Injury Control and Probabilistic Linkage

Lenora Olson

Larry Cook

Gordon Smith

Michael Bauer

Andrea Thomas

Introduction to Probabilistic Linkage

Larry Cook
Utah CODES Project
Intermountain Control Research Center
University of Utah

Probabilistic Linkage

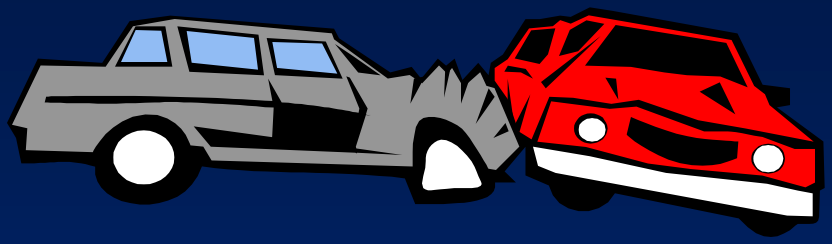
- Probabilistic is a method that uses properties of variables common to databases to determine the probability that two records refer to the same person and/or event
- Probabilistic linkage is NOT a method that produces results that are probably right

Utah CODES

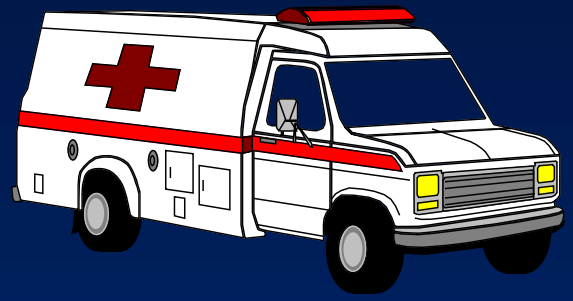
- Crash Outcome Data Evaluation System
- Funded by the National Highway Traffic Safety Administration (NHTSA)
- Goal – better understand the medical outcomes of crashes

Crash Outcomes

- 1996 CODES Report to Congress
- Are safety belts and motorcycle helmets effective at preventing injuries resulting from motor vehicle crashes?
  - Safety belts are 45% effective
  - Motorcycle helmets 26% effective
    - 66% effective at preventing brain injuries



Crash



EMS

Analysis
Database

ED

Inpatient

Probabilistic Linkage Studies

Shop Class Injuries

One Year ED

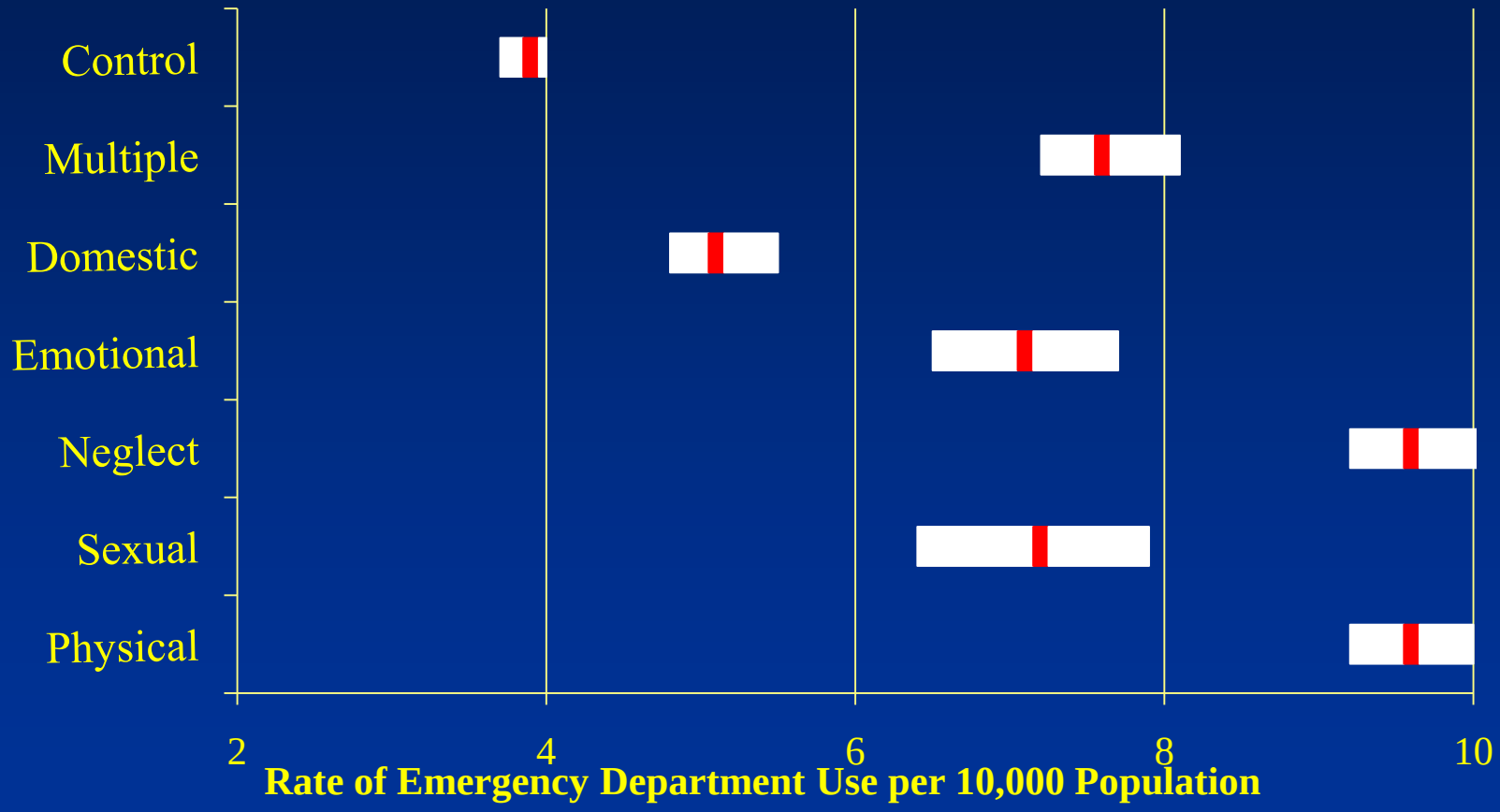
- 167 in class injuries
- 45 seen at ED
- ½ were saw related
- $16,571 ED charges

