Automated Patient Linking for Electronic Health Record and Child Welfare Databases

Judith W. Dexheimer, PhD[1,2,3], **Sarah J. Beal, PhD**[3,4], **Parth Divekar, MIS**[2], **Eric S. Hall, PhD**[2,3,5], **Vikash Patel**[6], **Mary V. Greiner, MD MS**[3,6]

[1]Division of Emergency Medicine, Cincinnati Children's Hospital Medical Center, Cincinnati, OH

[2]Department of Biomedical Informatics, University of Cincinnati College of Medicine, Cincinnati, OH

[3]Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH

[4]Division of Behavioral Medicine and Clinical Psychology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH

[5]Perinatal Institute, Cincinnati Children's Hospital Medical Center, Cincinnati, OH

[6]Division of General and Community Pediatrics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH

Abstract

There are 427,000 children in protective custody in the United States. A lack of integration between the child welfare data system and electronic health record systems complicates the communication of critical health history details to caregivers. We created and evaluated automated ten custom algorithms linking these data. Deterministic matching was performed using combinations of first and last name, date of birth, and gender. If unmatched, a non-deterministic algorithm allowed for punctuation differences and letter transpositions. Of the children linked deterministically, 91.3% were linked. Of the ones undergoing non-deterministic matching, 71.3% were linked. Sharing integrated data is the first step in systematically improving health outcomes for children in protective custody. This approach represents an automatable and scalable solution that could help merge data from two disparate sources.

Keywords

Correspondence should be addressed to Judith W Dexheimer, PhD, Division of Emergency Medicine, Cincinnati Children's Hospital Medical Center, MLC 2008, 3333 Burnet Avenue, Cincinnati, OH 45229-3039, Telephone: 513-803-2962, Judith.Dexheimer@cchmc.org.

Conflicts of Interest: The authors have no conflicts of interest, financial, or other to declare.

Problem

In the United States, 427,000 youth (US Department of Health & Human Services, 2016) are in protective custody (e.g., foster care, kinship care). Children in protective custody are significantly more likely to suffer from emotional, behavioral, medical, and dental problems that require health care services than children in the general population. Caseworkers have insufficient health information for these youth, including missing acute and chronic diagnoses, medication lists, allergies, and immunization records. (Greiner, Ross, Brown, Beal, & Sherman, 2015; Risley-Curtiss & Kronenfeld, 2001) Missing information contributes to poor chronic condition management (Ahrens et al., 2011; Christian & Schwarz, 2010) and hinders prevention. (Ahrens et al., 2010) While health information is stored by medical providers in the electronic health record (EHR), child welfare maintains separate administrative records (e.g., State Automated Child Welfare Information System; SACWIS). EHR data and child welfare data are often shared manually, where a health care provider contacts the county to access child welfare information, and the caseworker contacts the health care provider or health care system to access health records however this process is inconsistent and rarely timely. SACWIS and the EHR are not electronically integrated, which complicates dissemination of health details to caregivers and caseworkers. Merging and displaying SACWIS and EHR data would support health information exchange, improve health care delivery, and promote continuity during changes in placement for children in protective custody. However, no unique identifier between data systems exists, making merging and sharing the data in an automated fashion technically difficult. In the absence of a shared unique identifier, the objective of this study was to evaluate the feasibility of using automated deterministic and non-deterministic algorithms to link records. While this application is specific to child welfare and EHR settings, similar processes could be used to link EHR data with other public sector datasets, including other state-based services (e.g., housing authority), employment data, or public health services.

Approach

Setting.

We used EHR records are from a large pediatric medical center in the Midwest. The medical center provides mandated healthcare visits for all youth in protective custody in the county. The hospital implemented a commercial, enterprise EHR in 2009. The county welfare system began using SACWIS in 2008. Child welfare records from SACWIS were provided for children in foster care in March 2017 and May 2017. The study was approved by the hospital's institutional review board and received a waiver of consent.

Study Design.

Ten custom algorithms were developed to link records for youth in protective custody represented in SACWIS and the EHR. Four algorithms were deterministic and six algorithms were non-deterministic. Once an EHR record was linked to a SACWIS record, the EHR record were removed from the candidate pool of EHR records for other SACWIS records, optimizing linking of one SACWIS record to one EHR record. To allow for the

potential of duplicate EHR records, SACWIS records were allowed to link to more than one EHR record.

