Birth of identity: understanding changes to birth certificates and their value for identity resolution

Jeffrey Duncan,[1] Scott P Narus,[1,2] Stephen Clyde,[3] Karen Eilbeck,[1] Sidney Thornton,[1,2] Catherine Staes[1]

[1]Department of Biomedical Informatics, University of Utah, Salt Lake City, Utah, USA
[2]Intermountain Healthcare, Salt Lake City, Utah, USA
[3]Department of Computer Science, Utah State University, Logan, Utah, USA

Correspondence to
Jeffrey Duncan, Department of Biomedical Informatics, University of Utah, 421 Wakara Way, Suite 140, Salt Lake City, UT 84108, USA; jeff.duncan@utah.edu

Received 4 March 2014
Revised 23 June 2014
Accepted 13 July 2014

ABSTRACT

Introduction Identity information is often used to link records within or among information systems in public health and clinical settings. The quality and stability of birth certificate identifiers impacts both the success of linkage efforts and the value of birth certificate registries for identity resolution.

Objective Our objectives were to describe: (1) the frequency and cause of changes to birth certificate identifiers as children age, and (2) the frequency of events (ie, adoptions, paternities, amendments) that may trigger changes and their impact on names.

Methods We obtained two de-identified datasets from the Utah birth certificate registry: (1) change history from 2000 to 2012, and (2) occurrences for adoptions, paternities, and amendments among births in 1987 and 2000. We conducted cohort analyses for births in 1987 and 2000, examining the number, reason, and extent of changes over time. We conducted cross-sectional analyses to assess the patterns of changes between 2000 and 2012.

Results In a cohort of 48 350 individuals born in 2000 in Utah, 3164 (6.5%) experienced a change in identifiers prior to their 13th birthday, with most changes occurring before 2 years of age. Cross-sectional analysis showed that identifiers are stable for individuals over 5 years of age, but patterns of changes fluctuate considerably over time, potentially due to policy and social factors.

Conclusions Identities represented in birth certificates change over time. Specific events that cause changes to birth certificates also fluctuate over time. Understanding these changes can help in the development of automated strategies to improve identity resolution.

BACKGROUND AND SIGNIFICANCE

In a 2013 recommendation to the US Congress, the Health Information and Management Systems Society (HIMSS) wrote, 'One of the largest unresolved issues in the safe and secure electronic exchange of health information is the need for a nationwide patient data matching strategy to ensure the accurate, timely, and efficient matching of patients with their healthcare data across different systems and settings of care.'[1] Despite considerable research in record linkage[2–9] and the creation of enterprise master person indices,[10–13] problems with identity resolution continue to challenge efforts to improve the delivery of quality healthcare.

With the growth of data sharing initiatives such as health information exchange[14] and comparative effectiveness research,[15] the need to link patient records across institutions and organizations presents a growing challenge that is exacerbated by the lack of a unique national identifier for healthcare in the USA.

The relevance of birth certificates

For the 95% of births occurring in hospitals in Utah,[16] birth certificates are reported to the state by hospital medical records staff. Each birth certificate includes demographic information obtained from parents who indicate the desired name of the child and other information such as race and ethnicity. This birth certificate identity often propagates to other public health information systems such as immunization registries, early hearing detection and intervention registries, and metabolic and other newborn screening systems. While the identity information submitted at birth identifies a newborn child's name, date of birth, and sex, that information is not necessarily permanent. In its role as a civil registry, and because birth certificates are a foundational identity document, vital statistics offices routinely correct birth certificate information to reflect changes in real identity and to correct mistakes on original records. Understanding these changes and their implications for the use of birth certificates in identity resolution is the goal of this analysis.

As a fundamental source of identity information, there are at least two possible roles birth registries may play in identity resolution for healthcare and public health. First, birth registries may be used as a source in a hierarchical master person index (MPI) incorporating clinical and public health sources such as Utah's statewide MPI.[17] Second, automated queries to birth registries may be used to facilitate the resolution of potential record matches.

While we know that birth certificate information is subject to change at any age and for multiple reasons, to date there has been no assessment of the number, frequency, reasons, and age distribution of changes on birth certificates. We propose that understanding changes made to the assigned identities on birth certificates could improve record matching strategies. The goal of this project was to document and understand the frequency and types of changes to birth certificates and to assess the value of this information for improving identity resolution across healthcare and public health. As the original source for and a registry of changes to an individual's legal name, date of birth, and sex, birth certificates have great potential for identity resolution. However, to fully utilize this potential it is important to understand the stability of birth certificate information. Therefore, the objectives of this analysis were to describe: (1) the frequency

To cite: Duncan J, Narus SP, Clyde S, *et al*. *J Am Med Inform Assoc* Published Online First: [*please include* Day Month Year] doi:10.1136/amiajnl-2014-002774

