

Comparison of Record Linkage Software
for De-duplicating Patient Identities
in California's Prescription Drug Monitoring Program

Tina Farales
California Department of Justice

Susan Stewart
University of California, Davis

ONC Symposium on Patient Matching for PDMPs
September 6, 2019



Outline

1. Background
2. Methods
3. Results
4. Discussion
5. Limitations
6. Conclusions

Background

- Accurate linkage of PDMP records is essential
 - To identify potentially risky prescribing and dispensing patterns and outlier patient behavior
 - To monitor potentially risky population trends

- Patient entity resolution is performed in CURES to provide the following features:
 - Patient safety alerts to prescribers (new alerts produced daily)
 - De-identified data for researchers
- CURES receives approximately 155K new prescription records daily.
- With this new data, the analytics engine must reconcile patient, prescriber, and dispenser entities across the 1TB database every night.

- Once the data is de-duplicated nightly, the analytics engine identifies the resolved persons' current prescriptions based on date filled and number of days supply.
- The resolved persons' current prescription medicinal therapy levels are calculated and compared against pre-established thresholds.
- Therapy levels exceeding those thresholds trigger Patient Safety Alerts to current prescribers.

- The de-duplicated data also contributes to the quarterly and annual systematic production of 58 California county and one statewide de-identified data sets for use by public health officers and researchers.
- This data enables counties to
 - calculate current rates of prescriptions,
 - examine variations within the state, and
 - track the impact of safe prescribing initiatives.

- CURES is a “home grown” PDMP system. This means that the CA PDMP has full access and visibility to how the CURES system operates and functions. After employing a custom-built entity resolution methodology, the CA PDMP wanted to have its de-duplication approach evaluated.
- One of the purposes of the evaluation is to help inform the CA PDMP on areas for strength and weakness. The CA PDMP plans to pursue implementing improvements in this challenging area.

Goal

- Compare record linkage programs with respect to
 - Accuracy in de-duplicating a subset of patient identities
 - Identification of excessive opioid use and outlier behavior
- Challenges
 - No unique patient identifier
 - Variation in identity fields for an individual
 - Hundreds of millions of records

Methods

Compare Record Linkage Programs

- CURES 2.0 custom-built program
 - SAS application
- The Link King: <http://www.the-link-king.com/index.html>
 - SAS application
- Link Plus: <http://www.cdc.gov/cancer/npcr/tools/registryplus/lp.htm>
 - Microsoft Windows stand-alone application
- LinkSolv: <http://www.strategicmatching.com/products.html>
 - Microsoft Access application

Approach

- Start with exact matching of prescription record identifiers
 - Decreases size to ~60 million records
- Link within smaller geographic areas
 - Test dataset: patient identities for prescriptions filled in 2013 in 2 zip3s
 - 1 in Northern California, 1 in Southern California
 - ~500,000 records

Entity resolution

- 1) Compare pairs of records to determine whether they match
- 2) Assign a score to indicate match quality
- 3) Determine which records correspond to the same entity based on match results

Fields Available to Match

- First name
- Last name
- Date of birth
- Gender
- Address
 - Street address
 - City
 - Zip code (5 digits)

Manual Review

- Matches identified by one or more of the programs at any level of certainty were included in the full dataset of paired records
- Paired records were stratified by level of certainty
 - From high to low confidence in a match
- 5 reviewers inspected a stratified random sample of 720 paired records
 - Blinded to software certainty ratings
 - “Truth” determined by majority opinion

Statistical Analysis

- Assessed accuracy of software using stratified sample weighted to full set of paired records
 - **Sensitivity**: proportion of true matches identified by the program (aka recall)
 - **Positive predictive value**: proportion of identified matches that are true matches (aka precision)
- Assessed relative importance of specific identity fields in distinguishing matches from non-matches by each program
- Computed PDMP patient alerts and CDC metrics for the patient entities identified by each program

Results

Agreement between Record Linkage Software and Manual Review of a Sample of Pairs of Identity Records from California's PDMP

