

# Year 1 Review of the North Carolina Crash Injury Surveillance System Project

KATHY PETICOLAS, MPS, PMP

MIKE DOLAN FLISS, PHD, MPS, MSW

KATHERINE J. HARMON, PHD

ANNA E. WALLER, SCD



THE UNIVERSITY  
*of* NORTH CAROLINA  
*at* CHAPEL HILL

This presentation was supported by the National Center for Injury Prevention and Control of the Centers for Disease Prevention and Control (CDC) under award number CE16-1602. The content is solely the responsibility of the authors and does not necessarily represent the official views of the CDC.

September 24, 2020

Revised March 16, 2021

# Agenda

- Implementation overview
- Successes
- Challenges
- Lessons learned

## NC DPH Data Attribution & Disclaimer

NC DETECT is a statewide public health syndromic surveillance system, funded by the NC Division of Public Health (NC DPH) Federal Public Health Emergency Preparedness Grant and managed through collaboration between NC DPH and UNC-CH Department of Emergency Medicine's Carolina Center for Health Informatics. The NC DETECT Data Oversight Committee does not take responsibility for the scientific validity or accuracy of methodology, results, statistical analyses, or conclusions presented.



# Project collaboration



## NC DHHS Injury and Violence Prevention Branch

- Ingrid Bou-Saada
- Alan Dellapenna
- Kendall Knuth
- **Kathy Peticolas\***
- Scott Proescholdbell

**\*Core team members**

## UNC Injury Prevention Research Center

- Steve Marshall
- **Mike Fliss\***

## UNC Highway Safety Research Center

- **Katie Harmon\***
- Nancy Lefler
- Eric Rodgman

## Carolina Center for Health Informatics

- **Anna Waller\***
- Clifton Barnett
- Dennis Falls
- Amy Ising

# We linked two health outcome data sources to crash report data.



2018 crash  
report data

- Chosen based on experience from prior data linkage projects
- Obtained data owner support prior to project