Five Year Hospital Discharge

- 1,008
- 7 admitted
- 6 table saw related
- 3 amputations
- $26,767 hospital charges

Knight S, Cook LJ, Nechodom PJ, Olson LM, Reading JC, Dean JM. (2001). Shoulder belts in motor vehicle crashes: a statewide analysis of restraint efficacy. *Accid Anal Prev*, *33*(1), 65-71.

Rates of ED Use by Type of Child Abuse



Guenther E, Knight S, Olson LM, Dean JM. Prediction of child abuse risk from emergency department use. *J Pediatrics*. 2008

Probabilistic Linkage Basics

Let’s Play 20 Questions

I’m thinking of a person

Probabilistic Linkage Theory

Reliability (m)

Probability that a common variable agrees on a **matched** pair. Approximately 1 - error rate.

Discriminating Power (u)

Probability that a common variable agrees on an **unmatched** pair. Approximately the probability of agreeing by chance.

Record Linkage with Imperfect Data

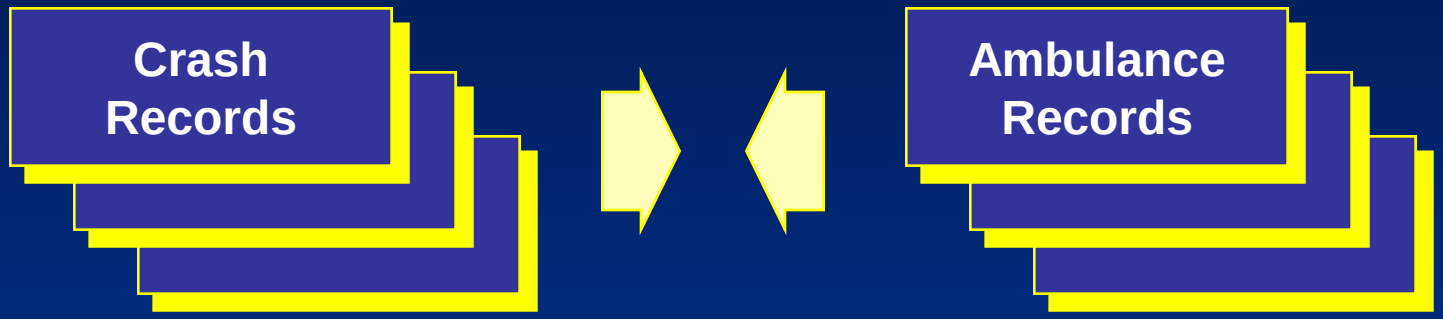
Crash Record

Mary Smith F 05/05/45 07/15/96 11:40 Weber US5 Seat=1 Belt=N

Ambulance Record

Mary Smith Sanchez F 05/05/44 07/15/96 11:51 Weber Fracture Mem Hosp

Probabilistic Record Linkage



If each ambulance record matches to one crash record in a file of 100,000 crashes then the odds for a match at random are 1:99,999 or probability of true match = 0.00001.

