534	8,546,506	9,671,040	17.3%	17.2%	-0.1%
S -East Central	417,245	2,777,163	3,194,408	6.4%	5.6%	-0.8%
S -West Central	878,160	5,895,711	6,773,871	13.5%	11.9%	-1.6%
W -Mountain	455,287	2,898,673	3,353,960	7.0%	5.8%	-1.2%
W -Pacific	1,066,498	8,409,670	9,476,168	16.4%	16.9%	0.5%
Unknown	80,536	291,890	372,426	1.2%	0.6%	-0.7%

Table A2 (Cont'd): Characteristics of Correct Kidlink Child-Parent Links, Non-restrictive Outcome Measure - Kidlink & 2010 Census Comparison

	Number			Percent[1]		
	Incorrect Link	Correct Link	Total	Incorrect Link	Correct Link	Pctg. Point Difference
Child Relationship						
Householder	1,419,879	158,706	1,578,585	21.9%	0.3%	-21.6%
Husband/Wife	256,240	74,545	330,785	3.9%	0.2%	-3.8%
Biological Son/Daug.	1,604,890	43,090,960	44,695,850	24.7%	86.9%	62.1%
Adopted Son/Daug.	93,212	738,280	831,492	1.4%	1.5%	0.1%
Stepson/Daughter	300,288	2,026,566	2,326,854	4.6%	4.1%	-0.5%
Brother/Sister	170,832	208,201	379,033	2.6%	0.4%	-2.2%
Father/Mother	8,055	81,191	89,246	0.1%	0.2%	0.0%
Grandchild	804,000	1,856,401	2,660,401	12.4%	3.7%	-8.6%
Parent-In-Law	2,618	2,737	5,355	0.0%	0.0%	0.0%
Son/Daughter-In-law	77,791	22,036	99,827	1.2%	0.0%	-1.2%
Other Relative	308,744	510,490	819,234	4.8%	1.0%	-3.7%
Not Related	1,443,296	844,648	2,287,944	22.2%	1.7%	-20.5%
Race						
White alone	4,326,248	33,736,798	38,063,046	66.7%	68.0%	1.3%
Black alone	969,000	6,275,897	7,244,897	14.9%	12.6%	-2.3%
AIAN alone	103,773	486,733	590,506	1.6%	1.0%	-0.6%
Asian alone	148,387	2,058,921	2,207,308	2.3%	4.1%	1.9%
NHPI alone	13,518	85,459	98,977	0.2%	0.2%	0.0%
SOR alone	346,850	2,380,802	2,727,652	5.3%	4.8%	-0.5%
Two or more	260,675	2,278,795	2,539,470	4.0%	4.6%	0.6%
Missing	321,394	2,311,356	2,632,750	5.0%	4.7%	-0.3%
Hispanic Origin						
Not Hispanic	5,085,156	39,205,895	44,291,051	78.4%	79.0%	0.7%
Hispanic	1,120,826	8,503,450	9,624,276	17.3%	17.1%	-0.1%
Missing	283,863	1,905,416	2,189,279	4.4%	3.8%	-0.5%

Sources: 2007 Kidlink file, 2010 Census.
[1] Column percentages may not add up to 100% due to rounding.

Table A3: Kidlink & 2010 Census Comparison – Characteristics of Correct Kidlink Child-Parent Links, Restrictive Outcome Measure