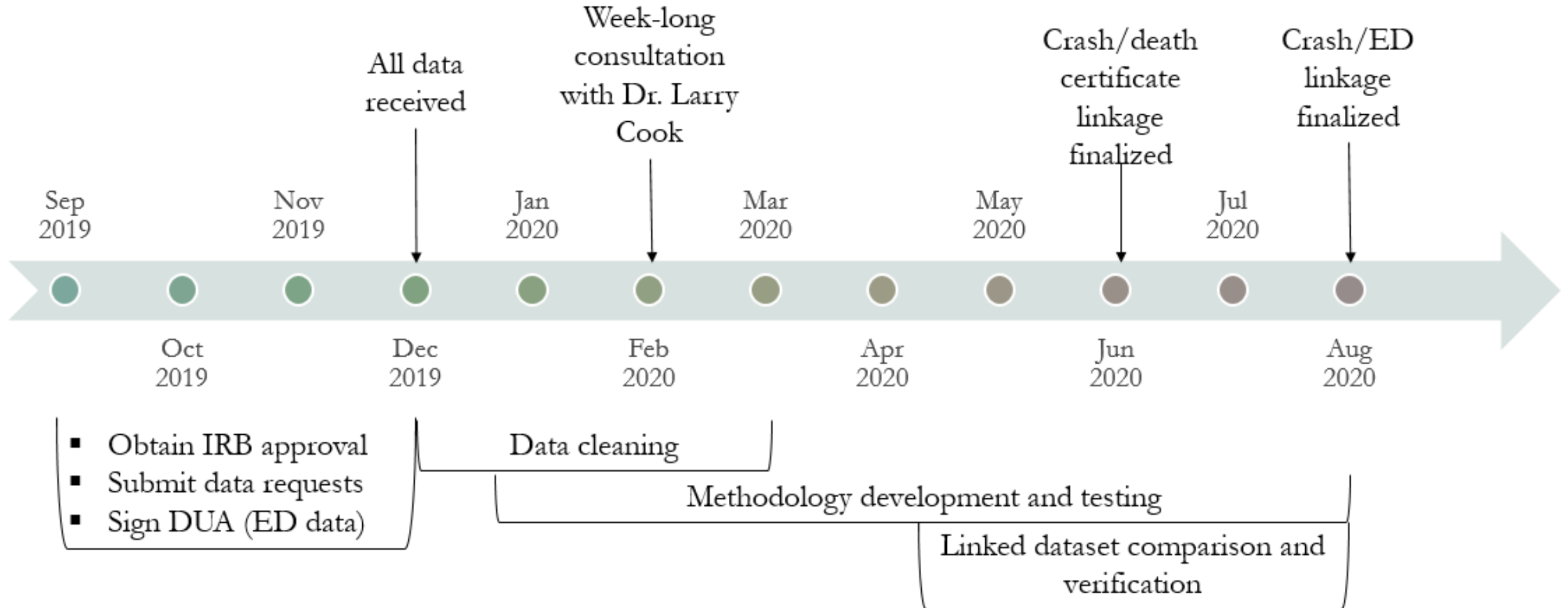


Death  
certificate data

Emergency  
department data  
(NC DETECT)



# The project timeline



# We created cleaned and standardized linkage variables.

Category	Examples
Person demographics	Date of birth Sex Race/ethnicity
Crash characteristics	Whether or not a crash is indicated Whether an injury is present The type of person in the crash (driver, pedestrian, etc.)
Place of residence	Residence zip Residence city
Place of crash	Crash county County of death/ED visit Latitudes/Longitudes
Time	Date of crash/date of death/ED visit

For health data, these were based on cause of death/ diagnosis codes.

We geocoded location data, when possible.

# We tried different linkage methodologies, but focused on deterministic linkage.

Linkage methods	Description
Hierarchical deterministic linkage	Matching up shared linkage variables in different patterns
Recursive partitioning trees	Calculating the 'distance' between linkage variables
Probabilistic linkage	Calculating the likelihood of records matching based on variable frequency
Hand review	Verifying and matching up records through individual review

## Strengths:

- Easy to explain to a multi-disciplinary audience
- High quality results
- Fast
- Replicable in many applications
- Usable for near-real-time linkages