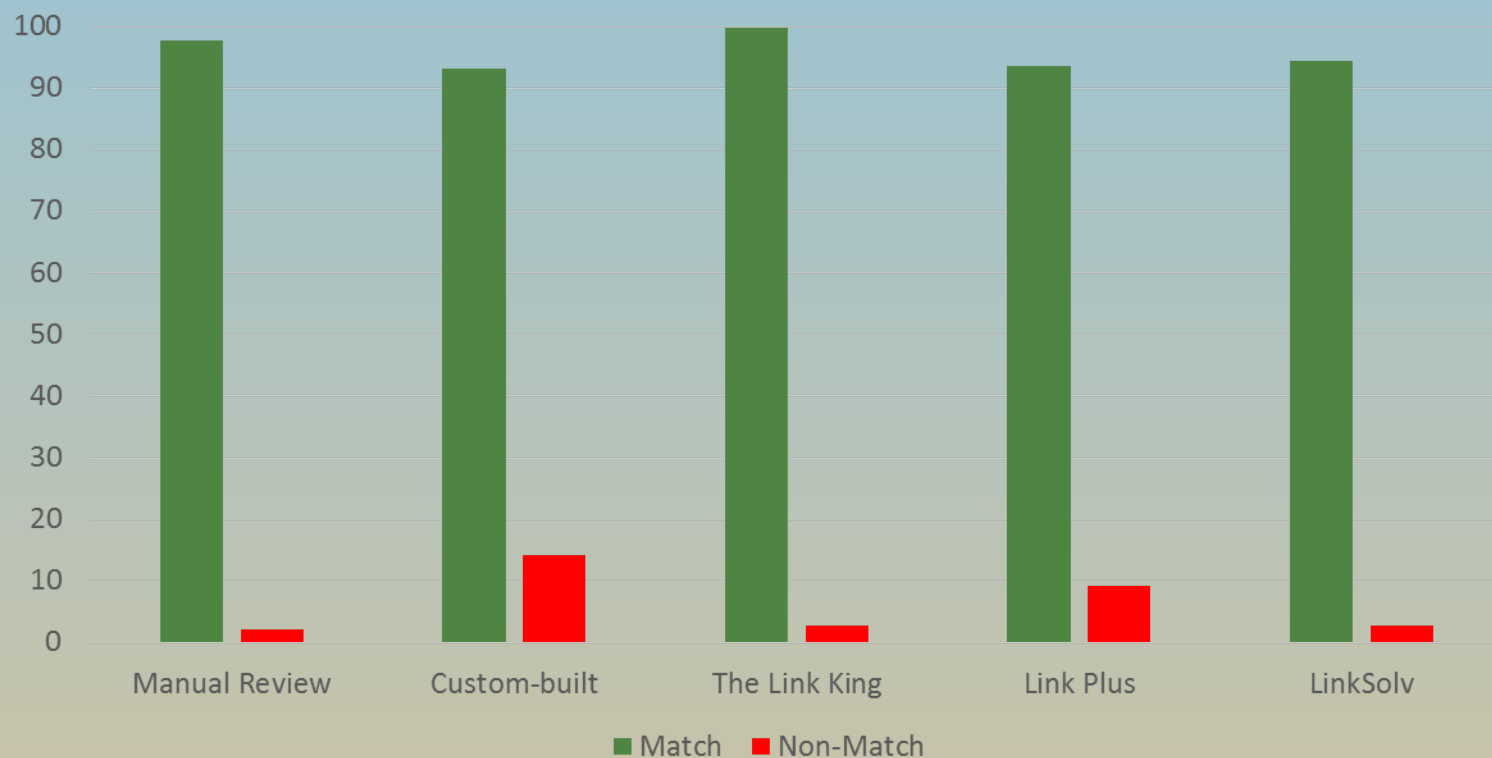
Software	PPV (%)		Sensitivity (%)	
	Est.	95% CI	Est.	95% CI
Custom-built	94.9	94.1-95.7	73.0	72.0-74.1
The Link King	97.9	96.7-99.2	94.8	93.8-95.8
Link Plus	93.5	92.3-94.7	83.6	81.5-85.8
LinkSolv	93.1	91.7-94.5	95.3	94.8-95.8