Table 1 Descriptions of change events that impact birth certificates

Event	Description	Effect on birth certificate
Amendment		
Amendment	A change to correct minor errors or omissions on birth certificates. An amendment requires a signed affidavit and may require documentary evidence.	Identity information is changed and amendment histories are documented on printed birth certificates.
Adoption		
Two-new-parent adoption	A court awards parental rights to two new parents, neither of whom is a biological parent of a child.	The original birth certificate is sealed and a new certificate is issued reflecting the names of the adoptive parents.
Step-parent adoption	A court awards parental rights to a step-parent.	The name of the step-parent is entered on the birth certificate, replacing a biological parent. The original certificate is sealed.
Family adoption	A court awards parental rights to a family member such as an older sibling, aunt, uncle, etc.	The names of the adoptive family member and spouse (if applicable) replace the names of the biological parents. The original certificate is sealed.
Single parent adoption	A court awards parental rights to a single person, either male or female.	The birth certificate is amended with the name of the single parent as father or mother, as appropriate, and the original certificate is sealed.
Paternity establishment		
Court-ordered paternity	A court determines biological fatherhood and orders a male's name entered as the father on a birth certificate.	The father is listed as the court order decrees. The decree may also change the name of the child.
Voluntary/administrative declaration of paternity	A male voluntarily acknowledges paternity of a child or is administratively determined to be the father by the state's child support enforcement agency.	The child's name on the birth certificate may change at the discretion of the parents.

and causes of changes to birth certificate identifiers as children get older, and (2) the frequency of events (ie, adoptions, paternities, and amendments) that may trigger changes and the impact of the different types of events on a name.

METHODS

Study population

Utah's birth registry system, maintained by the Utah Department of Health, includes information about births in Utah since statewide registration first began in 1905. In 2000, the system started tracking changes (ie, updates) to the information included in the registry. In 2009, the system started tracking more detailed information about adoptions and paternities. The types of change events identified on birth certificates are described in table 1. Each change event may result in changes to the facts recorded on a birth certificate.

Identity classification

Three tiers of identity have been described by Durand (figure 1).[18] Tier 1 is a person's *real* identity: the identity that is

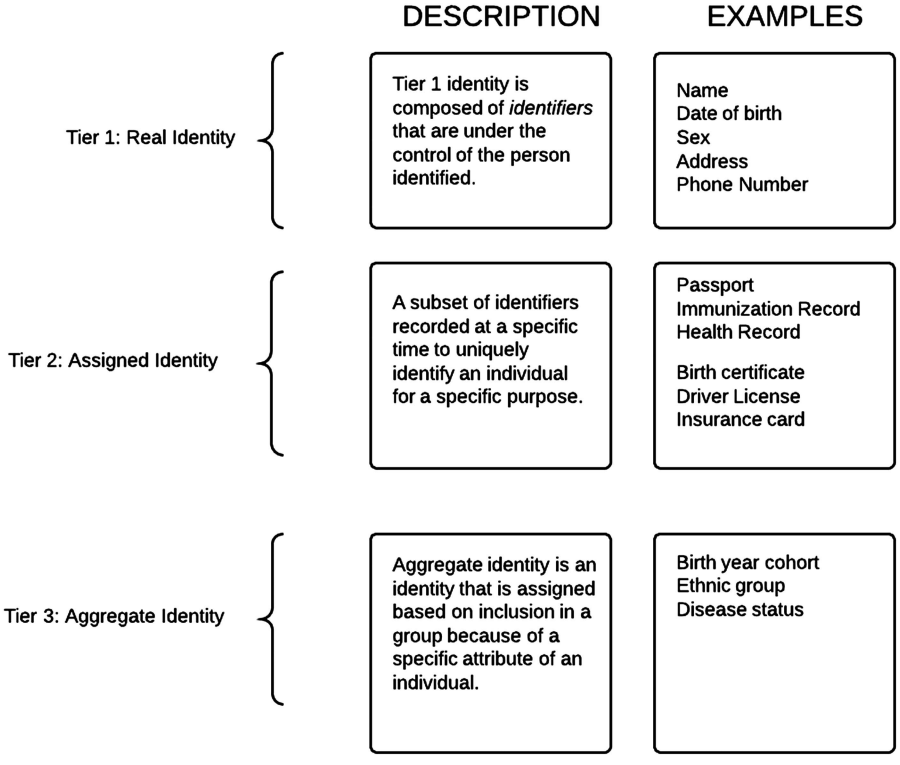

Figure 1 Three tiers of identity with examples.

owned by and completely under the control of the individual or the individual's agent (eg, parent). Tier 2 is an *assigned* identity: an identity that is created by some entity for a specific context or purpose. Tier 3 is an *aggregate* identity: an identity that is assigned based on inclusion in a group because of a specific attribute of the individual.