## Challenge:

A sufficient and representative linkage rate

# We compared and verified the results of different methodologies.

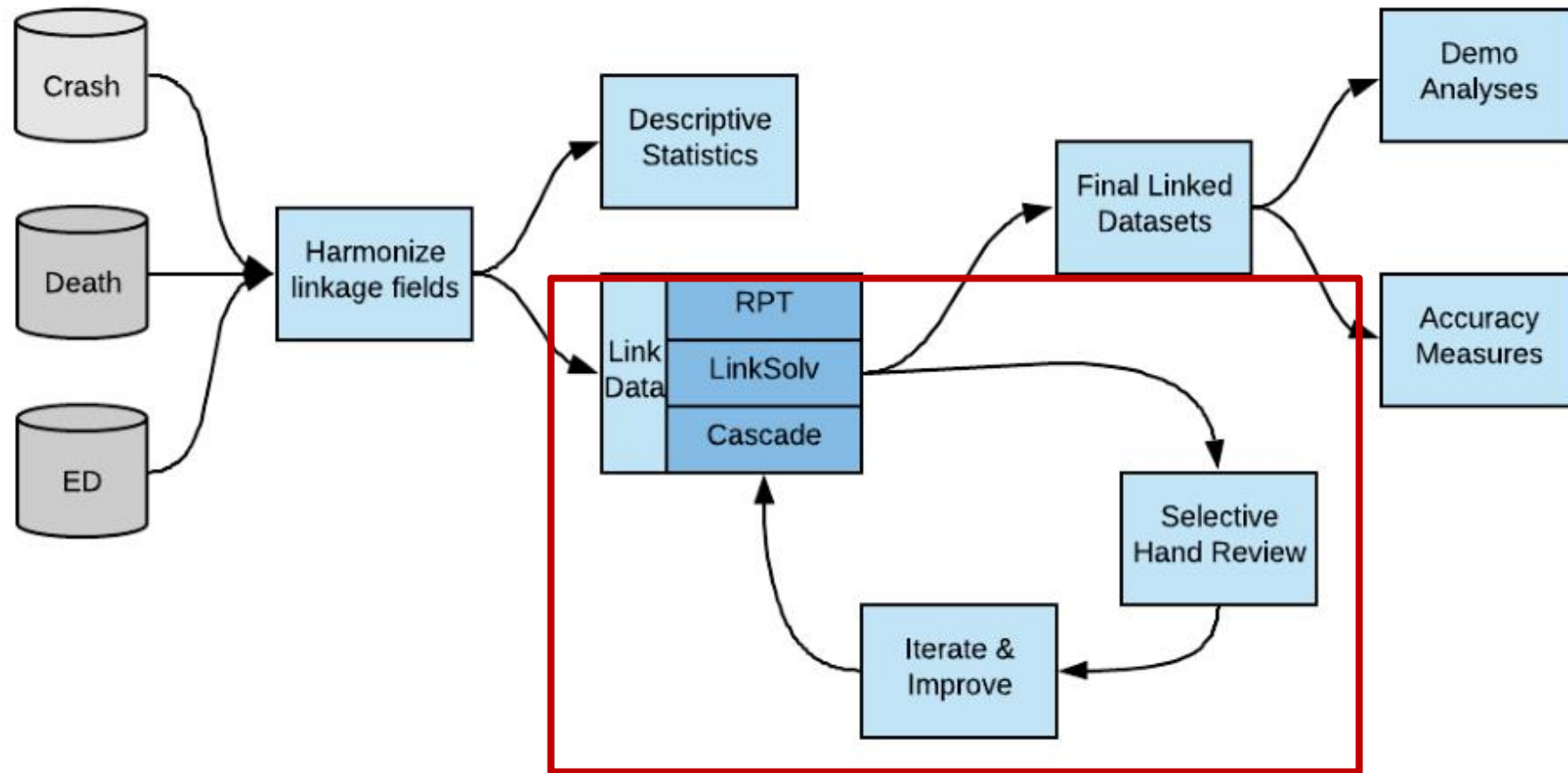
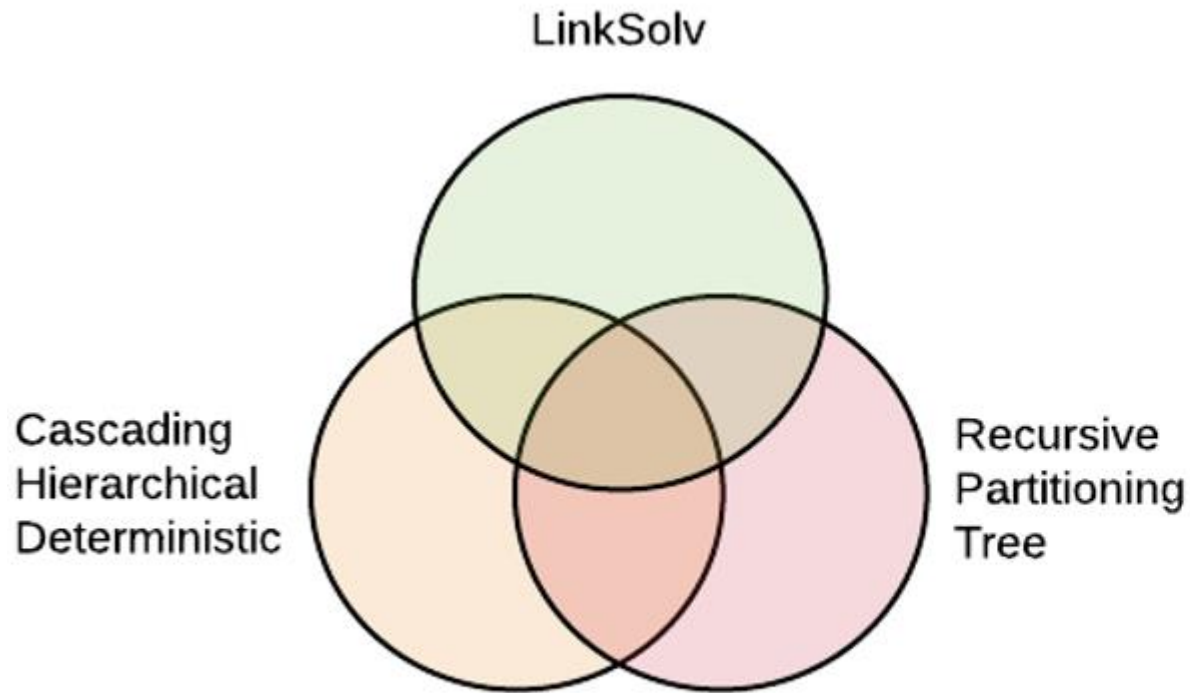