	Number			Percent[1]		
	Incorrect Link	Correct Link	Total	Incorrect Link	Correc t Link	Pctg. Point Difference
Total	10,417,412	45,687,194	56,104,606	100%	100%	
Child Age (as of 2010)						
3	131,995	707,998	839,993	1.3%	1.5%	0.3
4	441,649	2,412,768	2,854,417	4.2%	5.3%	1.0
5	427,250	2,495,362	2,922,612	4.1%	5.5%	1.4
6	410,044	2,558,346	2,968,390	3.9%	5.6%	1.7
7	392,223	2,559,143	2,951,366	3.8%	5.6%	1.8
8	381,741	2,551,787	2,933,528	3.7%	5.6%	1.9
9	385,501	2,621,630	3,007,131	3.7%	5.7%	2.0
10	379,204	2,651,568	3,030,772	3.6%	5.8%	2.2
11	360,922	2,642,770	3,003,692	3.5%	5.8%	2.3
12	356,906	2,648,427	3,005,333	3.4%	5.8%	2.4
13	360,157	2,634,431	2,994,588	3.5%	5.8%	2.3
14	370,154	2,631,708	3,001,862	3.6%	5.8%	2.2
15	389,576	2,651,750	3,041,326	3.7%	5.8%	2.1
16	415,522	2,671,410	3,086,932	4.0%	5.8%	1.9
17	450,435	2,676,604	3,127,039	4.3%	5.9%	1.5
18	517,462	2,482,273	2,999,735	5.0%	5.4%	0.5
19	671,700	1,882,792	2,554,492	6.4%	4.1%	-2.3
20	956,707	1,572,996	2,529,703	9.2%	3.4%	-5.7
21	1,229,785	1,389,445	2,619,230	11.8%	3.0%	-8.8
22	1,388,479	1,243,986	2,632,465	13.3%	2.7%	-10.6
Child Gender						
Female	5,302,944	22,025,272	27,328,216	50.9%	48.2%	-2.7
Male	5,052,667	23,381,793	28,434,460	48.5%	51.2%	2.7
Missing	61,801	280,129	341,930	0.6%	0.6%	0.0
Citizenship Status						
Citizen/Unknown	10,196,474	44,595,118	54,791,592	97.9%	97.6%	-0.3
Noncitizen – authorized	211,052	1,064,389	1,275,441	2.0%	2.3%	0.3
Noncitizen –other	9,886	27,687	37,573	0.1%	0.1%	0.0
Age SSN received						
<1	8,805,039	42,291,056	51,096,095	84.5%	92.6%	8.0
1-2	614,878	1,419,381	2,034,259	5.9%	3.1%	-2.8
2-3	511,863	767,589	1,279,452	4.9%	1.7%	-3.2
3-4	204,779	328,111	532,890	2.0%	0.7%	-1.2
4-5	80,488	173,663	254,151	0.8%	0.4%	-0.4
5+	200,365	707,394	907,759	1.9%	1.5%	-0.4
Region SSN received						
NE -New England	447,438	2,188,478	2,635,916	4.3%	4.8%	0.5
NE -Mid Atlantic	1,006,849	6,462,958	7,469,807	9.7%	14.1%	4.5
Midwest -East	1,694,065	7,495,306	9,189,371	16.3%	16.4%	0.1
Midwest -West	770,641	3,196,998	3,967,639	7.4%	7.0%	-0.4
S -Atlantic	1,846,698	7,824,342	9,671,040	17.7%	17.1%	-0.6
S -East Central	663,431	2,530,977	3,194,408	6.4%	5.5%	-0.8
S -West Central	1,395,599	5,378,272	6,773,871	13.4%	11.8%	-1.6
W -Mountain	713,090	2,640,870	3,353,960	6.8%	5.8%	-1.1
W -Pacific	1,764,487	7,711,681	9,476,168	16.9%	16.9%	-0.1
Unknown	115,114	257,312	372,426	1.1%	0.6%	-0.5

Table A3 (cont'd): Kidlink & 2010 Census Comparison – Characteristics of Correct Kidlink Child-Parent Links, Restrictive Outcome Measure

