Matching patients across institutions without definitive patient identifiers

November 20, 2012

Jim Naessens, Stephanie Peterson, Ahmed Rahman, Matt Johnson, Diane Olson, Sue Visscher, Kyle Koenig

Agenda

- Purpose
- Literature Review
- Proposed Algorithms
- Results
- Summary
- Next Steps

Agenda

- **Purpose**
- Literature Review
- Proposed Algorithms
- Results
- Summary
- Next Steps

Purpose

Link 30 day readmissions and deaths for discharges between MN hospitals to:

- Determine the clinical effectiveness of current medical therapies with comprehensive “real life” databases
- Provide healthcare professionals with information that supports the highest possible quality of care.
- Linkage of data is a fundamental building block for a health information exchange.

Agenda

- Purpose
- **Literature Review**
- Proposed Algorithms
- Results
- Summary
- Next Steps

Literature Review: Record linkage

- The term “record linkage” was first used to describe the process of connecting two or more medical record documents by Halbert L. Dunn in 1946 [1] in the late 1950s.

Literature Review- Record Linkage, cont.

At a high level:

- **Deterministic matching**
  - All or none matching – based on a single unique identification number was found to be insufficient due to missing values and typographical errors

- **Probabilistic matching**
  - Produces sensitivity and specificity based on the frequency and uniqueness of the data elements

Sensitivity: Test's ability to identify positive results

$$\text{sensitivity} = \frac{\text{number of true positives}}{\text{number of true positives} + \text{number of false negatives}}$$

$$= \text{probability of a positive test given that the patient is ill}$$

Example:
If 100 screening tests were done for strep throat, and 60 patients have strep throat, but only 45 patients were identified as having strep throat:

Truth x test	Test – strep throat=no	Test – strep throat=yes	
Strep throat=no	32	8	40
Strep throat=yes	15	45 (row percent 45/(45 + 15) = .75)	60
Total	47	53	100

Specificity: Ability of the test to identify negative results

$$\text{specificity} = \frac{\text{number of true negatives}}{\text{number of true negatives} + \text{number of false positives}}$$

$$= \text{probability of a negative test given that the patient is well}$$

Example:
If 100 screening tests were done for strep throat, and 40 patients truly did not have strep throat, but only 32 patients were identified as not having strep throat:

Truth x test	Test – strep throat=no	Test – strep throat=yes	
Strep throat=no	32(row percent 32/(32 + 8) = .80)	8	40
Strep throat=yes	15	45	60
Total	47	53	100

Recent Literature

- Tromp, et all[1] determined that probabilistic matching on four variables outperformed deterministic matching
- 2009 systematic review by Silviera et al[2]
  - Record and field quality is a better determinant of accuracy than database size
  - More studies are needed to determine the accuracy of linkage procedures

[1] Tromp, M, Ravelli, AC , Bonsel, GJ, Hasman, A, Reitsma, JB, Results form simulated data sets: probabilistic record linage outperforms deterministic record linkage. J Clin Epidemiol. 2011 May; 64(5):565-72

[2] Silveira DP, Artmann E., Accuracy of probabilistic record linkage applied to health databases: systematic review. Rev Saude Publica. 2009 Oct;43(5):875-82

Agenda

- Purpose
- Literature Review
- **Proposed Algorithms**
- Results
- Summary
- Next Steps

Proposed Algorithms: REP Exploration

Rochester Epidemiology Project (REP):

- Unique research infrastructure system
- Links together medical records from SE MN Providers
- Conducts population-based descriptive, case-control, historical and prospective cohort, and cross-sectional research studies

What have we learned?