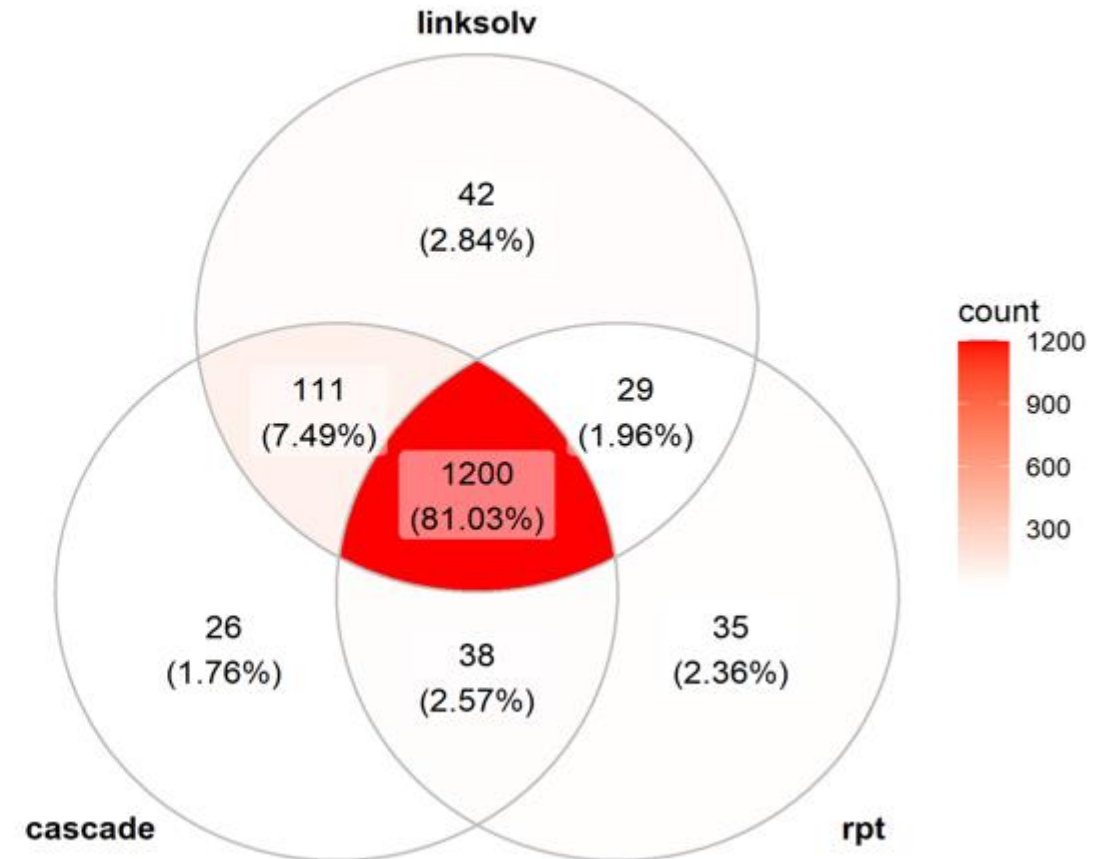
Diagram by Mike Fliss, PhD, MPS, MSW



# We examined the differences between methodologies.



Crash-Death Pair ID Venn



Not to scale; diagrams by Mike Fliss, PhD, MPS, MSW

The linkage 'instructions' are stored separately from the coding for the linkage.

