Reducing Duplicate Patient Creation Using a Probabilistic Matching Algorithm in an Open-access Community Data Sharing Environment

Sidney N. Thornton, Ph.D.1,2, Shannon K. Hood1

1Department of Medical Informatics, Intermountain Health Care, Salt Lake City, Utah
2Department of Medical Informatics, University of Utah, Salt Lake City, Utah

In an open-access community data sharing environment, Intermountain Health Care (IHC) is managing the creation of duplicate patient records through a probabilistic matching algorithm that allows the threshold limits for the returned set to be dynamically assigned to the source system. For internal hospital systems, the rate of duplicate creation was cut 30% in the first 6 months. For IHC’s first community data sharing partner, the rate of duplicate creation has been maintained below the acceptable range for the internal Health Plans rate.

INTRODUCTION

Successful community data sharing within a longitudinal patient record is dependent on accurate patient identification. Without accurate patient identification portions of the longitudinal patient health record may be fragmented into isolated episodes (duplicates) or attached to the wrong patient (overlays). The incompleteness negates the full value of the community data sharing initiative.

Data sharing across IHC’s 21 hospitals, 94 ambulatory care settings, and Health Plans follows an open-access model for data sharing, where source information systems can add/update members in the Enterprise Master Person Index (EMPI)1. Using an algorithm to determine the probability of matching demographics within the EMPI either adds the data to an existing member or creates a new member. If the added member demographics are sufficiently close to matching with an existing member, the case is manually resolved. Recently, IHC began sharing data with a Health Care Organization (HCO) outside the IHC network using the same open-access model.

If unmanaged the problem of the open-access model is the uncontrolled creation of duplicate records. Our approach to managing duplicate patient creation is the integration of a third-party matching algorithm2 into IHC’s proprietary patient lookup services, and modify the lookup services to provide dynamically assigned, high and low thresholds for the returned set of potential patient matches. The relative trust of the data source determines the matching thresholds. A relatively low-trusted system would be assigned high matching thresholds, creating more duplicates but reducing the risk of overlaying the new information on the incorrect patient. In all cases, the duplicates are manually resolved at $60 per case.

METHODS

To evaluate the impact of the new matching algorithm, a quasi-experimental design was used to compare the rate of duplicate records per new member creation. The implementation of the matching algorithm was made in two steps: 1) enterprise-wide including all hospital registration systems (11K adds/month), and 2) HCO scheduling records (400 adds/month). Matched controls were approximated using 1) the ambulatory additions and 2) Health Plans member load. In both controls the algorithm did not impact registration or enrollment workflow. The ambulatory duplicate rate has historically tracked the hospital system. The Health Plans load, like the HCO load is asynchronous to the EMPI with approximately the same trust level.

RESULTS AND DISCUSSION

For the 6 months following the enterprise-wide intervention, the rate of duplicate records per new member creation for the internal hospital systems has decreased more than 30% after correcting for a 13% decrease in the control ambulatory clinic rate. At first the algorithm was unacceptably slow, but was later improved through full EMPI caching. For the 12 weeks since accepting HCO scheduling records, the HCO duplicate rate has been 17% below the rate for the Health Plans member load. The HCO is very highly aligned with IHC’s pediatric hospital and may not be fully representative of the data sharing community.

CONCLUSIONS

The probabilistic matching algorithm facilitates the management of duplicate patient creation and positions IHC to further tune and refine the EMPI processes. Duplicates, however, are still generated, even if at lower rates. As more and more HCO’s begin to share data with IHC, the cost of duplicate management will grow.

REFERENCES

2. QuadraMed Corporation, San Rafael, CA, USA