Note: CI=confidence interval; PPV=positive predictive value

Match by manual review: at least 3 of 5 reviewers rated pair as probably or definitely the same person

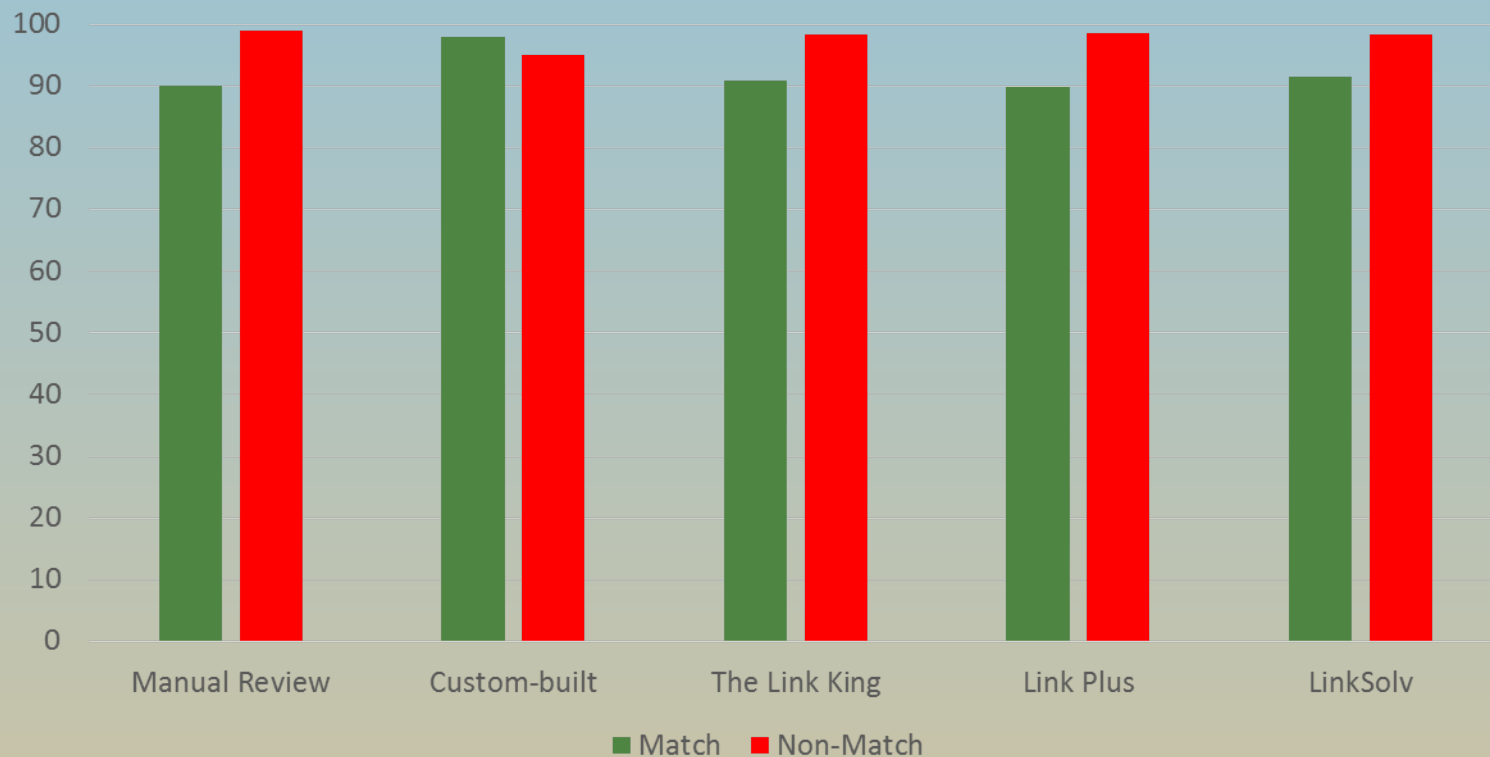
Importance of Date of Birth

Percent of Paired Identities with the Same DOB
by Match Status



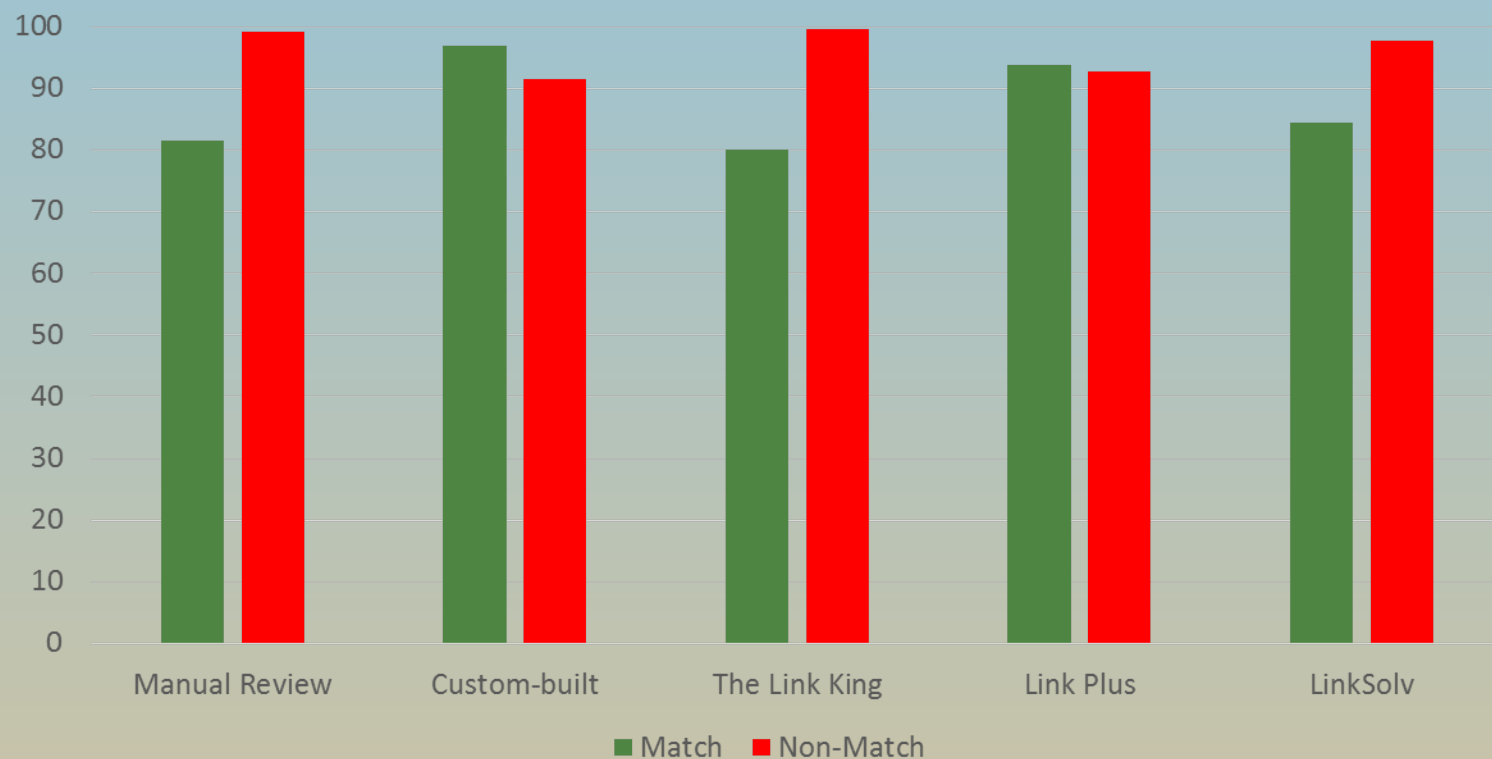
Importance of Last Name

Percent of Paired Identities with the Same Last Name
by Match Status



Importance of Zip Code

Percent of Paired Identities with the Same Zip Code
by Match Status



Number of Patient Alerts

PDMP Alert Scenario	Software	Patient Entities	
		n	%diff.
Currently prescribed >90 MMEs/day	Custom-built	3426	0
	The Link King	3434	0.2
	Link Plus	3444	0.5
	LinkSolv	3435	0.3
Obtained prescriptions from ≥6 prescribers or ≥6 pharmacies in last 6 months	Custom-built	1993	0
	The Link King	2211	10.9
	Link Plus	2524	26.6
	LinkSolv	2329	16.9
Currently prescribed opioids >90 consecutive days	Custom-built	3039	0
	The Link King	3138	3.3
	Link Plus	3097	1.9
	LinkSolv	3140	3.3
Currently prescribed both benzodiazepines and opioids	Custom-built	2923	0
	The Link King	2955	1.1
	Link Plus	2989	2.3
	LinkSolv	2976	1.8

CDC Metrics

CDC Metric	Software	Value per Quarter or 6-Month Period			
		Period 1	%diff.	Period 2	%diff.
Average dose of > 90 MMEs in quarter*	Custom-built	8.89	0	8.33	0