	Number			Percent		
	Incorrect Link	Correct Link	Total	Incorrect Link	Correct Link	Pctg. Point Difference
Race						
White alone	7,064,344	30,998,702	38,063,046	67.8%	67.8%	0.0
Black alone	1,410,923	5,833,974	7,244,897	13.5%	12.8%	-0.8
AIAN alone	173,091	417,415	590,506	1.7%	0.9%	-0.7
Asian alone	250,917	1,956,391	2,207,308	2.4%	4.3%	1.9
NHPI alone	21,356	77,621	98,977	0.2%	0.2%	0.0
SOR alone	535,501	2,192,151	2,727,652	5.1%	4.8%	-0.3
Two or more	455,082	2,084,388	2,539,470	4.4%	4.6%	0.2
Missing	506,198	2,126,552	2,632,750	4.9%	4.7%	-0.2
Hispanic Origin						
Not Hispanic	8,156,642	36,172,037	44,328,679	78.3%	79.2%	0.9
Hispanic	1,819,187	7,767,461	9,586,648	17.5%	17.0%	-0.5
Missing	441,583	1,747,696	2,189,279	4.2%	3.8%	-0.4
Relationship to Householder						
Householder	1,449,597	133,789	1,583,386	13.9%	0.3%	-13.60%
Husband/Wife	277,337	53,640	330,977	2.7%	0.1%	-2.60%
Biological Son/D	4,694,358	39,981,925	44,676,283	45.1%	87.5%	42.40%
Adopted Son/Daughter	134,905	696,510	831,415	1.3%	1.5%	0.20%
Stepson/Daughter	509,915	1,816,099	2,326,014	4.9%	4.0%	-0.90%
Brother/Sister	198,614	181,166	379,780	1.9%	0.4%	-1.50%
Father/Mother	15,822	73,380	89,202	0.2%	0.2%	0.00%
Grandchild	1,095,078	1,570,802	2,665,880	10.5%	3.4%	-7.10%
Parent-In-Law	3,100	2,262	5,362	0.0%	0.0%	0.00%
Son/Daughter-In-	80,664	19,513	100,177	0.8%	0.0%	-0.80%
Other Relative	333,778	284,669	618,447	3.2%	0.6%	-2.60%
Extended roster: related	42,470	159,851	202,321	0.4%	0.3%	-0.10%
Not Related	1,581,774	713,588	2,295,362	15.2%	1.6%	-13.60%

Sources: 2007 Kidlink file, 2010 Census.
[1] Column percentages may not add up to 100% due to rounding.

Table A4: Kidlink & 2010 Census Comparison – Characteristics of Correct Kidlink Child-Parent Links, Non-restrictive Outcome Measure, Biological & Adopted Children Only

	Number			Percent[1]		
	Incorrect Link	Correct Link	Total	Incorrect Link	Correct Link	Pctg. Point Difference
Total	1,698,102	43,829,240	45,527,342	100%	100%	
Child Age (as of 2010)						
3	29,598	690,015	719,613	1.7%	1.6%	-0.2%
4	93,564	2,353,328	2,446,892	5.5%	5.4%	-0.1%
5	81,536	2,437,707	2,519,243	4.8%	5.6%	0.8%
6	74,817	2,496,607	2,571,424	4.4%	5.7%	1.3%
7	72,480	2,491,333	2,563,813	4.3%	5.7%	1.4%
8	73,453	2,477,044	2,550,497	4.3%	5.7%	1.3%
9	77,797	2,537,184	2,614,981	4.6%	5.8%	1.2%
10	82,497	2,550,522	2,633,019	4.9%	5.8%	1.0%
11	85,755	2,522,673	2,608,428	5.1%	5.8%	0.7%
12	89,194	2,519,767	2,608,961	5.3%	5.7%	0.5%
13	93,543	2,504,293	2,597,836	5.5%	5.7%	0.2%
14	99,135	2,499,155	2,598,290	5.8%	5.7%	-0.1%
15	107,472	2,514,276	2,621,748	6.3%	5.7%	-0.6%
16	115,278	2,531,115	2,646,393	6.8%	5.8%	-1.0%
17	122,609	2,535,631	2,658,240	7.2%	5.8%	-1.4%
18	118,695	2,354,262	2,472,957	7.0%	5.4%	-1.6%
19	97,605	1,779,695	1,877,300	5.7%	4.1%	-1.7%
20	78,364	1,494,509	1,572,873	4.6%	3.4%	-1.2%
21	59,594	1,335,018	1,394,612	3.5%	3.0%	-0.5%
22	45,116	1,205,106	1,250,222	2.7%	2.7%	0.1%
Child Gender						
Female	791,108	21,222,027	22,013,135	46.6%	48.4%	1.8%
Male	906,982	22,607,061	23,514,043	53.4%	51.6%	-1.8%
Missing	12	152	164	0.0%	0.0%	0.0%
Citizenship Status						
Citizen/Unknown	1,664,935	42,831,421	44,496,356	98.0%	97.7%	-0.3%
Noncitizen – authorized	30,986	972,691	1,003,677	1.8%	2.2%	0.4%
Noncitizen –other	2,181	25,128	27,309	0.1%	0.1%	-0.1%
Age SSN received						
<1	1,571,654	40,553,549	42,125,203	92.6%	92.5%	-0.03%
1-2	54,352	1,396,856	1,451,208	3.2%	3.2%	-0.01%
2-3	29,034	759,404	788,438	1.7%	1.7%	0.02%
3-4	12,842	320,593	333,435	0.8%	0.7%	-0.02%
4-5	6,560	166,398	172,958	0.4%	0.4%	-0.01%
5+	23,660	632,440	656,100	1.4%	1.4%	0.05%
Region SSN received						
NE -New England	71,385	2,182,634	2,254,019	4.2%	5.0%	0.8%
NE -Mid Atlantic	203,905	6,120,500	6,324,405	12.0%	14.0%	2.0%
Midwest -East	253,567	7,275,646	7,529,213	14.9%	16.6%	1.7%
Midwest -West	102,619	3,134,976	3,237,595	6.0%	7.2%	1.1%
S -Atlantic	292,807	7,465,149	7,757,956	17.2%	17.0%	-0.2%
S -East Central	110,244	2,406,473	2,516,717	6.5%	5.5%	-1.0%
S -West Central	229,873	5,087,857	5,317,730	13.5%	11.6%	-1.9%
W -Mountain	106,642	2,561,058	2,667,700	6.3%	5.8%	-0.4%
W -Pacific	282,177	7,338,276	7,620,453	16.6%	16.7%	0.1%
Unknown	44,883	256,671	301,554	2.6%	0.6%	-2.1%