Proposed Algorithms: Approach

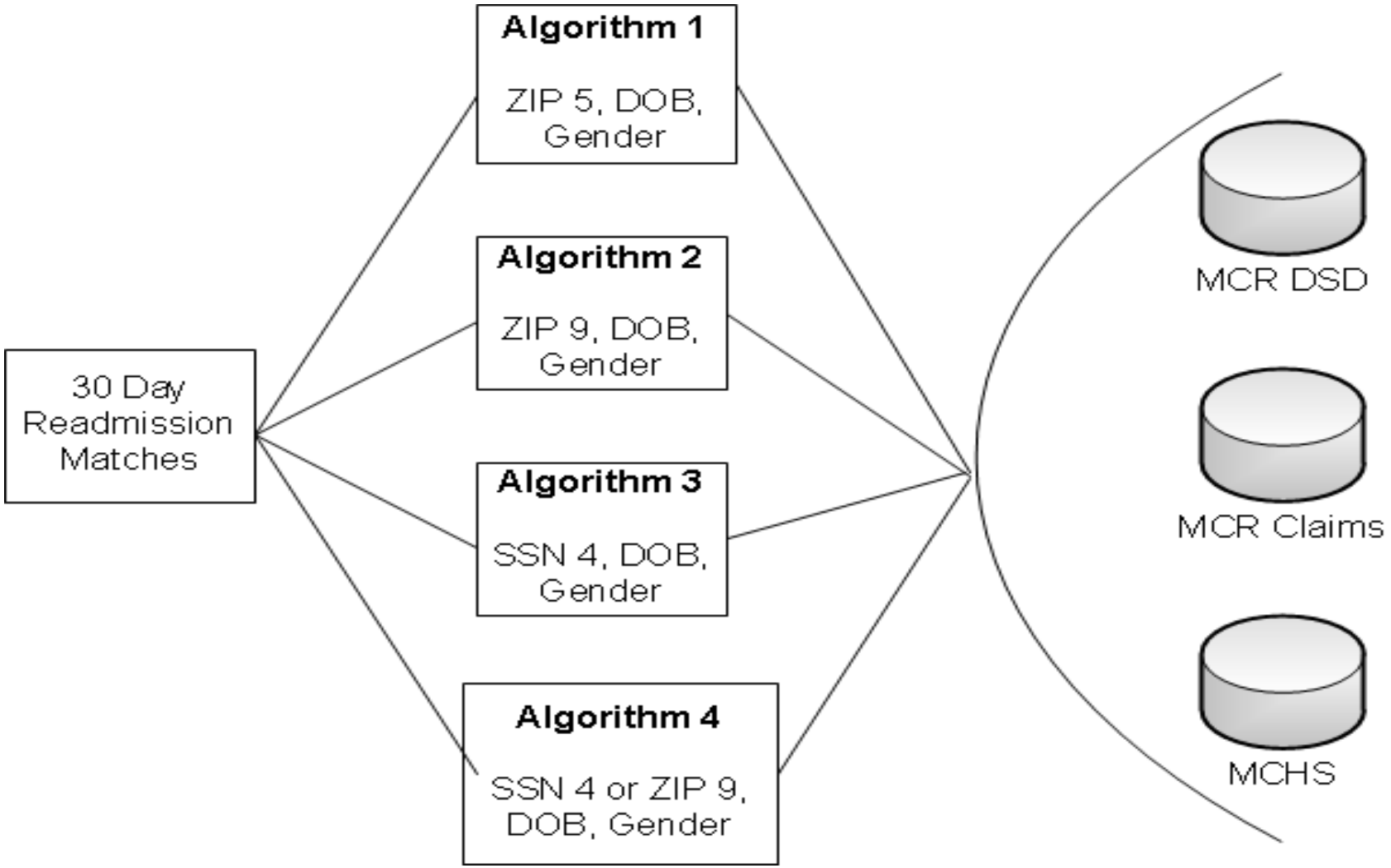
- **We tested 4 different algorithms**:
  - Algorithm 1: DOB, gender, 5-digit zip
  - Algorithm 2: DOB, gender, 9-digit zip
  - Algorithm 3: DOB, gender, 4-digit SSN (required for a match)
  - Algorithm 4: DOB, gender, 4-digit SSN when available, otherwise zip code