	The Link King	8.76	-1.5	8.22	-1.3
	Link Plus	8.91	0.2	8.33	0.0
	LinkSolv	8.78	-1.2	8.25	-1.0
Obtained prescriptions from ≥5 prescribers and ≥5 pharmacies in 6 months†	Custom-built	18.15	0	13.68	0
	The Link King	20.44	12.6	16.74	22.4
	Link Plus	25.16	38.6	20.34	48.7
	LinkSolv	22.39	23.4	18.25	33.4
Overlap of opioid prescriptions in quarter‡	Custom-built	16.70	0	17.53	0
	The Link King	17.14	2.6	18.04	2.9
	Link Plus	17.55	5.1	18.45	5.2
	LinkSolv	17.30	3.6	18.20	3.8
Overlap of benzodiazepine and opioid prescriptions in quarter‡	Custom-built	9.72	0	9.96	0
	The Link King	9.89	1.7	10.15	1.9
	Link Plus	10.12	4.1	10.38	4.2
	LinkSolv	9.97	2.6	10.24	2.8

*% of patients

†per 100,000 population

‡% of patient prescription days

Discussion

- All 4 record linkage programs were reasonably accurate in identifying matches and non-matches
 - Most accurate: the Link King and LinkSolv
 - Least accurate: custom-built program

Importance of Matching Fields

- Date of birth was very important to human reviewers, but less so to the custom-built program and Link Plus
- Agreement in last name was more important to the custom-built program than to human reviewers and the other 3 programs
 - Double last names and switched first & last names were less likely to be included in matches by the custom-built program
- Agreement in zip code was more important to the custom-built program and Link Plus than to the others

Patient Alerts and Metrics

- Effects of using specific software were greatest on the identification of outlier patients who obtained prescriptions from a large number of prescribers and/or pharmacies
 - Prescriptions from multiple prescribers and/or pharmacies are likely to result in multiple identity records, which must be linked

Limitations

- Small scope of evaluation
 - Half a million records from geographically separated areas
 - Used default settings where available
- Changes to current linkage methods in production would require further testing for feasibility and accuracy

Conclusions

- Certain publicly and commercially available record linkage programs linked identity records more accurately than a custom-built application
 - It is not necessary to build a record linkage system from the ground up
 - It is necessary to conduct a test of any proposed software with manual review of matches to ascertain their accuracy

Thank you!

- For further information, please contact
 - Tina Farales, CURES Manager
 - Phone: 916-210-3171
 - Email: Tina.Farales@doj.ca.gov
 - Susan Stewart, Professor of Biostatistics
 - Phone: 916-734-7217
 - Email: slstewart@ucdavis.edu