Table A4 (cont'd): Kidlink & 2010 Census Comparison – Characteristics of Correct Kidlink Child-Parent Links, Non-restrictive Outcome Measure, Biological & Adopted Children Only

	Number			Percent		
	Incorrect Link	Correct Link	Total	Incorrect Link	Correct Link	Pctg. Point Difference
Race						
White alone	1,022,455	30,250,454	31,272,909	60.2%	69.0%	8.8%
Black alone	301,762	5,352,182	5,653,944	17.8%	12.2%	-5.6%
AIAN alone	28,355	405,653	434,008	1.7%	0.9%	-0.7%
Asian alone	40,585	1,875,151	1,915,736	2.4%	4.3%	1.9%
NHPI alone	3,236	69,360	72,596	0.2%	0.2%	0.0%
SOR alone	124,735	2,052,037	2,176,772	7.3%	4.7%	-2.7%
Two or more	82,231	1,996,077	2,078,308	4.8%	4.6%	-0.3%
Missing	94,743	1,828,326	1,923,069	5.6%	4.2%	-1.4%
Hispanic Origin						
Not Hispanic	1,216,115	35,107,919	36,324,034	71.6%	80.1%	8.5%
Hispanic	414,309	7,282,554	7,696,863	24.4%	16.6%	-7.8%
Missing	67,678	1,438,767	1,506,445	4.0%	3.3%	-0.7%

Sources: 2007 Kidlink file, 2010 Census.
[1] Column percentages may not add up to 100% due to rounding.

Table A5: Logit Results – Child-Parent Kidlink & 2010 Census Comparison – Restrictive Outcome Measure