Real (Tier 1) identities are described by non-unique identifiers, a subset of which are recorded at a point in time to form an assigned (Tier 2) identity. For example, a person's real identity is described by identifiers such as name, sex, date of birth, address, social security number, and so on. Many of these identifiers, including name and address, are subject to change over time at the direction of the individual who owns the identity. Although rare, persons may change their gender identity and the sex recorded in a record, creating challenges for electronic health records.[19]

Tier 2 identities are records or snapshots of Tier 1 identity taken at specific points in time, for a specific purpose. A birth certificate is thus an assigned identity documenting a child's first legal name, sex, date of birth, place of birth, and parentage. For most newborn children born in the USA since 1989, in a process known as Enumeration at Birth,[20] state birth certificate registrars electronically apply for social security numbers by providing identity information from the birth certificate. A driver license is an assigned identity typically created when an individual is around 16 years of age using the birth certificate and social security card as the source of identity, especially the name, date of birth, and sex.

A record in a hospital information system is a Tier 2 identity. As part of the intake process for clinical visits, a person presents identifying information that may include a driver license and an insurance card. Either a new assigned identity is created in the hospital's information system, or information from an existing record is verified and updated if necessary.

Record linkage attempts to associate records identifying the same individual in different electronic systems. De-duplication refers to linkage techniques used to identify multiple instances of the same individual in a single system. Using Durand's model as a framework for understanding, it can be said that record linkage attempts to locate and link two or more different *assigned identities* for a single *real identity*. If two or more assigned identities in the same database refer to the same real identity, they are duplicate records. If assigned identities in two or more separate systems are equivalent, they are usually assumed to refer to the same real identity. Integrated health delivery systems[21] typically use an enterprise master person index (EMPI) to continuously search for, de-duplicate, and link identities across the multiple information systems within the organization. Most individuals can be identified and linked by matching identifiers. Problems arise when assigned identities for the same real identity do not match.

When performing record linkage, two identities assigned to one real individual at different times, t1 and t2, may fail to be identical for two reasons:

> Identity Problem 1: The individual's identifiers changed between t1 and t2. A person may change names, address, phone number, and other identifiers.

> Identity Problem 2: Identifiers listed for the assigned identity were recorded incorrectly at t1, t2, or both. This may be caused by a data entry error, or incomplete or incorrect submission of information from the source. Data entry errors during inpatient registration are a common source of duplicate records in electronic health record systems.[22]

A third problem that confounds linkage methods, which we will call Identity Problem 3, occurs when two assigned identities that appear to belong to the same person in fact do not. This situation may occur when two or more newborns (eg, as a result of multiple births) have the same last name, date of birth, and sex, and different but similar first names, or when a child shares a first and last name, and even a month and day of birth, with a parent.

Of the events identified in table 1, adoptions and paternities reflect changes to a person's Tier 1 identity, resulting in Identity Problem 1. Amendments reflect corrections to the Tier 2 identity recorded on the birth certificate resulting from Identity Problem 2.

Records and fields abstracted for analysis

In November 2013, we abstracted information from Utah's birth registry system, which includes information about all births registered since 1905. We obtained detailed change history information including the date of change, field changed, and reason for changes processed from January 1, 2000 to December 31, 2012. To preserve the confidentiality of persons identified in the records, we obtained information about the fact of change but not the actual pre- or post-change values. The following fields in the record were considered to be 'identify fields': first name, middle name, last name, date of birth, and sex.

We also obtained de-identified information about the dates and occurrence of an adoption, paternity establishment, or any amendments for two birth cohorts (1987 and 2000). This information was linked to change information using each birth certificate's unique file number.

Cohort analysis

We analyzed two cohorts of births (babies born in 1987 and in 2000) to identify the distribution and sequence of identity change events over time. Detailed change history from the Utah birth record system was only available from 2000 onward; therefore, we reviewed records for the cohort born in 2000 to understand changes documented during the first 12 years of life, that is, between 2000 and 2012. Because we were limited to detailed change history after 2000, we reviewed records for the cohort born in 1987 to document changes during the second 12 years of life (ie, when the cohort born in 1987 were 13–25 years of age between 2000 and 2012). We conducted non-parametric univariate survival analysis using SAS software's PROC LIFETEST, using each individual's year of age, not calendar year, as the time variable.[23] For the survival model, we calculated time to the first change to identity fields on the birth certificate for any reason. In the event that individuals experience multiple changes to identity fields at different times, we only used the time to the first change in our analysis. Understanding multiple changes to an individual requires more complex statistical models and is beyond the scope of this analysis. Birth records are flagged upon death to limit illegal use, thus death was used as a censoring event in our analysis.

Cross-sectional analysis

We performed a cross-sectional analysis of changes made to birth certificates between 2000 and 2012 for births in any prior year to identify the frequency, distribution, and reasons for changes to identity information. We summarized the changes recorded to any existing birth certificate during each year. We aimed to understand whether the rate of change by type varied over time. We stratified the findings by age groups 0–2, 3–5, and 6 years and older for each year analyzed. Additional detail captured after upgrades to Utah's birth registry in 2009 allowed us to analyze the frequency of changes to first, middle, and last

Table 2 Summary of analyses and data used in this study

Analysis strategy (results)	Birth certificates in Utah's birth registry included in analysis	Years analyzed for changes to a birth record	Purpose
Cohort (table 3)	All births in 1987 and 2000	2000–2012	To describe the frequency of changes in birth certificate data as children grow older
Cross-sectional (figure 2)	All births in any year prior to the year of analysis that were never marked as deceased	2000–2012	To describe the patterns of changes to birth certificates between 2000 and 2012 for all births with no record of having died in any prior year to the year of analysis
Cross-sectional (figure 3)	All births prior to 2010 that were never marked as deceased	2010	To analyze changes related to adoptions and paternities enabled by changes to Utah's birth registry implemented in 2009

names by event type for all adoptions, paternities, and amendments to births occurring in 2010 in which a name change occurred. Table 2 summarizes the types of analyses, birth records used, and time periods addressed in this study.

