

Turning Junk Data into Value

Using 9-digit Mixed Identifiers to Enhance Linkage Results for Utah Controlled Substance Database

Yukiko Yoneoka¹, MS, Wu Xu¹, PhD, Zhiwei Liu¹, MS, Brian Sauer², PhD, Robert Rolfs¹, MD
¹Utah Department of Health and ²University of Utah, Salt Lake City, Utah

Background

Public Health Problem

In the past seven years, the state of Utah has experienced increasing numbers of unintentional deaths due to prescription pain medication overdoses.

Controlled Substance Database (CSDB)

Pharmacies submit controlled substance dispensing records with patient information to the CSDB without strict data standards. As a result, considerable variability is found in required patient information fields.

Challenge: Customer ID

Since standardization is not enforced for the customer ID field, it contains different types of IDs such as pharmacy specific ID, SSN, drivers license number, passport number, phone number, names and other text. This inconsistency makes it challenging for researchers to accurately construct patient-centered prescription records across pharmacy records.

Objective

The purpose of this study was to salvage and utilize all 9-digit ID numbers from the ID variable. We then examined how adding the 9-digit mixed ID as a linkage variable affected linkage results.

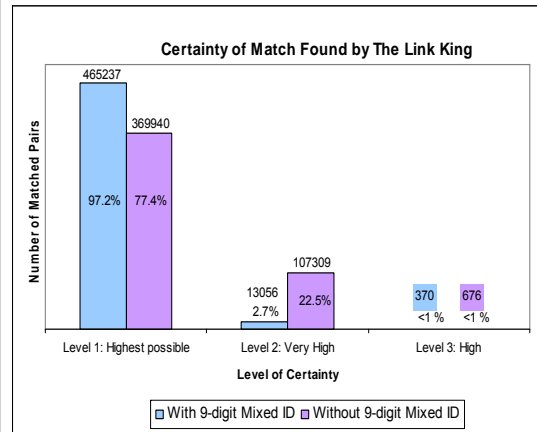
Acknowledgements

This study is supported by CDC Grant No. P01 CD000284-03, P.I. Matthew Samore, Utah Research Center for Excellence in Public Health Informatics. Many thanks to Nancy McConnell for her valuable suggestions.

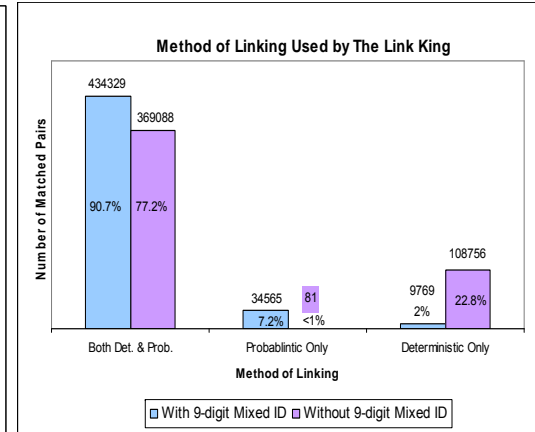
Contact Yukiko Yoneoka
y.yoneoka@utah.gov

Result

How Using 9-digit Mixed ID Enhanced Linkage Results



About **20% more** matched pairs were found with highest possible level of certainty by using 9-digit mixed IDs.



About **14% more** matches were found by both deterministic and probabilistic linking methods and **7% more** by probabilistic method by using 9-digit mixed IDs. Without, it heavily depended on deterministic method only.

Major blocking criteria where match was found by The Link King	With 9-digit Mixed ID	Without 9-digit Mixed ID
9-digit Mixed ID only	60,031 (12.5%)	N/A
Last Name & DOB	407,105 (85%)	464,732 (97.2%)
First, Middle and Last Names only	7,879 (1.7%)	9,613 (2%)
First Name & DOB	1,941 (0.4%)	1,934 (0.4%)
First and Last Names & Birth Year	1,028 (0.2%)	990 (0.2%)
First and Last Names & Birth Month	575 (0.1%)	565 (0.1%)

In blocking process, about **13% of all matches** were found by 9-digit mixed ID match alone.

Conclusion

Retrieving 9-digit IDs from a mix of data collected in a customer ID field to create a mixed ID field as a linkage variable would be a worthwhile practice, considering the enhanced quality of the linkage results.

Methods

Data

CSDB 2006-2007 contained 9,342,994 prescription records. Of those, 6,212 records (0.07%) from veterinarian or other non medical prescribers were excluded from analysis. A total of 9,336,782 prescription records were used.

The 9-digit Mixed ID

The 9-digit mixed ID was populated with customer ID content, using:

- 1) Length of 9 after stripping off attached characters (e.g., SSN, SS#, UTDL, UDL);
 - 2) Drivers license number validation by an algorithm that assigns Utah 9-digit drivers license numbers (9, 8, 7, 6, 5, 4, 3, 2, 1) • (d1, d2, d3, d4, d5, d6, d7, d8, d9) = 0 mod 10; and,
 - 3) Valid range of first 3 digits of SSN.
- About 35.5% (3,313,731) of all prescription records carried some type of 9-digit ID. The distribution of content of the 9-digit mixed ID is indicated in the table below.

Break down of 9-digit mixed ID	Number (%)
SSN	1,108,388 (33.4)
UTDL	1,070,071 (32.3)
Other 9-digit ID	1,135,272 (34.3)
Total	3,313,731 (100.0)

Linkage

The **Link King® v.7**, a free SAS based linkage software, was used on a platform of SAS v.9.1.3. Variables used for linkage were: First, middle and last names; date of birth (DOB); gender; zip code; and, 9-digit mixed ID. The data was linked first with the 9-digit mixed ID then without. The results were compared based on The Link King's result reports.