Census	All				Both Parents			Only Mother			Only Father		
Parent link													
Both													
Father	0.5***	0.5***	0.6***	0.6***									
Mother	1.4***	1.4***	3.0***	3.0***									
Child Age (as of 2007)													
<1													
1-3	1.1***	1.1***	1.1***	1.1***	1.0	1.0***	1.0***	1.4***	1.4***	1.4***	1.1***	1.1***	1.1***
4-7	1.2***	1.2***	1.1***	1.1***	1.1***	1.1***	1.0***	2.2***	2.2***	2.0***	1.0	1.0	1.0
8-15	1.1***	1.2***	1.0***	1.0***	1.0*	1.0	0.9***	1.9***	1.9***	1.6***	0.8***	0.8***	0.9***
16+	0.2***	0.3***	0.5***	0.5***	0.2***	0.2***	0.5***	0.3***	0.3***	0.5***	0.3***	0.3***	0.6***
Gender													
Male													
Female	0.9***	0.9***	1.0***	1.0***	0.9***	0.9***	1.0***	0.9***	0.9***	1.0***	0.8***	0.8***	0.9***
Citizenship Status													
Citizen/Unknown													
Noncitizen – authorized	2.0***	2.3***	2.0***	2.0***	1.9***	1.7***	1.7***	2.1***	1.6***	2.0***	2.9***	2.8***	2.1***
Noncitizen –other	1.9***	2.1***	2.0***	2.0***	1.3***	1.2***	1.3***	1.4***	1.1***	1.2***	1.8***	1.7***	1.4***
Region SSN received													
Midwest -East													
Midwest -West	0.9***	0.9***	1.0***	1.0***	0.9***	0.9***	1.0***	0.8***	0.8***	0.9***	0.9***	0.9***	1.0*
NE -Mid Atlantic	1.4***	1.4***	1.1***	1.1***	1.0***	1.0***	1.0***	1.7***	1.8***	1.4***	1.4***	1.4***	1.3***
NE -New England	1.1***	1.1***	1.0***	1.0***	1.0***	1.0***	1.0***	1.2***	1.2***	1.1***	1.1***	1.1***	1.0
S -Atlantic	0.9***	0.9***	0.9***	0.9***	0.8***	0.8***	0.9***	1.0	1.0**	1.0***	1.2***	1.2***	1.2***
S -East Central	0.8***	0.8***	0.8***	0.8***	0.8***	0.8***	0.8***	0.9***	0.9***	0.9***	0.9***	0.9***	1.0*
S -West Central	0.8***	0.8***	0.9***	0.9***	0.8***	0.8***	0.8***	0.9***	0.9***	0.8***	1.3***	1.3***	1.2***
W -Mountain	0.8***	0.8***	0.9***	0.9***	0.8***	0.8***	0.9***	0.8***	0.8***	0.8***	1.1***	1.1***	1.1***
W -Pacific	1.1***	1.1***	1.1***	1.1***	0.9***	0.9***	1.0***	1.0***	1.0***	1.0	1.9***	1.9***	1.5***
Unknown	0.5***	0.5***	0.4***	0.4***	0.4***	0.4***	0.4***	0.9***	0.9***	0.6***	0.6***	0.6***	0.4***

Table A5 (cont'd) : Logit Results – Child-Parent Kidlink & 2010 Census Comparison – Restrictive Outcome Measure

Census	All				Both Parents			Only Mother			Only Father		
Age SSN received													
<1													
1-2		0.9***		1.0***	0.9***		1.0***	0.8***		0.9***	1.0***		1.0***
2-3		0.8***		0.9***	0.8***		0.9***	0.7***		0.8***	0.9***		1.0*
3-4		0.7***		0.9***	0.8***		0.9***	0.7***		0.8***	0.9***		1.1***
4-5		0.8***		1.0***	0.8***		1.0***	0.7***		0.9***	0.9***		1.1***
5+		0.8***		1.1***	0.9***		1.1***	0.7***		0.9***	0.9***		1.2***
Short parent name	1.0***	1.0***	1.0***	1.0***	1.1***	1.1***	1.0	1.1***	1.1***	1.1***	0.9***	0.9***	0.9***
Race													
White alone													
Black alone			0.7***	0.7***			0.4***			1.2***			0.7***
AIAN/NHPI alone			0.6***	0.6***			0.5***			0.8***			1.0
Asian alone			1.9***	1.9***			1.6***			1.6***			3.0***
SOR alone			1.0***	1.0***			0.9***			1.0***			1.1***
Two or more			0.9***	0.9***			0.8***			1.1***			1.0**
Missing			1.2***	1.3***			1.2***			1.2***			1.4***
Hispanic Origin													
Not Hispanic													
Hispanic			0.9***	0.9***			0.8***			1.2***			1.3***
Missing			1.3***	1.3***			1.3***			1.2***			1.5***
Child Relation to Householder													
Biological Son/Daughter													
Adopted Son/Daughter			0.5***	0.5***			0.9***			0.2***			0.8***
Brother/Sister			0.1***	0.1***			0.1***			0.1***			0.2***
Grandchild			0.1***	0.1***			0.1***			0.1***			0.2***
Non relative/Unknown			0.1***	0.1***			0.05***			0.1***			0.1***
Other relative			0.1***	0.1***			0.1***			0.1***			0.1***
Parent figure			0.03***	0.02***			0.02***			0.02***			0.1***
Stepson/Daughter			0.3***	0.3***			0.01***			0.9***			0.2***

