Match Point- Data Matching Across Source

4:15PM EST | November 17th, 2020 | 2020 SyS Symposium

Moderated by Kelley Chester, DrPH, MPH – Thought Bridge, LLC

Speakers:
Josh Levy, MS – Levy Informatics, LLC + Kelley Chester, DrPH, MPH – Thought Bridge, LLC
Dana Quesinberry, JD – Kentucky Injury Prevention and Research Center
Madison Schaeffer, MS, MPH; Keaton Hughes, MPH; & Lori Zigich, MPH – New Mexico Department of Health

National Syndromic Surveillance Program Community of Practice
Advancing the Science and Practice of Syndromic Surveillance

KAHUINA CONSULTING

CSTE
Council of State and Territorial Epidemiologists

Josh Levy, MS, Levy Informatics, LLC
Kelley Chester, DrPH, MPH, Thought Bridge, LLC

CSTE

KAHUINA
CONSULTING

Integrating State EMS Data into the BioSense Platform

Josh Levy & Barbara Massoudi

Surveillance Practice and Implementation Subcommittee

June 5, 2020

THOUGHTBRIDGE

INITIAL PROBABALISTIC LINKING ALGORITHM DEVELOPMENT

THOUGHTBRIDGE

Probabilistic Linkage Process

Pre-linkage	Identify fields for linking
Test Data	Create "gold standard" hand-evaluated test data set
Parameters	Iterate through probabilistic linkage test runs to find proper parameters
Cutoffs	Determine proper cutoffs to maximize TP and minimize FP and FN
Full Linkage	Run linkage on full data set and evaluate results

THOUGHTBRIDGE

Suitability of fields for linking

Field	EMS	ED
Birth date	> 99% complete	< 50% complete
Age	> 99% complete	100% complete
Gender	> 99% complete	> 99% complete
Patient zip code	92% complete (100% after dropping missing)	> 99% complete
Hospital code	100% complete	100% complete
Arrival time*	> 99% complete	100% complete
Injury present†	95% complete	13% complete
Text fields‡	Too different for matching	

* ED data does not have an arrival time – it has C_Visit_Date_Time: the time of the first EMR message per patient.
† ED data has a text indicator in the Category_flat field, but this is not standardized and is inconsistently recorded.
‡ Text fields checked are in supplementary slide 1

THOUGHTBRIDGE

Test sets for evaluation of probabilistic linkage performance

- Test set 1:
  - Matches from simple fuzzy inner join with age, arrival time, gender, zip code, and hospital
- Test set 2:
  - Manual evaluation and classification of one month of probabilistic linkage matches using age, arrival time, gender, zip code, and hospital

Iterating probabilistic linkage using the test sets on ONE month of data

- Tested many different conditions for matching variables
  - Example: Arrival time
    - varied windows for full/partial matching
    - Idaho has two time zones, and ED records do not record the actual arrival time
- Goals
  - Creation of an optimal EM object to score matches in the full data set
  - Determination of proper weight to give priors