RESULTS

Study population

A total of 2 589 265 births have been registered in Utah since 1905. The annual number of births registered in Utah has steadily increased each year from 1406 in 1905 to 51 439 in 2012. Between 2000 and 2012, a total of 685 984 birth records were added to the Utah birth certificate registry. For the cohort analysis, we used the 35 285 births that occurred in 1987 and the 48 350 births that occurred in 2000. For the cross-sectional analysis, we used all birth records never marked as deceased for the denominator.

Cohort analysis: distribution of changes as children age

During the first 13 years of life, 3147 (6.5%) of the 48 350 children born in 2000 had changes to their birth certificate (table 3). Using non-parametric univariate survival analysis, children in their first year of life faced the highest likelihood of change (5%); the probability decreased markedly after the first year to 0.2% during the 13th year of life. The second cohort born in 1987 demonstrated a similar rate of change events (0.1%) during the 13th year of life. The rates of changes remained low with the exception of increases observed for those between 15 and 19 years of age, which are the years when persons usually apply for their first driver's license. Amendments and adoptions triggered over 90% of identity changes for nearly every age group, but the proportions varied by age (table 3).

Cross-sectional analysis: frequency of change events over time

Amendment and adoption events in the first 2 years of life caused the greatest rates of name changes on birth certificates, with changes due to amendments showing the greatest variation: from 1110 per 100 000 births in 2001 to a high of 2736 per 100 000 births in 2010 (figure 2A). The rate of name changes due to adoptions in the first 2 years of life were more consistent, ranging between 350 and 450 per 100 000 births (figure 2B). In contrast, the rate of name changes due to paternity acknowledgment, although relatively low for all age groups, showed considerable variation over time for all age groups (figure 2C).

Detailed cross-sectional analysis: changes in identity fields by event type

The likelihood and extent of a name change varied considerably depending on the type of change event. Figure 3 illustrates changes due to amendments, adoptions, and paternities finalized in 2010 to births in any prior year. Among all records that changed, 55% (n=6459) were due to amendments with changes most frequently occurring to the middle or last name.

Of the 5341 paternity and adoption events that were processed in 2010 to births in any prior year, 2750 (51%) were voluntary/administrative paternities (table 4). Despite being the most frequently occurring event, voluntary paternities cause the fewest name changes compared to other types (figure 3). Among adoption events, two-new-parent adoptions occur more frequently than other types. Step-parent adoptions most frequently result in changes to a child's last name, but not first or middle names. In contrast, nearly two-thirds of two-new-parent adoptions result in changes to both first and last names and over half include changes to a child's first name (figure 3).

DISCUSSION

Our findings indicate that birth certificate identities change over time, particularly in the first years of life, but also the patterns of change fluctuate temporally potentially due to societal factors. The birth record is not static and is subject to change for multiple reasons. Updated information in a birth certificate may reflect corrections or changes in Tier 1 (real) identity. This information may be useful for identity resolution in electronic healthcare and public health record systems.

The timing of changes to birth certificates may be impacted by a variety of processes. Minor typographical errors are often corrected by parents when they request a certified copy of the certificate from the state. In the case of births to unmarried mothers, most jurisdictions will subsequently add a biological father to the birth certificate only after paternity establishment, either voluntary/administrative or court-ordered. Adoption decrees do not automatically trigger immediate changes to birth certificates in Utah, and likely in many other states that follow standard practices under the Model State Vital Statistics Act. Adoption decrees are issued to adoptive parents who then must request a *supplementary* birth certificate from the state vital records office and pay a fee to have the new birth certificate issued and the original record sealed. In practice this often may happen years after the adoption takes place, for example when a child turns 16 and wants to obtain a driver license. Thus, to interpret identity changes due to adoptions, it is important to understand that the year of change due to adoption is the year that the adoptive parents present a court's decree of adoption to obtain a new birth certificate, not necessarily the year when the court issues the decree. In other words, there may be a lag of several years between the time when a court changes a child's Tier 1 identity and the time when his or her parents change the birth certificate (Tier 2) identity. In addition, child adoption

Table 3 Results of non-parametric survival analysis to describe identity change events* that occurred between 2000 and 2012 for two birth cohorts