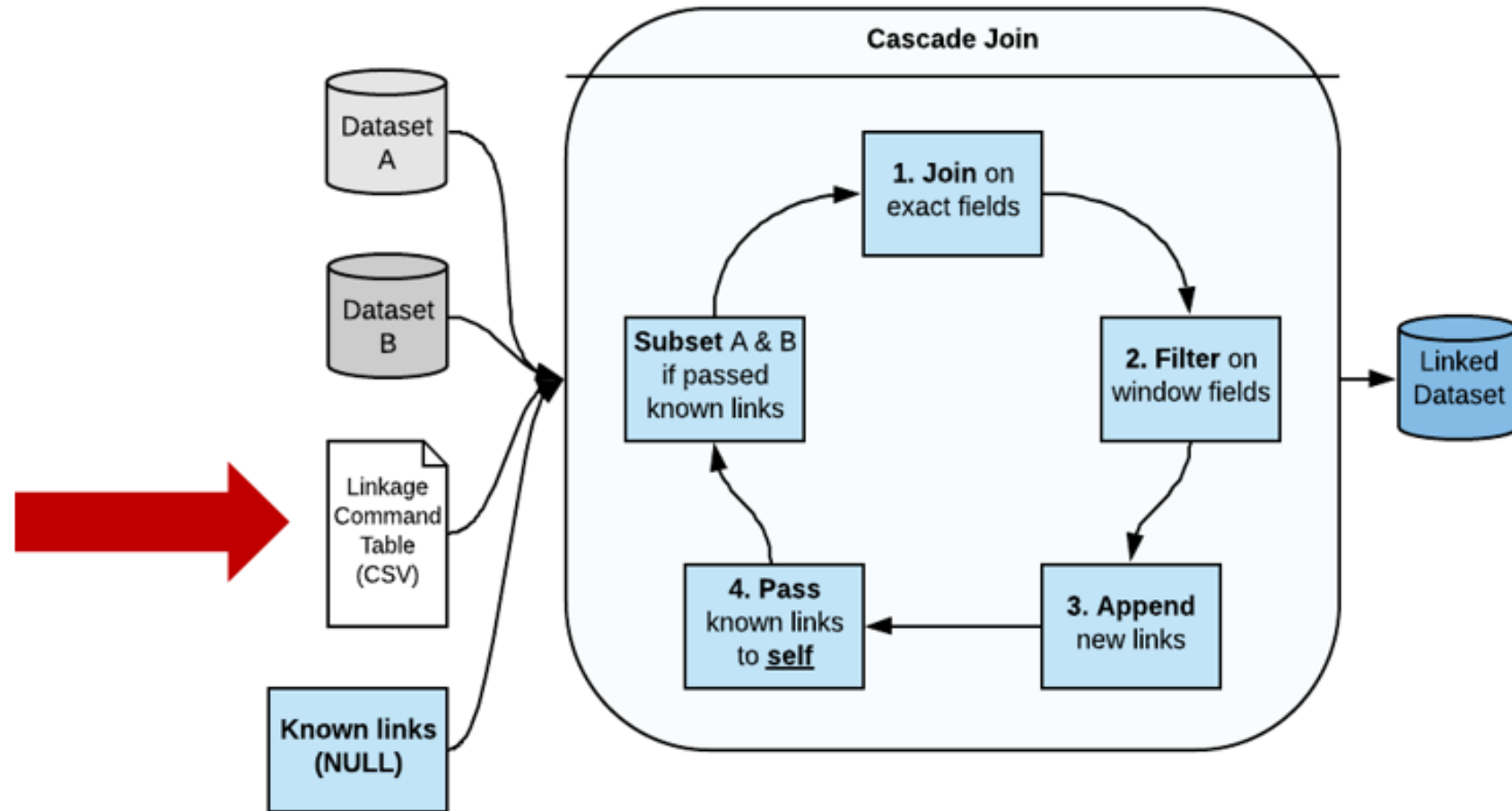


Diagram by Mike Fliss, PhD, MPS, MSW

# The linkage command file lists the exact linkage requirements for each deterministic linkage pass.

link_id	link_name	l_age_num	l_dob_date	l_dobmd_fct	l_gender_fct	l_raceeth_fct	l_fatal_lgl	l_isinj_lgl	l_issevere_lgl	l_iscrash_lgl	l_crashpos_fct	l_rstate_fct	l_rcounty_fct	l_rzip5_fct	l_rzip3_fct	l_rcity_fct	l_estate_fct	l_ccounty_fct	l_acc_date	l_accmd_fct
1	Exact: completely matching	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	Place-2: Drop R city, R zip	0	0	0	0	0	0	0	0	0	0	0	0	NA	NA	NA	0	0	0	0
3	Place-2: Drop R city, C county, R county	0	0	0	0	0	0	0	0	0	0	0	NA	0	0	0	0	NA	0	0
4	Demo-1: drop race	0	0	0	0	NA	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	Crash-1: crash pos	0	0	0	0	0	0	0	0	0	NA	0	0	0	0	0	0	0	0	0
6	Crash-2: crash & crash pos	0	0	0	0	0	0	0	0	NA	NA	0	0	0	0	0	0	0	0	0
7	Date-1: Day 30 away	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	30	NA