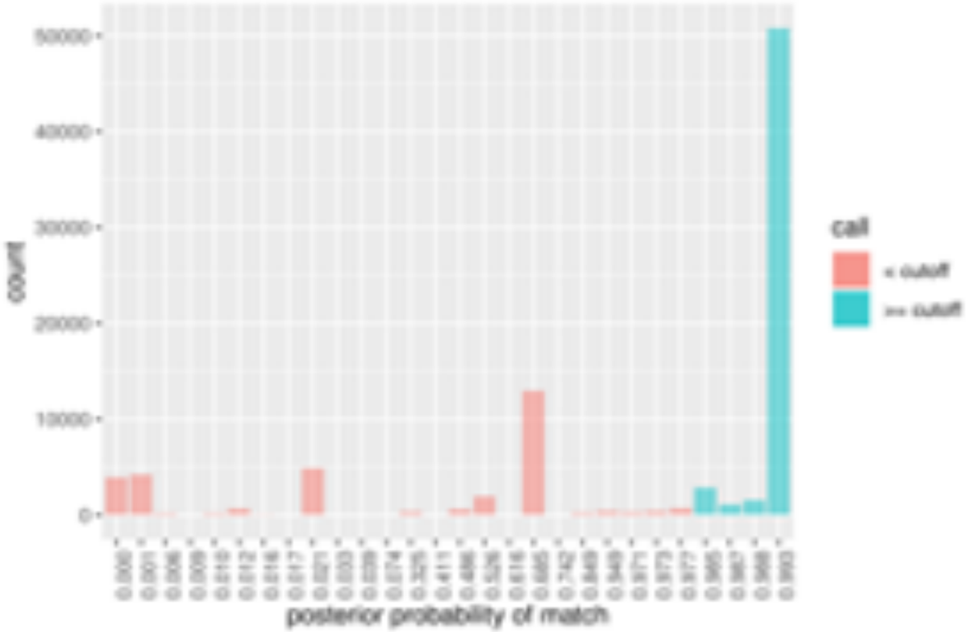
Scaling up to the full data set

Methodology and details

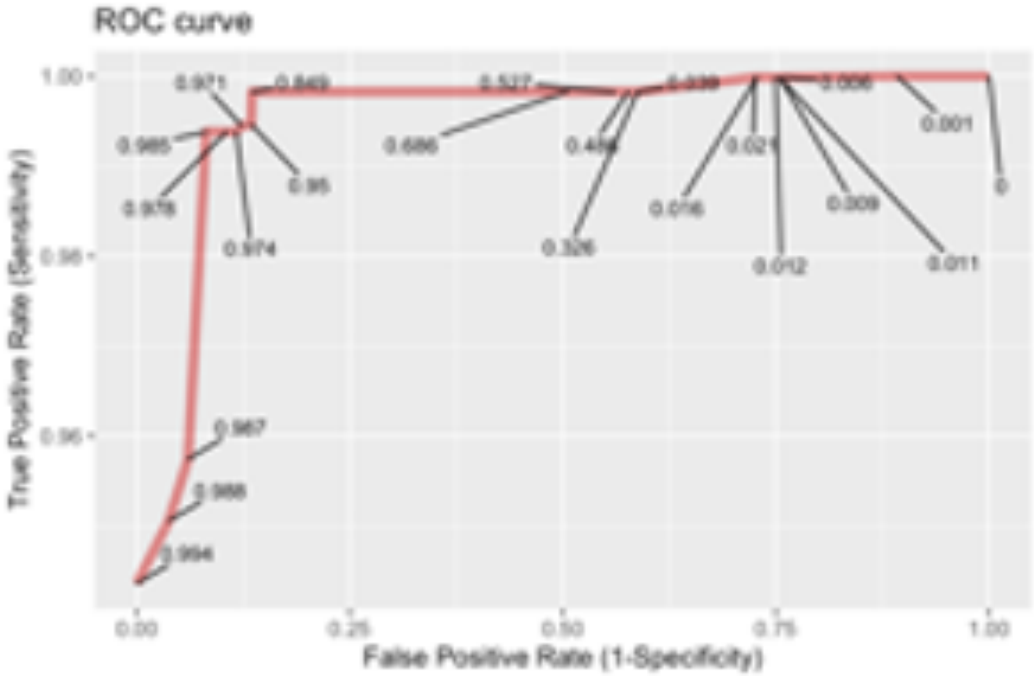
- Applied EM to full data set
- Had to run as twelve chunks otherwise the system crashed
- For each EMS run, took the matches with the highest posterior match probability (prob_m)
  - Sometimes, there are multiple matches for an EMS run with the same prob_m
  - Spot-checking these, there sometimes do appear to be multiple ED visit entries for what seems to be the same patient
  - Others of these just have certain match patterns with an equivalent score

Evaluation of Matching Algorithm

Match rate = matches/records = 52,369/64,629 = **81%**



Our chosen cutoff of 0.985 shows a very high TP rate and lies right before a large FP jump. (AUC = 0.99)