Deterministic matching algorithms required complete and exact matches of a) the first and last name, date of birth (DOB), and gender; or b) an alias first and last name, DOB, and gender. (Hall et al., 2014) Punctuation was eliminated to use true alphabets and to overcome any errors in entry of punctuation. Aliases listed in the EHR are provided at the discretion of registration personnel, including when patients have a name change (e.g., change from baby girl to first name, change from trauma to name, changes to spelling of a first or last name). Deterministic matches were also allowed if all of the patient's first name (or alias first name) matched the SACWIS record name and the first part of the last name matched. DOB and gender were required to match exactly for all deterministically-linked records. A random ten percent of deterministically-linked records were validated via manual chart review in both SACWIS and the EHR, where the reviewers were blinded to which records were linked or whether a match was found.

Non-deterministic algorithms allowed for minor letter transpositions in text including first and last name, parent names, aliases, DOB, and gender. Punctuation was again eliminated to use true alphabets. Non-deterministic algorithms allowed for DOB or gender mismatches if the patient's first name and last name (or alias first and last name) matched exactly. All non-deterministic matches were reviewed for accuracy. A manual chart review was conducted for patients without a match identified.

Outcomes

There were 2,209 youth in protective custody in March. Of those, 2,017 (91.3%) were linked deterministically. In May there were 2,270 youth in protective custody, and 2,053 (90.4%) were linked deterministically. Blinded chart review revealed 100% accuracy for deterministic linking in both sets.

Of 192 (March) and 217 (May) records undergoing non-deterministic linking, due to lack of deterministic match, 137 (71.3%) and 153 (70.5%) were linked after manual review of proposed matches. Figure 1 depicts algorithm linking results and errors. Six incorrect non-deterministic linked records were found, five were DOB mismatches and one was an incorrect spelling.

Fifty-two youth (March; 2.4%) and sixty-two youth (May; 3.0%) had no algorithm link but were found in a manual EHR chart review. Youth that were missed in the automated matching but found in manual review included primarily those with differing last names and dates of birth between the SACWIS and EHR databases. The most common reason for youth having no algorithm link was having a last name entirely different or hyphenated in their EHR than in SACWIS. This fell out of the algorithm's threshold for misspellings and therefore caused the lack of a algorithmic link. Across both months, five youth had no data in the EHR. These data were used to calculate the true positive rate (i.e., sensitivity) and true negative rate (i.e., specificity) of non-deterministic algorithms, using manual chart review as

the gold standard. Non-deterministic algorithms had a sensitivity of 0.98 (March) and 0.97 (May) and specificity of 0.95 (March) and 0.40 (May).

Deterministic and non-deterministic linking took less than 25 minutes for all algorithms, indicating that producing deterministic linked records and candidate non-deterministic linked records on a more frequent (e.g., daily) basis is feasible once an automated process is in place.

Next Steps

This study identified an automated and scalable approach for linking patient records between medical and child welfare systems, with demonstrated reliability. This first step is essential to support an integrated, computerized display system of two disparate data sources. Our findings demonstrate that linking patient records between medical and child welfare systems is feasible and linking algorithms can be executed in a short period of time. Non-deterministic linked records (9% of records) require additional verification by an administrator to ensure valid record-linking; however, this represents a reduced burden on healthcare providers already attempting to link across systems manually. Finding multiple and merging records for one child helps to improve the quality of EHR demographics data by identifying the duplicate records. These duplicates in the EHR could otherwise cause a loss of important health data for the youth.

By linking child welfare and EHR data, an integrated data platform can support healthcare providers by ensuring that they receive essential information for care delivery, including custody information pertinent for medical consent and social history which is beneficial for developing treatment plans. Linked data also allow medical histories, assessments, and recommendations to quickly reach children's services workers, ensuring that children entering or living in foster care maintain needed healthcare services. This healthcare data can also be beneficial for setting up services with other systems, i.e. enrollment in school EHRs are used by 80% of health care providers (Lehmann, O'Connor, Shorte, & Johnson, 2014), and SACWIS is operational in 34 (68%) states (Children's Bureau, 2015) making our method applicable to a large number of institutions. Additionally, these algorithms can be adapted and applied to other public and private sector data systems.