Probabilistic Record Linkage

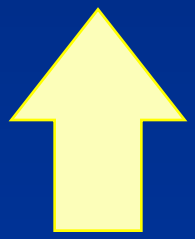
Crash Record

Mary Smith /15/10 11:47 Weber US5 Seat=1 Belt=N

Probability of
true match = 0.0009

Ambulance

Mary Smith S /15/10 11:55 Weber Fracture Mem Hosp



Probabilistic Record Linkage

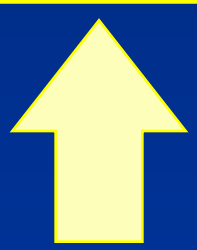
Crash Record

Mary Smith F ... 7 Weber US5 Seat=1 Belt=N

Probability of
true match = .0192

Ambulance Record

Mary Smith Sanchez F ... 5 Weber Fracture Mem Hosp



Probabilistic Record Linkage

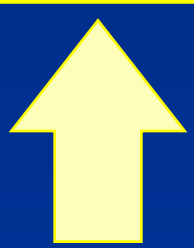
Crash Record

Mary Smith F 05/05/45 US5 Seat=1 Belt=N

Probability of
true match = .0385

Ambulance Record

Mary Smith Sanchez F 05/05/44 Fracture Mem Hosp



Probabilistic Record Linkage

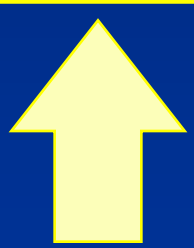
Crash Record

Mary Smith F 05/05/45 07/1 Seat=1 Belt=N

Ambulance Record

Mary Smith Sanchez F 05/05/44 07/1 ture Mem Hosp

Probability of a
true match = 0.1429



Probabilistic Record Linkage

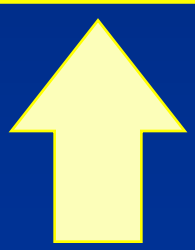
Crash Record

Mary Smith F 05/05/45 07/15/10 11 Belt=N

Ambulance Record

Mary Smith Sanchez F 05/05/44 07/15/10 11 m Hosp

Probability of a
true match = 0.9836



Probabilistic Record Linkage

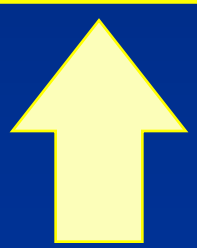
Crash Record

Mary Smith ... 10 11:47 Weber US5 Seat=1 Belt=N

Probability of a true match = 0.9817

Ambulance Re

Mary Smith Sanc ... 10 11:55 Weber Fracture Mem Hosp



Probabilistic Record Linkage

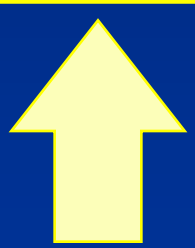
Crash Record

Mary Smith F ...7 Weber US5 Seat=1 Belt=N

Probability of a
true match = 0.9999

Ambulance Record

Mary Smith Sanchez F ...5 Weber Fracture Mem Hosp



Probabilistic Record Linkage

Crash Record

Mary Smith	F	05/05/45	07/15/10	11:47	Weber	US5	Seat=1	Belt=N

Ambulance Record

Mary Smith Sanchez	F	05/05/44	07/15/10	11:55	Weber	Fracture	Mem Hosp

This pair of records has both agreements and disagreements. Our calculations say that the odds are p = 0.9999 that the records refer to the same individual and crash event.

What Do You Need For Probabilistic Linkage

Data Files

- IRBs
- Data sharing agreements

- Variables common to both files
- Variable definitions same on each file

Are Names Necessary for Probabilistic Linkage?

