MEDINFO 2007
HISA Ltd Australia
© 2007 The authors. All rights reserved.

Understanding the Profile of Errors That Cause Duplicate Entries in a Patient Registry

Scott L. DuVall[a], Janice Conrads[b], Alison Fraser[b], Geraldine Mineau[b,c]

[a] *Department of Biomedical Informatics, University of Utah, United States of America*
[b] *Pedigree and Population Resource, Huntsman Cancer Institute, University of Utah, United States of America*
[c] *Department of Oncological Sciences, University of Utah, United States of America*

Abstract

Duplicate entries in patient registries are a significant problem facing healthcare master patient indices and public health data sources. Patterns exist in the demographic information of these records that can be used to better understand what causes duplicate entries to appear. The University of Utah contracts the Pedigree and Population Resource group at Huntsman Cancer Institute to link demographic records from its Enterprise Data Warehouse (EDW) to the Utah Population Database (UPDB). An additional benefit of this linkage is the discovery of duplicate records in the EDW. From one million EDW records, 76,922 have been identified as duplicate. A profiler was built to extract the patterns of errors from these duplicate records. Frequencies of error types found in the EDW were compared with those in existing literature.

Keywords:

record linkage, patient record management, patient registries

Introduction

Duplicate records are detrimental to the cost-effective and efficient delivery of health care.[1] Manually identifying and resolving duplicates can cost $60 per case.[2] Patterns have been found in the types of errors that occur in patient registries, suggesting that undetected duplicate records may be similar to those already identified.[3,4] The purpose of this study was to compare the frequency of error types in the University of Utah Enterprise Data Warehouse (EDW) with existing literature.

Methods

Patterns of errors in existing literature were extracted. State machine[5] templates were created for each error type as Java classes. As a potential duplicate record pair was examined, an instance of each state machine was created with one record and tested with the other record. A test resulting in an "accept" state resulted in the pair being classified as that error type.

Results

As shown in Table 1, the EDW records show some differences from published literature in patient name types.

	EDW	Friedman[4]
Extra names and titles	34.3%	36.9%
Nicknames, spelling variations	21.8%	13.9%
One letter substitutions	13.6%	13.7%
One letter added or deleted	7.6%	12.9%
Punctuation or spaces	1.9%	11.8%
Different last names for females	12.9%	7.8%
Permuted parts of names	3.2%	1.4%
Different first names	2.8%	1.4%
One letter transposed	1.9%	0.8%

Table 1 – Error types of patient names

Similar frequency discrepancies were found in error types of other demographic fields, including Social Security Number.

Discussion

Despite differences found in the frequency of error types that occur in the EDW data set from published literature, it is clear that patterns of error types do occur in duplicate records. Understanding which errors are likely to occur in a dataset can aid in the detection of duplicate records.

Acknowledgments

Research supported by a training grant from the National Library of Medicine and Robert Wood Johnson Foundation.

References

[1] Mays S, Swetnich D, Gorken L. Toward a Unique Patient Identifier. Health Manag Technol. 2002 Mar;23(3):42-4.
[2] Thornton SN, Hood SK. Reducing Duplicate Patient Creation Using a Probabilistic Matching Algorithm in an Open-access Community Data Sharing Environment. Proc AMIA Symp 2005:1135.
[3] Friedman C, Sideli R. Tolerating spelling errors during patient validation. Comput Biomed Res 1992;25:486-509.
[4] Grannis SJ, Overhage JM, McDonald CJ. Analysis of Identifier Performance using a Deterministic Linkage Algorithm. Proc AMIA Symp 2002:305-9.

[5] Finite state machine. In *Wikipedia, The Free Encyclopedia*. Retrieved December 4, 2006, from http://en.wikipedia.org/w/index.php?title=Finite_state_machine&oldid=91809582