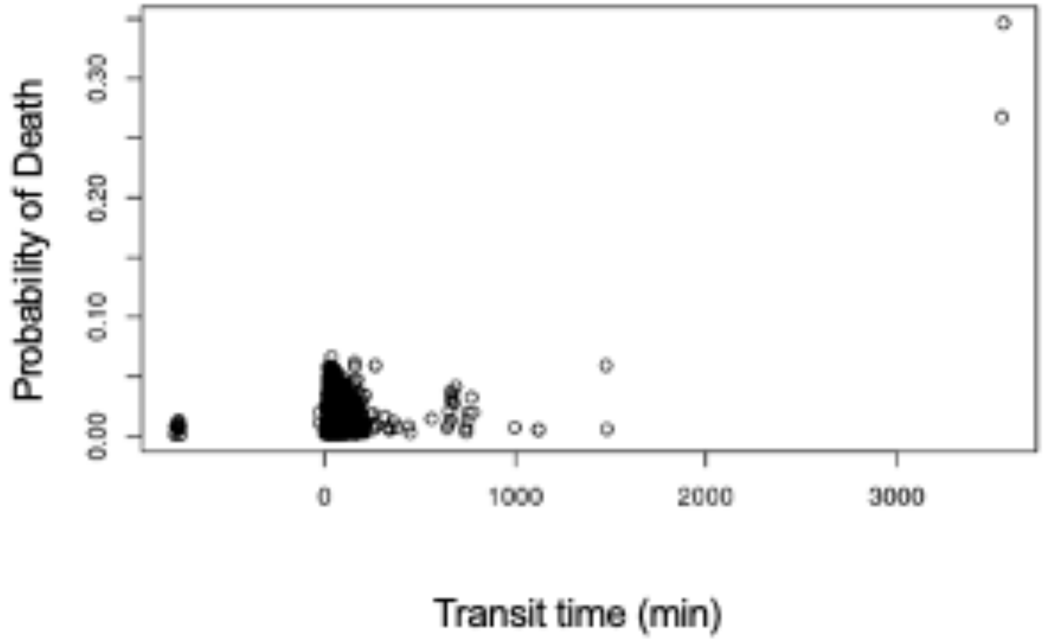


What percentage of Narcan administrations end up being actual opioid overdoses?

- 435 of the 52,369 matched EMS-ED records had a naloxone administration in relation to overdose in the EMS record during the time period examined
- ICD-10-CM codes from the ED record diagnoses were manually examined and used to classify the naloxone cases as either opioid poisoning, other toxic substance effect, or neither
- 77 out of 435 records (17.7%) had an opioid poisoning diagnosis
- 51 (11.7%) matched toxic effects of unknown substance
- Opioid overdoses may be overestimated by up to 70%

Does a longer trip to the ER increase the risk of death?

- ED record discharge disposition was used to create a binary variable for death
- Multiple logistic regression was used to test death + age + sex vs. transit time
- We were able to reject the null hypothesis, but overall there is only a 0.08% increase in death occurrence per additional minute of transit time – a weak effect
- The plot shows that probability of death does not really increase except for extremely long (>1500 min) transit times



THOUGHTBRIDGE

Acknowledgements

- Idaho Department of Health and Welfare, Division of Public Health
  - Kris K. Carter, DVM, MPVM, DACVPM
  - John Kramer
- Centers for Disease Control
  - Mike Coletta
  - Aaron Kite-Powell

Thought Bridge Team

Kelley Chester, MPH, DrPH
Co-founder/Principal

Barbara Massoudi, MPH, PhD
Co-founder/Principal

Josh Levy, MS
Data Scientist

THOUGHTBRIDGE

Wrap-up

Questions?

josh@levyinformatics.com
919-672-4192
kelley.chester@thoughtbridgellc.com
404-242-8740
barbara.massoudi@thoughtbridgellc.com
770-296-1810

THOUGHTBRIDGE

Dana Quesinberry, JD

Kentucky Injury Prevention and Research Center

CSTE

KAHUINA CONSULTING

Madison Schaeffer, MS, MPH; Keaton Hughes, MPH; & Lori Zigich, MPH
New Mexico Department of Health

CSTE

KAHUINA
CONSULTING

NEW MEXICO DEPARTMENT OF HEALTH

Investing for tomorrow, delivering today.