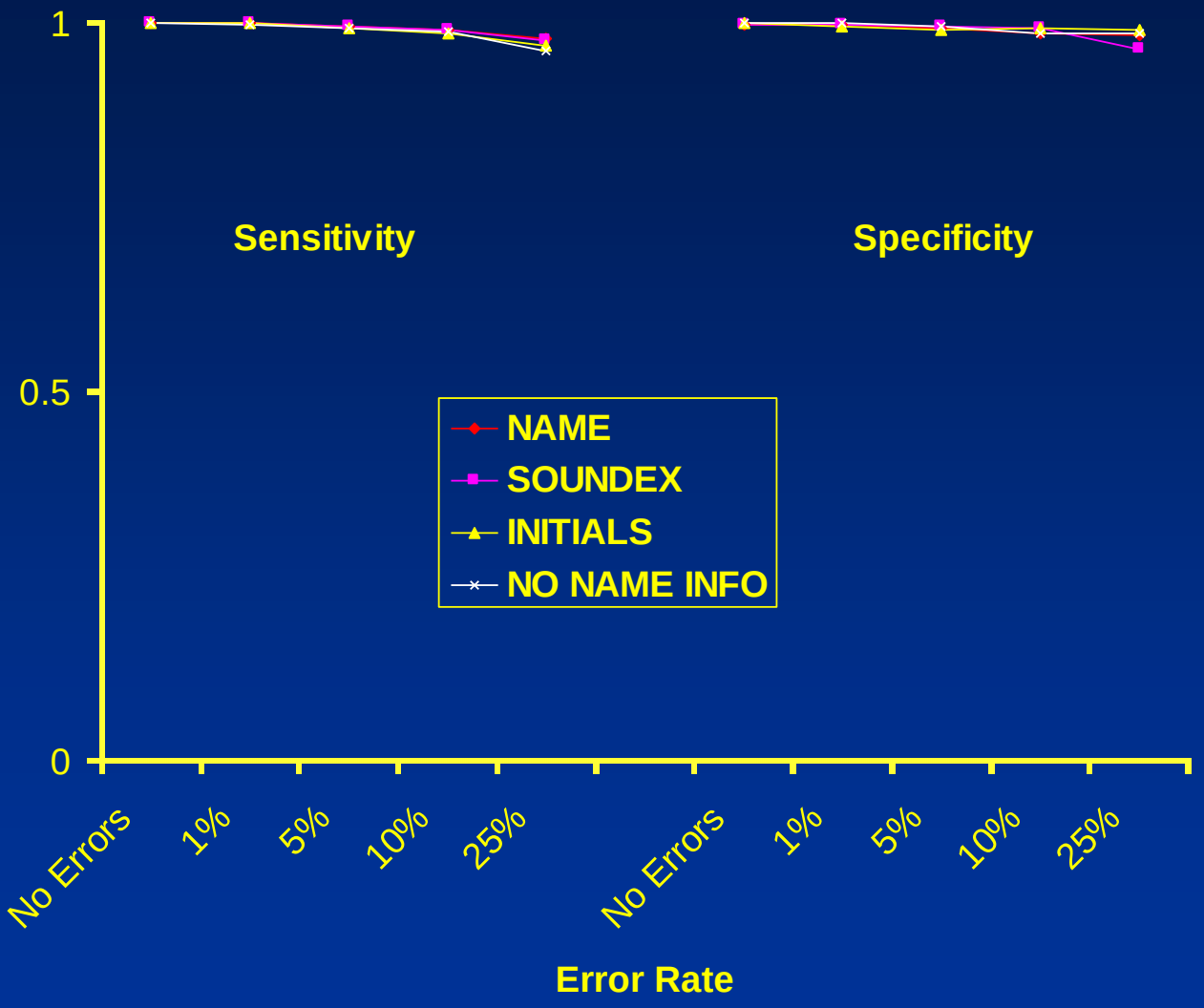
Name Dilema

- Name are powerful identifiers
- Confidentiality concerns
- Names may not be collected in database
- Simulation study to determine effect of name information on linkage projects
  - We know the answers