Age in years (lower, upper)	Number of changes to the following fields			Number of records censored due to death	Effective sample size†	Conditional probability of change (p)	Proportion of reasons for changes observed		
	Name	DOB	Sex				Amendments (%)	Paternity (%)	Adoptions (%)
Cohort born during 2000, with changes recorded January 1, 2000, through December 31, 2012									
0–<1	1831	0	2	264	48 350	0.05	88	10	2
1–<2	346	0	3	18	45 818	0.014	87	11	2
2–<3	159	4	2	14	46 146	0.007	74	20	6
3–<4	131	2	2	11	45 907	0.006	53	11	35
4–<5	135	0	5	11	45 690	0.005	46	8	46
5–<6	207	3	4	6	45 493	0.006	43	12	45
6–<7	105	0	0	9	45 236	0.004	30	3	67
7–<8	98	0	0	8	45 081	0.003	21	3	76
8–<9	54	0	3	7	44 972	0.003	33	5	61
9–<10	18	1	2	5	44 846	0.003	95	5	0
10–<11	17	1	0	8	44 748	0.002	94	6	0
11–<12	10	2	0	2	44 660	0.003	80	20	0
12–<13	16	0	1	2	44 558	0.002	94	6	0
Cohort born in 1987, with changes recorded January 1, 2000 through December 31, 2012									
12–<13	52	0	0	4	36 301	0.001	94	2	4
13–<14	39	0	0	9	36 243	0.001	97	0	3
14–<15	48	0	0	11	36 194	0.001	92	2	6
15–<16	69	0	0	15	36 133	0.002	83	1	16
16–<17	106	1	0	20	36 046	0.003	75	2	23
17–<18	53	1	0	15	35 922	0.002	54	2	44
18–<19	64	1	0	21	35 850	0.002	69	0	31
19–<20	45	0	0	26	35 761	0.001	67	0	33
20–<21	34	1	0	27	35 690	0.001	74	3	23
21–<22	18	1	0	24	35 629	0.001	74	0	26
22–<23	4	0	0	20	35 588	0.000	100	0	0
23–<24	11	0	0	27	35 561	0.000	100	0	0
24–<25	13	0	1	26	35 523	0.000	100	0	0

*Records were censored after the first change event or when death was recorded.
†Effective sample size is the number of persons entering each interval minus half the number censored during the interval.

proceedings in most states are 'closed,' meaning that a new birth certificate is created and the original is sealed when an adoption occurs.[24] In Utah's birth registry, this policy is implemented by changing information in the electronic record, recording a change event, and severing links to the record's previous history.

Birth records are mutable, but the changes to identity fields occur for a limited number of reasons that reflect changes to a person's real identity or corrections to the assigned birth certificate identity. Changes to the date of birth or sex represented in the birth record likely reflect errors (Identify Problem 2) that are corrected after errors are noticed when birth certificates are obtained by parents or teenagers for other reasons. The slight increase in the number of changes for 5-year-olds and 16-year-olds shown in table 3 is likely due to this phenomenon: birth certificates are obtained at these ages when children enter elementary school or teenagers obtain a driver license.

Table 4 Number and percentage of adoption and paternity events processed in Utah in 2010 for births in any prior year

Event	Number	Percentage of events
Two-new-parent adoption	1069	20
Step-parent adoption	845	16
Family adoption	263	5
Other adoption*	203	4
Voluntary/administrative paternity	2750	51
Court-ordered paternity	211	4
Total	5341	100

*Other adoption includes legitimations, single-parent adoptions, and foreign adoptions.

While the overall rate of change events for persons over 5 years of age is very low, the rate of change events for children in the first 2 years of life may be substantial. Our analysis also showed that the rate of events that cause identity changes fluctuates considerably for children in the first 2 years of life. The two peaks in name changes associated with adoptions for infants in 2001 and 2009, shown in figure 2B, correspond to increases in adoptions finalized in Utah courts for the same years.[25] In general, the rate of adoptions of children in foster care in the USA is increasing,[26] and paternities are increasing due to the increase in extramarital births and federal welfare reform laws encouraging state child support enforcement agencies to increase paternity establishment.[27] The sharp increase in name changes due to paternities reflected in figure 2C can be attributed to such efforts in Utah beginning in 2004. Since paternities only result in name changes when filed after birth registration,