This novel demonstration project evaluated the feasibility of linking data housed by multiple systems in an automated fashion. Linking cross-system data is a critical first step in creating a platform that shares complete and up-to-date health information with caseworkers, health care providers, and other stakeholders. By merging child welfare data with the EHR, health care providers will have information at their fingertips which is essential for delivery of health care, including custody information to know who can give permission to treat and who is authorized to receive health care information, including processing the large data sets to help improve foster care (Brindley, Heyes, & Booker, 2018). An accurate social, family, and maltreatment history will give health care providers the needed information to make decisions regarding health risks (i.e. sexually transmitted infections) and needs for trauma informed mental health services. Merged datasets will also allow a child's medical history, health assessments, and provider recommendations to quickly reach children's services

workers, ensuring that children entering or living in protective custody maintain needed health care services even when placements are disrupted. Allergies, medications, chronic disease management, and upcoming health appointment information can stay with a child coming into custody and changing placements while in custody. This can eliminate a gap in health care delivery and assist with improving preventive health care and chronic health condition management. Complete, integrated health records are the first step in understanding the medical and social complexity for children in foster care, allowing opportunities for the healthcare system to work with community partners to provide comprehensive and continuous care to improve the health status of this vulnerable population. A similar model can be replicated with other populations, to improve the health care safety net for vulnerable children and adults who interact with social services in the United States.

Acknowledgements:

Abbreviations:

EHR:	Electronic Health Record
SACWIS:	State Automated Child Welfare Information System
DOB:	Date of Birth

REFERENCES

Ahrens KR, DuBois DL, Garrison M, Spencer R, Richardson LP, & Lozano P (2011). Qualitative exploration of relationships with important non-parental adults in the lives of youth in foster care. Children and youth services review, 33(6), 1012–1023. [PubMed: 22661797]

Ahrens KR, Richardson LP, Courtney ME, McCarty C, Simoni J, & Katon W (2010). Laboratory-diagnosed sexually transmitted infections in former foster youth compared with peers. *Pediatrics*, peds. 2009–2424.

Brindley M, Heyes JP, & Booker D (2018). Can Machine Learning Create an Advocate for Foster Youth? Journal of Technology in Human Services, 36(1), 31–36.

Children's Bureau. (2015, 22 June 2018). SACWIS Status. Retrieved from https://www.acf.hhs.gov/cb/resource/sacwis-status

Christian CW, & Schwarz DF (2010). Child maltreatment and the transition to adult-based medical and mental health care. *Pediatrics*, peds. 2010–2297.

Greiner MV, Ross J, Brown CM, Beal SJ, & Sherman SN (2015). Foster caregivers' perspectives on the medical challenges of children placed in their care: implications for pediatricians caring for children in foster care. Clinical pediatrics, 54(9), 853–861. [PubMed: 25561697]

Hall ES, Goyal NK, Ammerman RT, Miller MM, Jones DE, Short JA, & Van Ginkel JB (2014). Development of a linked perinatal data resource from state administrative and community-based program data. Maternal and child health journal, 18(1), 316–325. [PubMed: 23420307]

Lehmann CU, O'Connor KG, Shorte VA, & Johnson TD (2014). Use of electronic health record systems by office-based pediatricians. *Pediatrics*, peds. 2014-1115.

Risley-Curtiss C, & Kronenfeld JJ (2001). Health care policies for children in out-of-home care. Child Welfare, 80(3), 325. [PubMed: 11380045]