Notes: *, ** and *** represent significance at the 10, 5 and 1 percent level respectively.
Sources: 2007 Kidlink file, 2010 Census.

Table A6: Logit Results – Child-Parent Kidlink & 2010 Census Comparison – Non-restrictive Outcome Measure

	All				Both Parents		Only Mother		Only Father	
Parent link										
Both										
Father	0.1***	0.1***	0.1***	0.1***						
Mother	0.3***	0.3***	0.4***	0.4***						
Child Age (as of 2007)										
<1										
1-3	1.2***	1.3***	1.3***	1.3***	1.4***	1.4***	1.4***	1.4***	1.1***	1.1***
4-7	1.6***	1.6***	1.6***	1.6***	2.1***	1.8***	2.2***	2.0***	1.0	1.0
8-15	1.3***	1.3***	1.2***	1.2***	1.5***	1.2***	1.9***	1.6***	0.8***	0.9***
16+	0.1***	0.2***	0.3***	0.3***	0.04***	0.1***	0.3***	0.5***	0.3***	0.6***
Gender										
Male										
Female	0.8***	0.8***	0.9***	0.9***	0.7***	0.8***	0.9***	1.0***	0.8***	0.9***
Citizenship Status										
Citizen/Unknown										
Noncitizen-authorized	2.4***	3.1***	2.4***	2.5***	2.2***	2.1***	2.1***	2.0***	2.9***	2.1***
Noncitizen-other	2.5***	3.2***	2.5***	2.8***	1.6***	1.3***	1.4***	1.2***	1.8***	1.4***
Region SSN received										
Midwest –East	0.8***	0.8***	0.9***	0.9***	0.7***	0.8***	0.8***	0.9***	0.9***	1.0*
Midwest –West	1.6***	1.6***	1.3***	1.3***	1.3***	1.2***	1.7***	1.4***	1.4***	1.3***
NE -Mid Atlantic	1.3***	1.3***	1.1***	1.1***	1.4***	1.2***	1.2***	1.1***	1.1***	1.0
NE -New England	1.0***	1.0***	1.0***	1.0***	1.0***	0.9***	1.0***	1.0***	1.2***	1.2***
S –Atlantic	0.9***	0.9***	0.9***	0.9***	0.9***	0.8***	0.9***	0.9***	0.9***	1.0***
S -East Central	0.9***	0.9***	0.9***	0.9***	0.8***	0.7***	0.9***	0.8***	1.3***	1.2***
S -West Central	0.8***	0.8***	0.8***	0.8***	0.7***	0.7***	0.8***	0.8***	1.1***	1.1***
W –Mountain	1.2***	1.3***	1.2***	1.2***	1.2***	1.1***	1.0***	1.0	1.9***	1.5***
W –Pacific	0.7***	0.7***	0.4***	0.4***	0.9***	0.3***	0.9***	0.6***	0.6***	0.4***
Unknown	1.0***	1.0***	1.0***	1.0***	1.0***	0.9***	1.0***	1.0***	1.2***	1.2***

Table A6 (cont'd): Logit Results – Child-Parent Kidlink & 2010 Census Comparison – Non-restrictive Outcome Measure