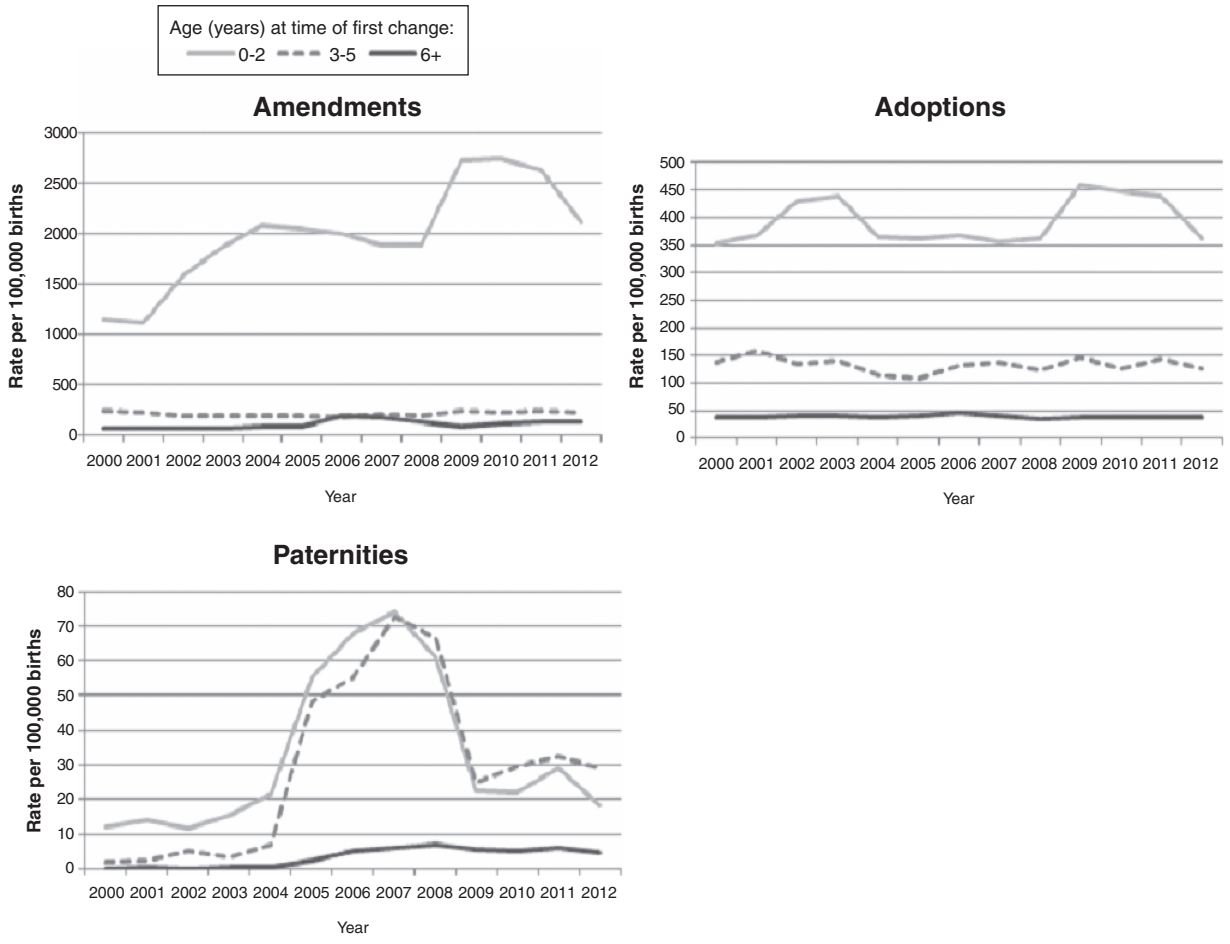

Figure 2 Rate of name changes due to amendments, adoptions, and paternities by year of change and age group of the child, recorded on birth certificates between 2000 and 2012 for births occurring in any year (n=2 589 265 births from 1905 to 2012).

the subsequent decline beginning in 2009 reflects the implementation of a new electronic birth registration system and a fax-to-image system that facilitated in-hospital paternity establishment prior to birth registration. Finally, the sharp increase in amendments beginning in 2009, shown in figure 2A, may be attributed to hospital staff learning the new birth registry software implemented in Utah in January 2009.

Adoptions present numerous challenges for data linkage, not only because names change, but also because those changes are often obscured by adoption privacy laws. In today's electronic exchanges, however, adoption privacy is often breeched when program staff investigate mismatched records and research potential matches, inadvertently discovering the identities of birth mothers. Our results show that the frequency and type of name change varies with the type of adoption. Currently, the records of adopted children must be linked manually, often resulting in inadvertent identification of birth mothers in violation of adoption privacy laws. Knowledge of the occurrence, type, and date of adoption or paternity, and probability of changes to name can aid in the development of automated strategies to improve identity resolution for adoptees while preserving confidentiality.

Record linkage methods can be deterministic or probabilistic.[28 29] Regardless of which methods are used, none achieve 100% sensitivity while maintaining the 100% specificity required for medical records, meaning there is always a non-zero number of real matches who fail for any number of reasons. Many of these records fail to match due to Identity Problems 1 and 2 described above, yet the problems preventing a match could be resolved in an automated way with queries to, or updates from, a birth registry. Currently, Integrating the Healthcare Enterprise (IHE) is working to develop standards and profiles for the collection of birth certificate information from electronic health records (EHRs) in order to automate the birth registration process.[30] Given that 95% of Utah births occur in hospitals, standards for reciprocal exchange may also make sense except in cases such as adoptions where a record is sealed by law. Enabling hospitals to receive allowed updates when birth records change is not currently considered in the standards being developed.