US Department of Health & Human Services. (2016). The AFCARS report: Preliminary FY 2015 estimates as of June 2016. No. 23. Washington, DC: Author Retrieved from https:// www.acf.hhs.gov/sites/default/files/cb/afcarsreport23.pdf
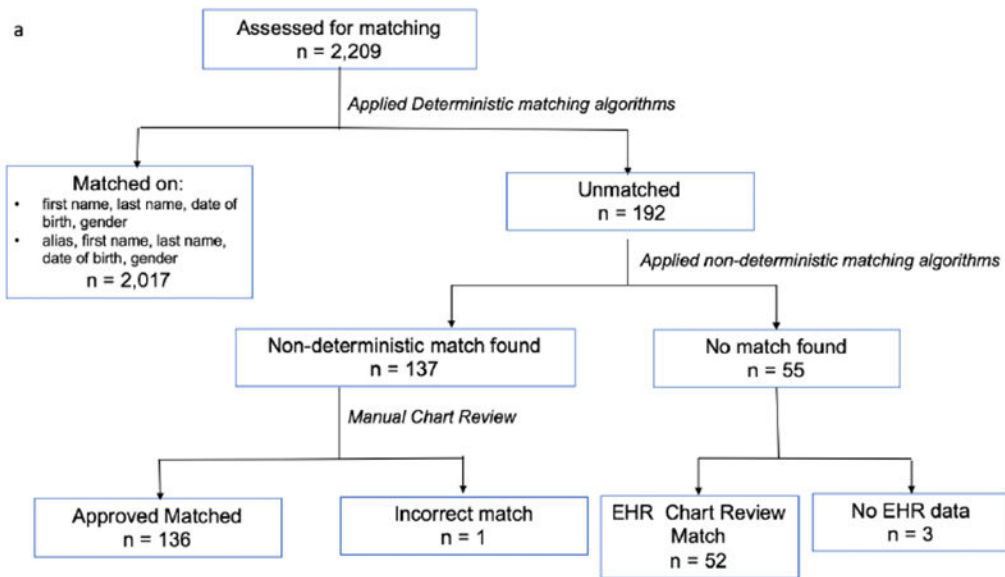
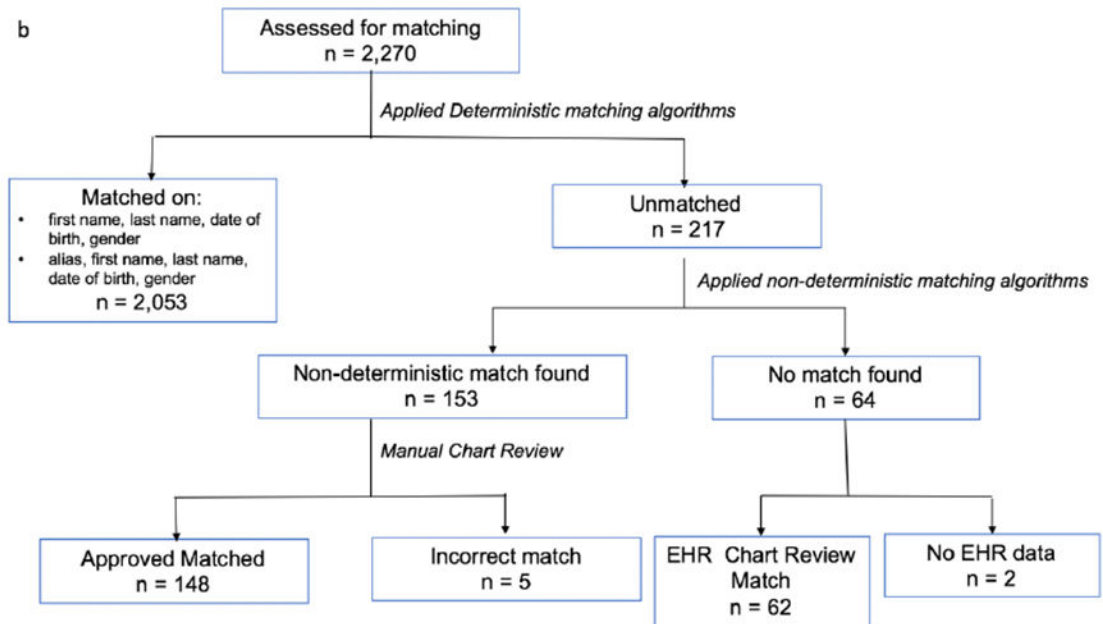## March 2017

May 2017

Figure 1.
Matching algorithm results for deterministic and non-deterministic algorithms for (a) March 2017 and (b) May 2017.