Proposed Algorithms: Methodology

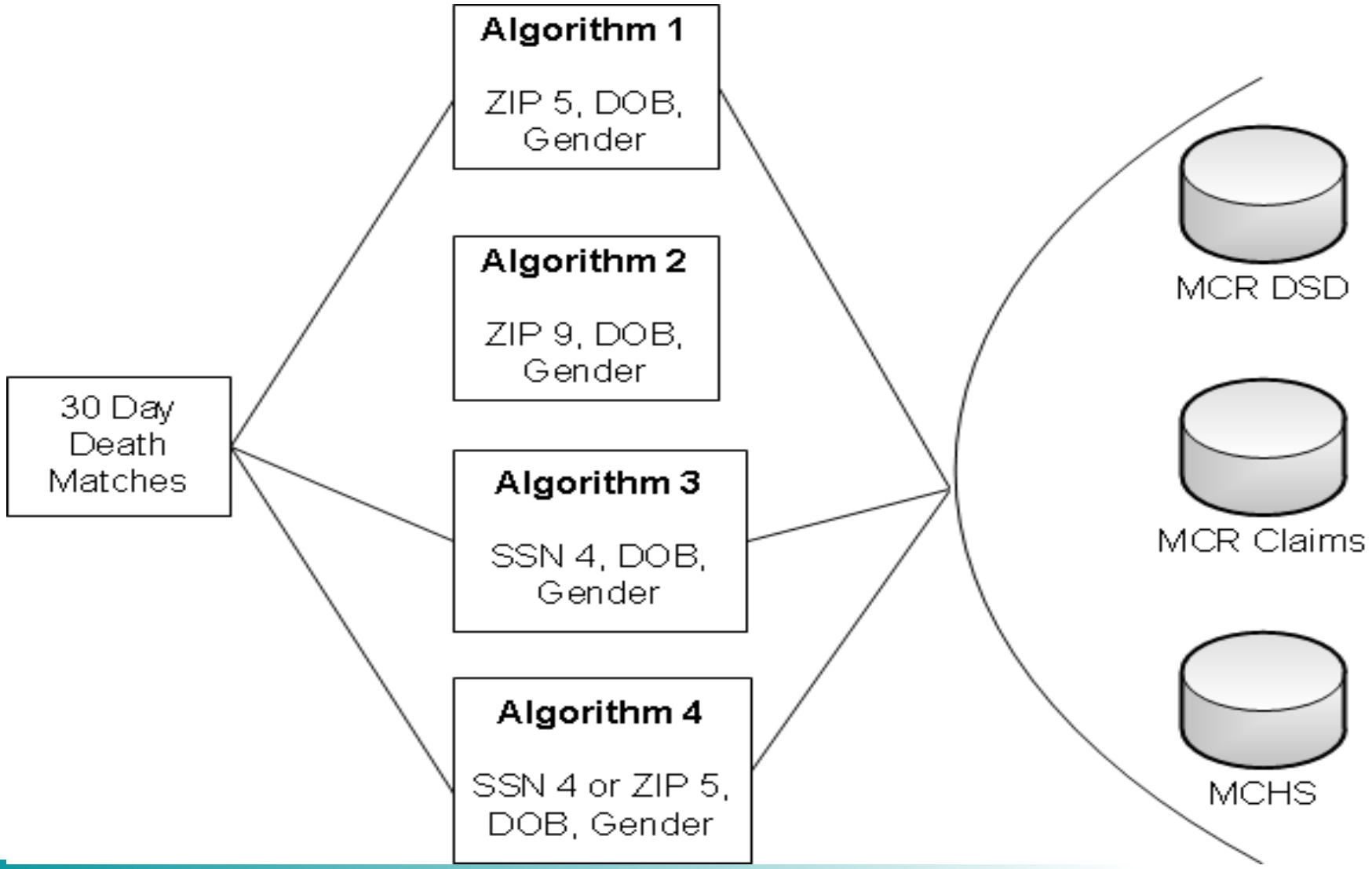
Data used to test approach:

- Decision Support Database (DSD-billing) within Mayo Clinic Rochester (MCR)
- Hospital claims data (data sent to MHA) from MCR
- Mayo Clinic Health System (MCHS) data

Proposed Algorithms: 30-Day Readmission



Proposed Algorithms: 30-Day Death



Agenda

- Purpose
- Literature Review
- Exploration through REP project
- Proposed Algorithms
- **Results**
- Summary
- Next Steps

Results: 30-Day Readmissions- Algorithm 1 - MCR DSD

- With this algorithm, we have a sensitivity of 98.4% and a specificity of 99.7%

Actual 30 Day Readmission (total %, row %, col %)	30 Day Readmission Algorithm		Total Discharges
	No readmit	Readmit	
No readmit	31,386	95	31,481
	83.73%	0.25%	83.98%
	99.70%	0.30%	
	99.70%	1.58%	
Readmits	96	5,908	6,004
	0.26%	15.76%	16.02%
	1.60%	98.40%	
	0.30%	98.42%	
Total Discharges	31,482	6,003	37,485
	83.99%	16.01%	100.00%

Results: 30-Day Readmissions- Algorithm 2- MCR DSD

- With this algorithm, we have a sensitivity of 97.3% and a specificity of 99.99%

Actual 30 Day Readmission(tot %, row %, col %)	30 Day Readmission Algorithm		Total Discharges
	No readmit	Readmit	
No readmit	31,479	2	31,481
	83.90%	0.01%	83.98%
	99.99%	0.01%	
	99.49%	0.03%	
Readmits	162	5,842	6,004
	0.43%	15.58%	16.02%
	2.70%	97.30%	
	0.51%	99.97%	
Total Discharges	31,641	5,844	37,485
	84.41%	15.59%	100.00%

Results: 30-Day Readmissions- Algorithm 3- MCR DSD

- With this algorithm, we have a sensitivity of 92.94% and a specificity of 100%

Actual 30 Day Readmission (tot %, row %, col %)	30 Day Readmission Algorithm		Total Discharges
	No readmit	Readmit	
No readmit	31,481	0	31,481
	83.98%	0.00%	83.98%
	100.00%	0.00%	
	98.67%	0.00%	
Readmits	424	5,580	6,004
	1.13%	14.89%	16.02%
	7.06%	92.94%	
	1.33%	100.00%	
Total Discharges	31,905	5,580	37,485
	85.11%	14.89%	100.00%