	All				Both Parents		Only Mother		Only Father	
Age SSN received										
<1										
1-2		0.8***		0.9***	0.8***	0.9***	0.8***	0.9***	1.0***	1.0***
2-3		0.6***		0.7***	0.6***	0.8***	0.7***	0.8***	0.9***	1.0**
3-4		0.6***		0.8***	0.6***	0.8***	0.7***	0.8***	0.9***	1.1***
4-5		0.7***		0.9***	0.7***	0.8***	0.7***	0.9***	0.9***	1.1***
5+		0.7***		1.0***	0.7***	1.0	0.7***	0.9***	0.9***	1.2***
Short Parent Name	0.9***	0.9***	0.9***	0.9***	1.1***	1.0	1.1***	1.1***	0.9***	0.9***
Race										
White alone										
Black alone			1.0***	1.0***		1.0***		1.2***		0.7***
AIAN/NHPI alone			0.7***	0.7***		0.5***		0.8***		1.0***
Asian alone			2.1***	2.1***		1.6***		1.6***		3.0***
SOR alone			1.1***	1.1***		1.0***		1.0***		1.1***
Two or more			1.0***	1.0***		1.0***		1.1***		1.0**
Missing			1.5***	1.5***		1.7***		1.2***		1.4***
Hispanic Origin										
Not Hispanic										
Hispanic			1.2***	1.2***		1.3***		1.2***		1.3***
Missing			1.6***	1.6***		2.2***		1.2***		1.5***
Child Relation to Householder										
Biological Son/Daughter										
Adopted Son/Daughter			0.3***	0.3***		0.2***		0.2***		0.8***
Brother/Sister			0.1***	0.1***		0.01***		0.1***		0.2***
Grandchild			0.1***	0.1***		0.01***		0.1***		0.2***
Non relative/Unknown			0.0***	0.03***		0.00***		0.1***		0.1***
Other relative			0.0***	0.04***		0.01***		0.1***		0.1***
Parent figure			0.0***	0.01***		0.00***		0.02***		0.1***
Stepson/Daughter			0.4***	0.4***		0.1***		0.9***		0.2***

Notes: *, ** and *** represent significance at the 10, 5 and 1 percent level respectively.
Sources: 2007 Kidlink file, 2010 Census.

Table A7: Logit Results, Kidlink & 2010 Census Comparison – Non Restrictive Outcome Measure, Only Biological & Adopted Children

	All				Both Parents		Only Mother		Only Father	
Parent link										
Both										
Father	0.02 ***	0.02 ***	0.02 ***	0.02 ***						
Mother	0.1 ***	0.1 ***	0.1 ***	0.1 ***						
Child Age (as of 2007)										
<1										
1-3	1.3 ***	1.3 ***	1.3 ***	1.3 ***	1.5 ***	1.5 ***	1.4 ***	1.4 ***	1.0 ***	1.0 ***
4-7	1.6 ***	1.6 ***	1.6 ***	1.6 ***	2.2 ***	2.3 ***	2.3 ***	2.3 ***	0.9 ***	0.9 ***
8-15	1.4 ***	1.4 ***	1.4 ***	1.4 ***	2.2 ***	2.2 ***	2.0 ***	2.0 ***	0.8 ***	0.8 ***
16+	1.2 ***	1.2 ***	1.3 ***	1.3 ***	1.5 ***	1.6 ***	1.6 ***	1.6 ***	0.8 ***	0.9 ***
Gender										
Male										
Female	1.1 ***	1.1 ***	1.1 ***	1.1 ***	1.0 ***	1.0 ***	1.2 ***	1.2 ***	0.9 ***	0.9 ***
Citizenship Status										
Citizen/Unknown										
Noncitizen-authorized	2.4 ***	2.4 ***	2.1 ***	2.1 ***	1.0	1.2 ***	1.8 ***	1.7 ***	2.6 ***	2.2 ***
Noncitizen-other	2.0 ***	1.9 ***	1.8 ***	1.8 ***	0.3 ***	0.5 ***	1.0	1.0	1.8 ***	1.5 ***
Region SSN received										
Midwest -East										
Midwest -West	0.9 ***	0.9 ***	1.0 ***	1.0 ***	0.9 ***	0.9 ***	0.8 ***	0.9 ***	1.0 ***	1.0 ***
NE -Mid Atlantic	1.5 ***	1.5 ***	1.5 ***	1.5 ***	0.7 ***	0.8 ***	1.5 ***	1.6 ***	1.3 ***	1.2 ***
NE -New England	1.1 ***	1.1 ***	1.0 ***	1.0 ***	0.9 ***	0.9 ***	1.1 ***	1.2 ***	1.0 ***	1.0 ***
S -Atlantic	1.0 ***	1.0 ***	1.0 ***	1.0 ***	0.6 ***	0.7 ***	1.0	1.0 ***	1.2 ***	1.3 ***
S -East Central	0.9 ***	0.9 ***	0.9 ***	0.9 ***	0.5 ***	0.5 ***	0.9 ***	0.9 ***	0.9 ***	1.1 ***
S -West Central	1.0	1.0 ***	1.0 ***	1.0 ***	0.4 ***	0.6 ***	0.8 ***	0.8 ***	1.4 ***	1.3 ***
W -Mountain	0.9 ***	0.9 ***	0.9 ***	0.9 ***	0.5 ***	0.6 ***	0.7 ***	0.8 ***	1.2 ***	1.1 ***
W -Pacific	1.3 ***	1.3 ***	1.1 ***	1.1 ***	0.6 ***	0.7 ***	0.9 ***	0.9 ***	1.9 ***	1.5 ***
Unknown	0.4 ***	0.4 ***	0.4 ***	0.4 ***	0.1 ***	0.2 ***	0.6 ***	0.6 ***	0.5 ***	0.4 ***