EMS Influenza-like Illness Case Definition: Prediction Validity Comparison [Syndromic Surveillance Emergency Department Visits to Laboratory Tests], New Mexico, 2018-2019

Madison Schaeffer, EMS Epidemiologist

Lori Zigich, Syndromic Surveillance Epidemiologist

Keaton Hughes, Syndromic Surveillance Epidemiologist

Objectives

- Syndromic surveillance database overview
- Emergency Medical Systems (EMS) database overview
- Linking ED visits and EMS runs
- Positive and negative predictive values
- Trend association analysis
- Real world implications

Syndromic Surveillance Representativeness

- 37 out of 38 emergency departments
  - Includes ED visits, inpatient admissions, and many outpatient visits
- 10+ urgent care centers
- 1 major primary care clinic in the Albuquerque region

1190 S. St. Francis Drive • Santa Fe, NM 87505 • Phone: 505-827-2613 • Fax: 505-827-2530 • nmhealth.org

Syndromic Surveillance Timeliness

- 90% of clinical information received within 48 hours

- Found some facilities batch submissions only once a day!?
  - Working with facilities to increase to 6-8 times daily

- Currently we send 3-4 batches to BioSense daily
  - Recently re-scheduled NM's repository so data is updated 8x daily
    - This is to better support more timely data linkages to the patient identifiers

Syndromic Surveillance Data Elements

- Demographics
  - Age
  - Gender
  - Race/Ethnicity
- Geographical Information
- Indicators
  - Chief complaint
  - Diagnoses codes
  - Triage notes/clinical impressions
- NM repository maintains patient identifiers to support data linkage efforts

1190 S. St. Francis Drive • Santa Fe, NM 87505 • Phone: 505-827-2613 • Fax: 505-827-2530 • nmhealth.org

Use Case: Influenza-like Illness

- Active surveillance of ILI-related ED visits
  - ILI CCDD v1

CCDD TERMS: (,^fever^,andnot,^denies fev^,andnot,^shot^,andnot,^afeb^,andnot,^no fev^,andnot,^no temp^,or,^chill^,or,^pyrexia^,or,^
febrile^,or,^high temp^,or,^elevated temp^,andnot,^denies any elevated temp^,andnot,^no elevated temp^,or,^feeling hot^,or,^feels
hot^,or,^fvr^,andnot,^denies fvr^,andnot,^no fvr^,),and,(,^cough^,andnot,^denies cough^,or,^sore ^th[o|r][a|o][a|r]t^,or,^soreth[o|r][a|o]
[a|r]t^,or,^strep^,or,^pharyn^,or,^upper resp^,),or,(,^[;/]J09^,or,^[;/]J10^,or,^[;/]J11^,or,^[;/]487.[018];^,or,^[;/]487[018];^,or,^[;/]487.[018],or,^[;/
]487[018],or,^[;/]488.[018][19];^,or,^[;/]488[018][19];^,or,^[;/]488.[018][19],or,^[;/]488[018]
[19],or,^442696006^,or,^442438000^,or,^6142004^,or,^195878008^,or,^influenza^,andnot,^vaccin^,andnot,^shot^,andnot,^immunizat^,or,^ flu
^,andnot,^shot^,andnot,^stomach^,andnot,^vaccin^,andnot,^immuniza^,or,^flu like^,or,^flulike^,or,^flu symptom^,),andnot,^[;/]A08.4^,andnot,^[;/
]A084^

- Utilized for:
  - Quality assurance to better understand the data
  - Augment traditional influenza surveillance
  - Validate other influenza surveillance sources