A major limitation of using an electronic birth registry as a data source for identity resolution arises from the fact that any state's birth registry only includes births occurring in the state, not necessarily residents of the state. Very often large segments of any state's residents were likely born in other states or countries, given the mobile nature of today's population. Even so, migration is likely lower during the first 5 years of life when changes to identity are more frequent and there are increased needs to link to the multitude of child health-related systems in public health, including registries for immunizations, newborn hearing screening, metabolic screening, and others.

This analysis has limitations. First, the findings reflect the experience in Utah, which may differ from other states. However, Utah has adopted the Model State Vital Statistics Act

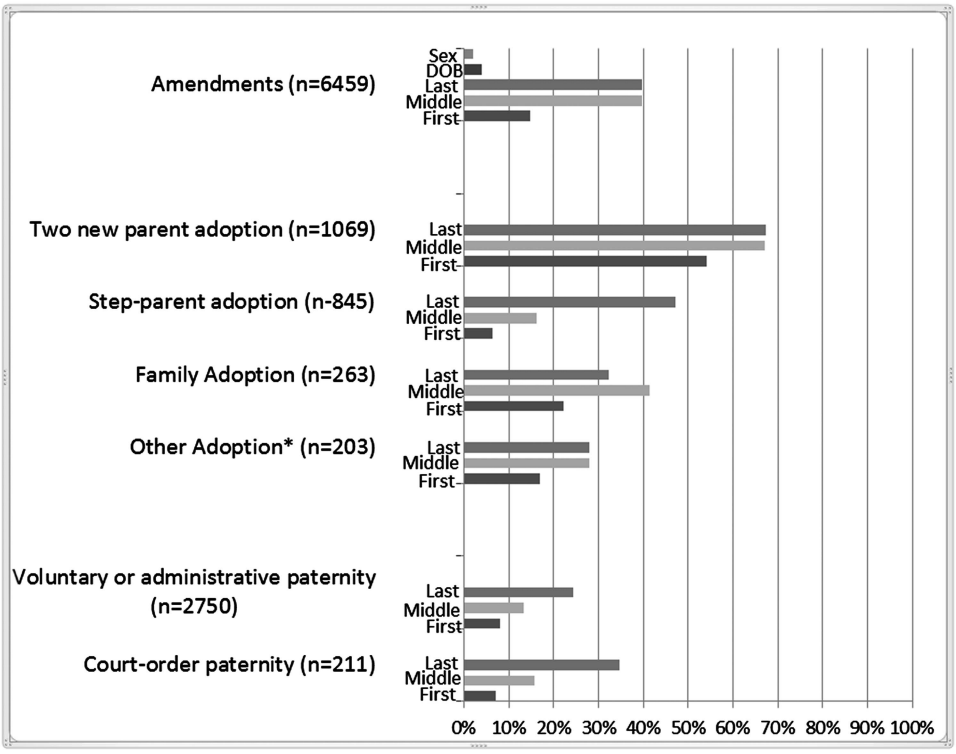

Figure 3 Frequency of changes to identity information, by field changed and type of change event for changes to a Utah birth certificate in 2010 for births in any prior year (n=2 589 265 births between 1905 and 2012). DOB, date of birth. Last, middle, first indicate last, middle, and first names.

so the findings should be similar in other states that have also adopted the vital statistics procedures defined in the law. Second, the data used for this study were limited to years after 2000 because of changes to Utah's vital statistics systems. Even so, the rates are likely valid due to the high number of records analyzed, and recent patterns are more relevant than previous patterns for addressing current identity management issues.

From a practical standpoint, these findings can influence practice in two ways. First, practitioners of data linkage in public health who currently use birth certificate data, such as immunization registries, can use knowledge of the age-dependent quality of birth identifiers to adjust blocking strategies and weight calculations. Second, birth certificate data are a potentially invaluable resource for informing identity resolution in healthcare and health information exchange settings, particularly in situations that involve identity changes due to adoptions. As societal trends such as gestational surrogacy and same-sex marriage may further contribute to fluctuations in patterns of identity change, more research will be needed to develop policies and technologies so that birth certificate information may be used to inform identity resolution while protecting the sensitive information of people identified.

CONCLUSIONS

Birth certificate identities change over time, particularly in the first years of life, but also patterns of change fluctuate temporally due to societal factors. The fact that changes to a birth certificate are overwhelmingly tied to changes in a person's real identity enhances the value of birth certificates for identity resolution in healthcare and public health information systems. Currently, system users struggle to link records that represent distinct snapshots of identity over time. Understanding the timing, frequency, and scope of these changes is an important first step in incorporating birth registries into data linkage strategies in healthcare and public health.

Contributors JD obtained the data, performed data analysis, and prepared the initial manuscript. SPN, SC, KE, ST, and CS contributed to study design, reviewed data analysis, revised the manuscript, and approved the final version.

Competing interests None.