Results: 30-Day Readmissions- Algorithm 4 - MCR DSD

- With this algorithm, we have a sensitivity of 99.83% and a specificity of almost 100%

Actual 30 Day Readmission (tot %, row %, col %)	30 Day Readmission Algorithm		Total Discharges
	No readmit	Readmit	
No readmit	31,480	1	31,481
	83.98%	0.00%	83.98%
	100.00%	0.00%	
	99.97%	0.02%	
Readmits	10	5,994	6,004
	0.03%	15.99%	16.02%
	0.17%	99.83%	
	0.03%	100.00%	
Total Discharges	31,490	5,995	37,485
	84.01%	15.99%	100.00%

Results: 30-Day Readmissions for MCR: DSD

		Sensitivity	# visits with a true readmit missed	Specificity	# falsely identified visits with a readmissions
Algorithm 1	DOB, gender, 5-digit zipcode	98.4%	96	99.7%	95
Algorithm 2	DOB, gender, 9-digit zipcode	97.3%	162	99.99%	2
Algorithm 3	DOB, gender, last 4 of SSN	92.9%	424	100%	0
Algorithm 4	DOB, gender, last 4 of SSN or 9-digit zip if unavailable	99.8%	10	100%	1

Results: 30-Day Readmissions for MCR: Claims

		Sensitivity	# visits with a readmissions algorithm missed	Specificity	# falsely identified visits with a readmissions
Algorithm 1	DOB, gender, 5-digit zipcode	97.9%	108	99.7%	81
Algorithm 2	DOB, gender, 9-digit zipcode	96.5%	174	100%	0
Algorithm 3	DOB, gender, last 4 of SSN	92.8%	358	100%	0
Algorithm 4	DOB, gender, last 4 of SSN or 9-digit zip if unavailable	99.8%	12	100%	0

Results: 30-Day Readmissions for MCHS

		Sensitivity	# visits with a readmissions algorithm missed	Specificity	# falsely identified visits with a readmissions
Algorithm 1	DOB, gender, 5-digit zipcode	97.99%	159	99.7%	169
Algorithm 2	DOB, gender, 9-digit zipcode	84.9%	1174	99.9%	29
Algorithm 3	DOB, gender, last 4 of SSN	93.5%	505	100%	0
Algorithm 4	DOB, gender, last 4 of SSN or 9-digit zip if unavailable	99.2%	64	99.97%	14

Results: 30-Day Deaths for MCR DSD

		Sensitivity	# visits with a death algorithm missed	Specificity	# falsely identified visits with a death
Algorithm 1	DOB, gender, 5-digit zipcode	85.6%	118	99.96%	15
Algorithm 2	DOB, gender, 9-digit zipcode	Cannot be applied because MN Death tapes do not provide 9-digit SSN			
Algorithm 3	DOB, gender, last 4 of SSN	94.1%	49	100%	0
Algorithm 4	DOB, gender, last 4 of SSN or 5-digit zip if unavailable	95.4%	38	99.99%	4

Results: 30-Day Deaths for MCR Claims

		Sensitivity	# visits with a death algorithm missed	Specificity	# falsely identified visits with a death
Algorithm 1	DOB, gender, 5-digit zipcode	84.9%	118	99.96%	14
Algorithm 2	DOB, gender, 9-digit zipcode	Cannot be applied because MN Death tapes do not provide 9-digit SSN			
Algorithm 3	DOB, gender, last 4 of SSN	93.39%	52	100%	0
Algorithm 4	DOB, gender, last 4 of SSN or 5 digit zip if unavailable	94.5%	41	99.99%	4

Results: 30-Day Deaths for MCHS Data

		Sensitivity	# visits with a death algorithm missed	Specificity	# falsely identified visits with a death
Algorithm 1	DOB, gender, 5-digit zipcode	84.4%	223	99.95%	25
Algorithm 2	DOB, gender, 9-digit zipcode	Cannot be applied because MN Death tapes do not provide 9-digit SSN			
Algorithm 3	DOB, gender, last 4 of SSN	94.5%	79	99.99%	4
Algorithm 4	DOB, gender, last 4 of SSN or 5-digit zip if unavailable	95.8%	60	99.98%	9

Agenda

- Purpose
- Literature Review
- Proposed Algorithms
- Results
- **Summary**
- Next Steps

Summary

- We presented 4 algorithms using 3 different data sources for post-hospitalization 30-day readmission and 30-day death

- DOB, gender, 4-digit SSN along with zip code appeared to improve the overall sensitivity and/or specificity
  - Proved true for all datasets
  - Both readmissions and death data

Agenda

- Purpose
- Literature Review
- Proposed Algorithms
- Results
- Summary
- **Next Steps**

Next Steps

- Will be sending early next week

Questions and Discussion