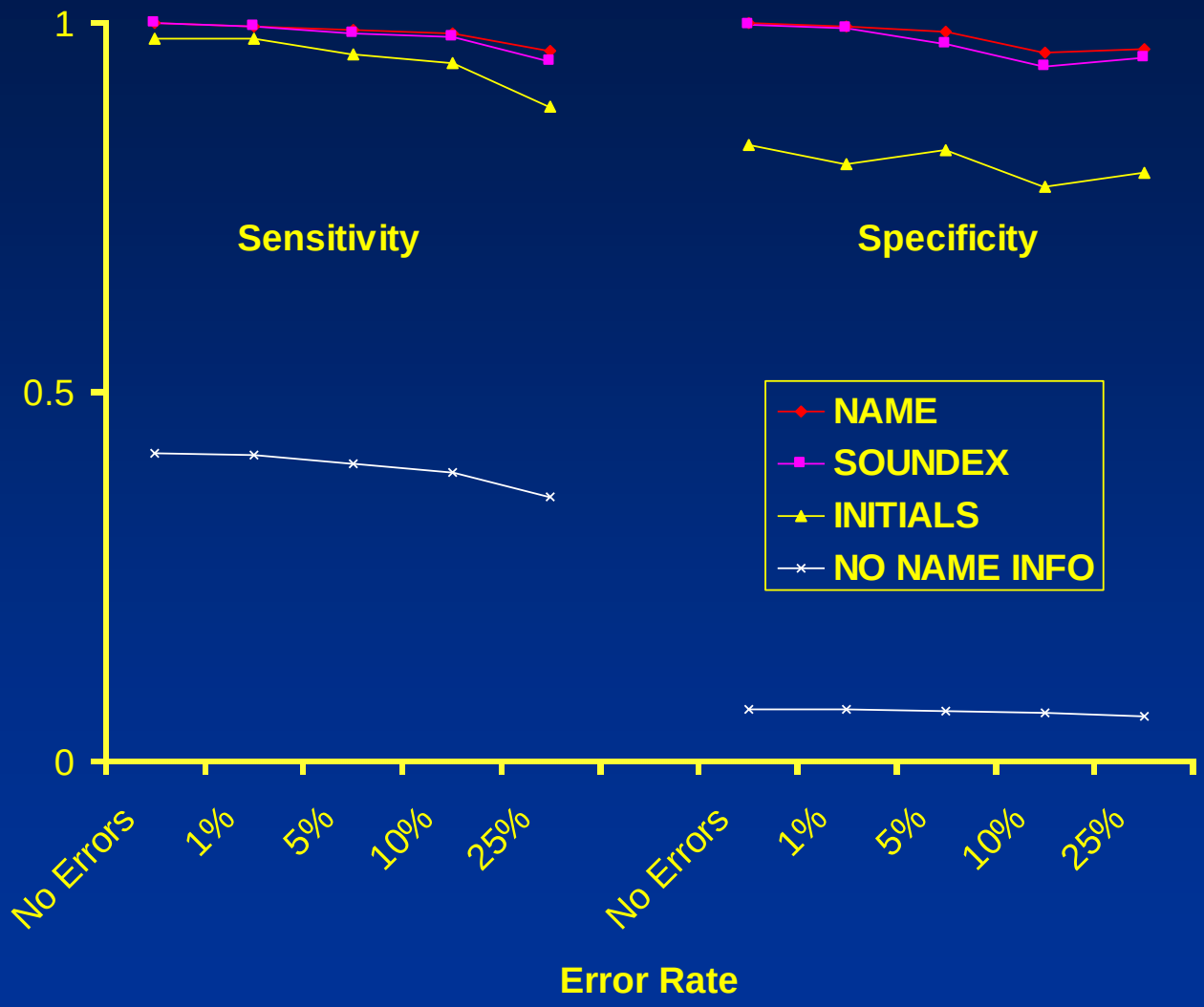
Linkage Performance Measures

- Sensitivity - Ability to recognize true matches
  - % of true matches identified
- Specificity - Ability to recognize incorrect matches
  - 1 – false positive rate

DOB, Gender, County, Time, Incident Date



DOB, Age, Gender, County, Time, Incident Date



Summary

- Is name information necessary?
  - If many non-name identifiers are available then name information may not be needed
  - If few non-name identifiers are available then name information becomes crucial
- Linkage feasibility test
  - Cook LJ, Olson LM, Dean JM. (2001). Probabilistic record linkage: relationships between file sizes, identifiers and match weights. *Methods Inf Med*, *40*(3), 196-203.

Software

Software

- CODES2000/LinkSolv
- Link Plus (CDC)
- Link King
- LinkageWiz
- Write your own
  - *Handbook of Record Linkage Methods for Health and Statistical Studies*, Howard Newcombe
- Pricing - Free to $$$$$

Software Checklist

- Add custom variable types and comparisons
- Unduplication / self match
- Link more than two files
- Size of databases
- Training and documentation

Other Record Linkage Hurdles

- Confidentiality concerns
  - IRBs & data sharing/use agreements
  - Separate tables of identifiers
- Databases
  - Missingness and accuracy of matching fields
  - Standardizing elements
  - Timeliness
- Analysis

Questions?

Larry Cook
295 Chipeta Way
Salt Lake City, UT 84158
801.585.9760
larry.cook@hsc.utah.edu