Ethics approval This study was approved by the University of Utah's Institutional Review Board.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1 HIMSS. Recommendation to Congress 2012 [cited 2013 10-11-2013]. http://www.himss.org/files/HIMSSorg/policy/d/2012_Ask1_PatientDataMatchingStrategy.pdf
2 Fellegi IP, Sunter AB. A theory for record linkage. *J Am Stat Assoc* 1969;64:1183–210.
3 Newcombe HB. *Handbook of record linkage: methods for health and statistical studies, administration, and business*. Oxford University Press, 1988.
4 Newcombe HB, Kennedy JM. Record linkage: making maximum use of the discriminating power of identifying information. *Commun ACM* 1962;5:563–6.
5 Winkler WE, ed. *Overview of record linkage and current research directions*. Bureau of the Census, 2006: Citeseer.
6 Acheson ED. Medical record linkage. Medical record linkage. London: Oxford University Press, 1967.
7 DuVall SL, Fraser AM, Kerber RA, *et al*. The impact of a growing minority population on identification of duplicate records in an enterprise data warehouse. *Stud Health Technol Inform* 2010;160(Pt 2):1122.
8 DuVall SL, Kerber RA, Thomas A. Extending the Fellegi–Sunter probabilistic record linkage method for approximate field comparators. *J Biomed Inform* 2010;43:24–30.
9 Sariyar M, Borg A, Pommerening K. Controlling false match rates in record linkage using extreme value theory. *J Biomed Inform* 2011;44:648–54.
10 Hillestad RJ, Bigelow JH, Chaudhry B, *et al*. *Identity crisis: an examination of the costs and benefits of a unique patient identifier for the US health care system*. RAND, 2008.
11 Bittle MJ, Charache P, Wassilchalk DM. Performance improvement: registration-associated patient misidentification in an academic medical center: causes and corrections. *Jt Comm J Qual Patient Saf* 2007;33:25–33.
12 Aung E, Rao C, Walker S. Teaching cause-of-death certification: lessons from international experience. *Postgrad Med J* 2010;86:143–52.
13 Mills ME. Linkage of patient records to support continuity of care: Issues and future directions. *Stud Health Technol Inform* 2006;122:320–4.
14 Menachemi N, Collum TH. Benefits and drawbacks of electronic health record systems. *Risk Manag Healthcare Policy* 2011;4:47.

15 Maro JC, Platt R, Holmes JH, *et al*. Design of a national distributed health data network. *Ann Intern Med* 2009;151:341–4.
16 Houston J. UOoVRa. Utah's Vital Statistics: Births and Deaths, Updated 2011 2012 [cited 2013 09/28/2013]. http://health.utah.gov/vitalrecords/2011bx_UpdatedFinal_03292013.pdf
17 Duncan J, Xu W, Narus SP, *et al*. A focus area maturity model for a statewide master person index. *Online J Public Health Inform* 2013;5:210.
18 Durand A. Three tiers of identity. *Digit Identity World* 2002:19–69.
19 Deutsch MB, Green J, Keatley J, *et al*. Electronic medical records and the transgender patient: recommendations from the World Professional Association for Transgender Health EMR Working Group. *J Am Med Inform Assoc* 2013;20:700–3.
20 Long WS. Social Security numbers issued: a 20-year review. *Soc Sec Bull* 1993;56:83.
21 Arellano MG, Weber GI. Issues in identification and linkage of patient records across an integrated delivery system. *J Healthc Inf Manag* 1998;12:43–52.
22 McClellan MA, ed. Duplicate medical records: a survey of twin cities healthcare organizations. *AMIA Annual Symposium Proceedings*; American Medical Informatics Association, 2009.
23 SAS Software 2014 [cited 2014 01-27-2014]. http://www.sas.com/en_us/home.html
24 Herman E. *The adoption history project: Ellen Herman*. Department of History, University of Oregon, 2003.
25 Utah District Courts. Utah State Courts Statistics 2014 [cited 2014 02/24/2014]. https://www.utcourts.gov/stats/
26 US Department of Health and Human Services. The Adoption and Foster Care Analysis and Reporting System (AFCARS) Report. U.S. Department of Health and Human Services, 2013 [cited 2014 02-03-2014]. http://www.acf.hhs.gov/sites/default/files/cb/afcarsreport20.pdf
27 Solomon-Fears C. *Paternity establishment: child support and beyond*. Nova Publishers. 2003.
28 Grannis SJ, Overhage JM, Hui S, McDonald CJ, eds. Analysis of a probabilistic record linkage technique without human review. *AMIA Annual Symposium Proceedings*. American Medical Informatics Association, 2003.
29 Tromp M, Ravelli AC, Bonsel GJ, *et al*. Results from simulated data sets: probabilistic record linkage outperforms deterministic record linkage. *J Clin Epidemiol* 2011;64:565–72.
30 Integrating the Healthcare Enterprise (IHE). *IHE quality, research, and public health technical framework supplement: maternal child health – fetal death reporting*. IHE International, Inc, 2013.

Birth of identity: understanding changes to birth certificates and their value for identity resolution

Jeffrey Duncan, Scott P Narus, Stephen Clyde, et al.

J Am Med Inform Assoc published online July 30, 2014
doi: 10.1136/amiajnl-2014-002774