Table A7 (cont'd): Logit Results, Kidlink & 2010 Census Comparison – Non Restrictive Outcome Measure, Only Biological & Adopted Children

	All				Both Parents		Only Mother		Only Father	
Age SSN received										
<1										
1-2	0.95 ***		0.95		1.0	1.0	0.9 ***	0.9 ***	1.1 ***	1.1 ***
2-3	0.99		1.00		1.0	1.0	0.9 ***	0.9 ***	1.2 ***	1.2 ***
3-4	1.00		1.01		0.9 ***	0.9 ***	0.8 ***	0.9 ***	1.2 ***	1.3 ***
4-5	1.02		1.04 *		0.8 ***	0.8 ***	0.8 ***	0.8 ***	1.3 ***	1.4 ***
5+	1.00		1.00		0.6 ***	0.6 ***	0.8 ***	0.8 ***	1.3 ***	1.4 ***
Short Parent Name	1.0	1.0	1.0 *	1.0 **	1.1 ***	1.0 ***	1.0 ***	1.1 ***	0.9 ***	0.9 ***
Race										
White alone										
Black alone			1.0 ***	1.0 ***		0.4 ***		1.5 ***		0.6 ***
AIAN/NHPI alone			0.7 ***	0.7 ***		0.2 ***		0.7 ***		0.9 ***
Asian alone			2.6 ***	2.6 ***		0.8 ***		1.4 ***		3.7 ***
SOR alone			1.0 ***	1.0 ***		0.5 ***		0.9 ***		1.1 ***
Two or more			1.1 ***	1.1 ***		0.7 ***		1.2 ***		1.0
Missing			1.1 ***	1.1 ***		0.7 ***		1.0 *		1.3 ***
Hispanic Origin										
Not Hispanic										
Hispanic			1.2 ***	1.2 ***		0.6 ***		1.2 ***		1.3 ***
Missing			1.0 ***	1.0 ***		0.6 ***		0.9 ***		1.1 ***

Notes: *, ** and *** represent significance at the 10, 5 and 1 percent level respectively.

Sources: 2007 Kidlink file, 2010 Census.

A8: Census Bureau Regions & Divisions

Region & Division	States
Northeast Region	
New England	Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, Vermont
Middle Atlantic	New Jersey, New York, Pennsylvania
Midwest	
East North Central	Indiana, Illinois, Michigan, Ohio, Wisconsin
West North Central	Iowa, Kansa, Minnesota, Missouri, Nebraska, North Dakota, South Dakota
South	
South Atlantic	Delaware, District of Columbia, Florida, Georgia, Maryland, Nrth Carolina, South Carolina, Virginia, West Virginia
East South Central	Alabama, Kentucky, Mississippi, Tennessee
West South Central	Arkansas, Louisiana, Oklahoma, Texas
West	
Mountain	Arizona, Colorado, Idaho, New Mexico, Montana, Utah, Nevada, Wyoming
Pacific	Alaska, California, Hawaii, Oregon, Washington

Source: U.S. Census Bureau