1190 S. St. Francis Drive • Santa Fe, NM 87505 • Phone: 505-827-2613 • Fax: 505-827-2530 • nmhealth.org

EMS Overview

- Represent runs from 300+ New Mexico EMS agencies (both ground and flight) including rural and tribal services
- Full patient care record
  - Demographics
  - Symptoms
  - Record of all procedures, medications
  - Geographic data
- Majority of records submitted within 24 hours
- Bias toward severe illness

ILI-related EMS Definition Development

- Impression fields
  - Important to remember that EMS does not diagnose (i.e. no rapid flu tests in the field)
  - Surveillance is based on symptoms and provider impression
  - Vitals fields are frequently incomplete
- Patient care narrative
- Refined based on SyS ILI definition after reviewing records

1190 S. St. Francis Drive • Santa Fe, NM 87505 • Phone: 505-827-2613 • Fax: 505-827-2530 • nmhealth.org

Use Case: Linking with Additional Data Sources

- ILI-related EMS runs
- ILI-related emergency department visits
- Positive laboratory tests at New Mexico Scientific Laboratory Division

Data Linkage Methodology

- Time period: 2018-19 CDC Flu Season, Sep 30, 2018 – May 18, 2019
- Variables supporting linkage:
  - First name, last name, DOB, SSN, gender, and dates of healthcare encounters
- Utilized Link King
  - Deterministic and probabilistic matching

1190 S. St. Francis Drive • Santa Fe, NM 87505 • Phone: 505-827-2613 • Fax: 505-827-2530 • nmhealth.org

Syndromic Surveillance-EMS Linkage Results (Preliminary)

- 204,233 Total EMS runs
  - 1,618 ILI-related EMS runs
- ~502,000 Total ED visits
  - 17,195 ILI-related ED visits
- Successfully linked 60,938 EMS runs with ED visits
  - Positive predictive value = 33.2%
  - Negative predictive value = 0.94%

	ED ILI positive	ED ILI negative	Total
EMS ILI positive	201	403	604
EMS ILI negative	575	60535	61110
Total	776	60938	

Analysis Methodology and Trend Association insight

Data sources:

- 2,150 suspected ILI-related EMS runs
- 19,065 ILI-related ED visits
- 1,650 positive laboratory tests at New Mexico Scientific Laboratory Division

For the 2018-2019 flu season

- EMS Data explained 79.7% of variance in PCR positive influenza laboratory tests ($R^2 = 0.79$, $p < 0.001$)
- EMS Data explained 84.7% of variance in SyS ILI ED visits ($R^2 = 0.84$, $p < 0.001$)

1190 S. St. Francis Drive • Santa Fe, NM 87505 • Phone: 505-827-2613 • Fax: 505-827-2530 • nmhealth.org

Real-world Implications

- EMS data are widely available in all 50 states and cover areas that ED SyS data do not
- Particularly valuable for tracking in frontier, rural, and tribal areas
- Could be expanded to other infectious disease tracking, e.g. COVID-19

Thank you and questions?

National Syndromic Surveillance Program

Community of Practice

Advancing the Science and Practice of Syndromic Surveillance

KAHUINA

CONSULTING

CSTE

Council of State and Territorial Epidemiologists

Block Schedule

Training | Methods | Innovation | Data Sharing

	Training	Methods	Innovation	Data Sharing
1:30-2PM EST	Keynote Address			
	...BREAK...			
2:15 - 3PM EST	Training: NSSP ESSENCE 101	Breakout: Spike, Alert, GO!	Breakout: Data Access and Integration	Breakout: Successful State-led Data Sharing Collaborations
		...BREAK...		
3:15-4PM EST		Lightning: Significant Findings	Roundtable: Bringing New Partners to the Table	Breakout: Can I Do That? Ask a Lawyer.
	...BREAK...			
4:15-5PM EST	Training: Custom Syndrome Definition in NSSP ESSENCE	Breakout: Match Point - Data Matching Across Sources		Roundtable: Establishing a Public Use Shared Data Set
	...BREAK...			
5:15-6PM EST	Virtual Happy Hour			

KAHUINA CONSULTING CSTE