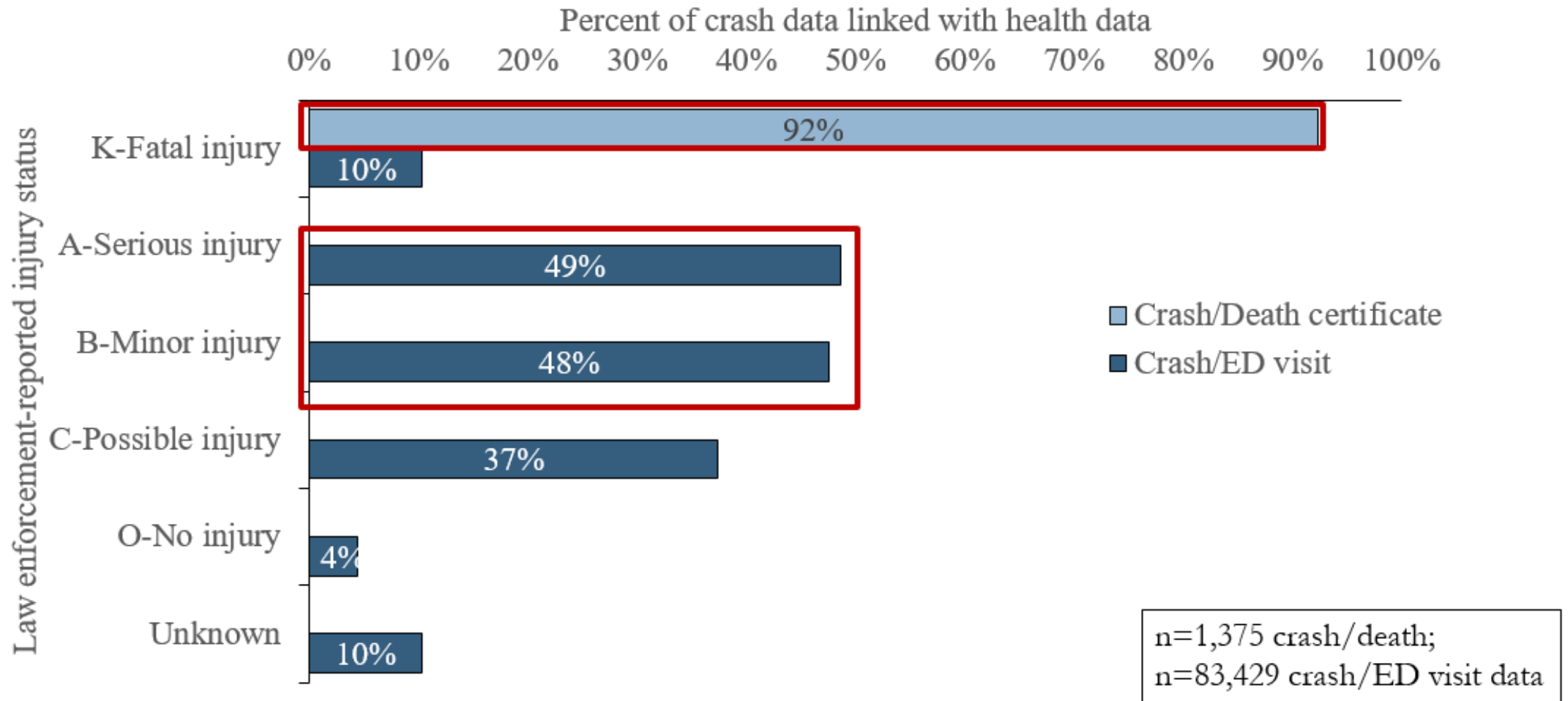
Green: Exact match

Red: Not used

Yellow: range of values allowed

Example linkage table by Mike Fliss, PhD, MPS, MSW

# We successfully linked both health outcome datasets using deterministic linkage.



# Other project accomplishments

## Presentations

- Traffic Records Forum 2020
- Transportation Research Board 2020
- Safe States 2020
- Council of State and Territorial Epidemiologists 2020

## Relationship building

- Dr. Larry Cook and others doing data linkage
- Other SVIPP MVC states
- CDC contacts
- Other NC stakeholders

## Documentation

- Data documentation for the three data sources
- Implementation plan
- Mid-year report
- Report outlining project barriers and facilitators
- Final report and handout ←

<http://cchi.web.unc.edu/transportation-health-data/>

# Challenges

- Imperfect source data, including limited unique identifiers (also a strength!)
- Limited time
- Lack of program ownership and stable funding

## Lessons learned: high value activities

- Identifying and cleaning linkage variables
- Comparing and verifying results across and between linkage methodologies
- Consulting with others doing data linkage
- Creating thorough documentation

# Thank You!

---

MIKE DOLAN FLISS (CO-INVESTIGATOR): [MIKE.DOLAN.FLISS@UNC.EDU](mailto:MIKE.DOLAN.FLISS@UNC.EDU)

KATIE HARMON (CO-INVESTIGATOR): [HARMON@HSRC.UNC.EDU](mailto:HARMON@HSRC.UNC.EDU)

KATHY PETICOLAS (PROGRAM MANAGER): [KATHY\\_PETICOLAS@MED.UNC.EDU](mailto:KATHY_PETICOLAS@MED.UNC.EDU)

ANNA WALLER (PI): [ANNA\\_WALLER@MED.UNC.EDU](mailto:ANNA_WALLER@MED.UNC.EDU)