Collaborative Sciences Center for ROAD SAFETY

www.roadsafety.unc.edu

Building a MVC injury system of linked data: Lessons learned & questions answered about pedestrian injuries

January 26, 2022

Introductions

Dr. Anna E. Waller, ScD

Dr. Katherine (Katie) J. Harmon, PhD

Acknowledgments

Funding:

Funding was provided by the Centers for Disease Control and Prevention, the NC Governor's Highway Safety Program, Collaborative Sciences Center for Road Safety, and NC Safe Routes to School.

Data attribution & disclaimer:

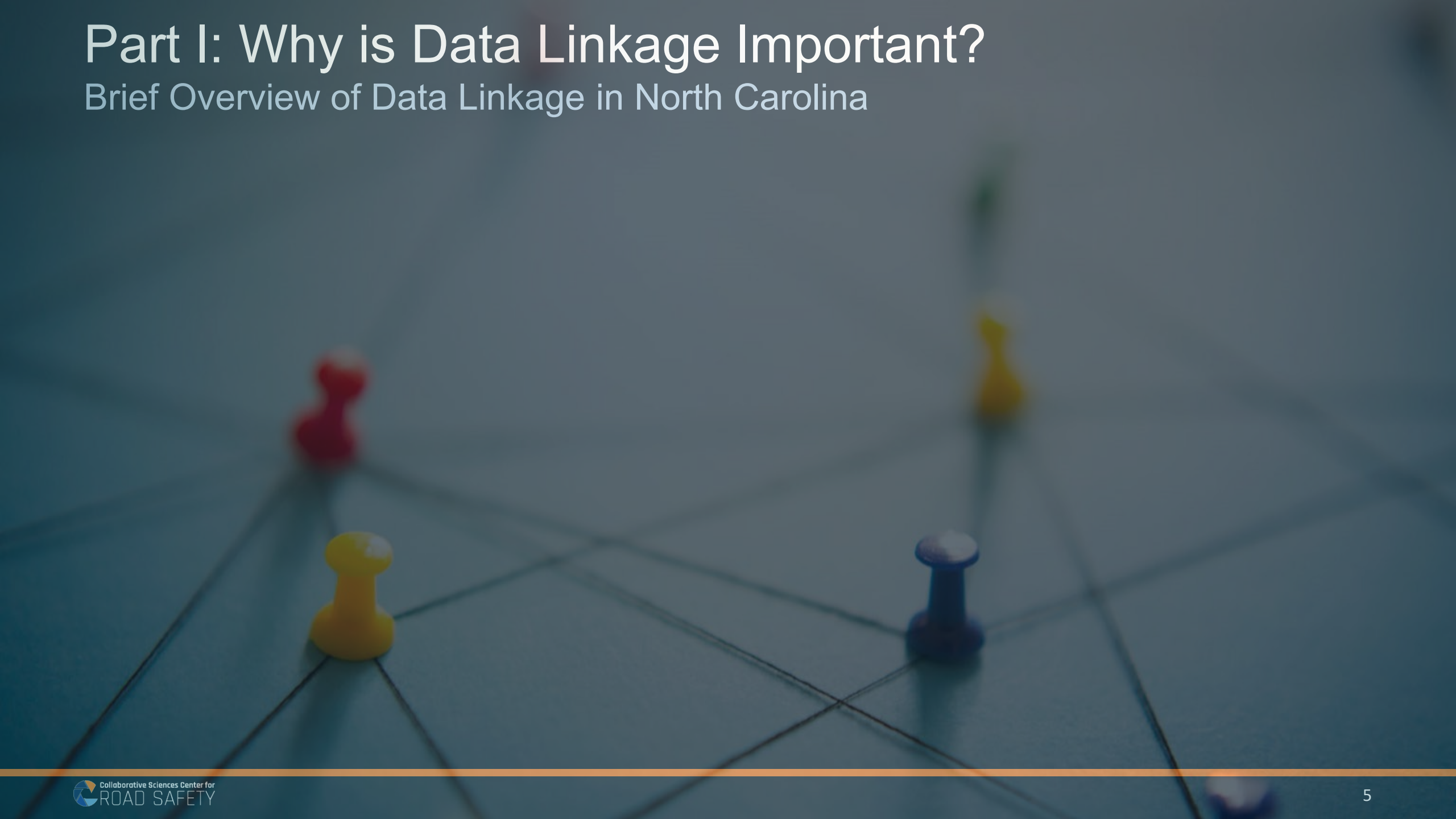
NC DETECT is a statewide public health syndromic surveillance system, funded by the NC Division of Public Health (NC DPH) Federal Public Health Emergency Preparedness Grant and managed through collaboration between NC DPH and UNC-CH Department of Emergency Medicine's Carolina Center for Health Informatics.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

Outline

- Part I: Why Is Data Linkage Important?
  - Overview of Data Linkage in North Carolina
  - What Did We Learn from Data Linkage?
    - Building Relationships Across Organizations
    - Data Quality
    - Methodology
    - NC Transportation Safety & Public Health Data Dashboard
    - Selected Results from NC Pedestrian and Bicycle Crash Injury Data Linkage Study
- Part II: Challenges to Sustainable Data Linkage
  - Obtaining/Maintaining Stakeholder Support
  - Obtaining Funding
  - Technical Issues

Part I: Why is Data Linkage Important?

Brief Overview of Data Linkage in North Carolina



NC Data Linkage Objectives

Goal: Establish an integrated statewide MVC injury surveillance system

Integrated MVC and Health Information has the potential to:

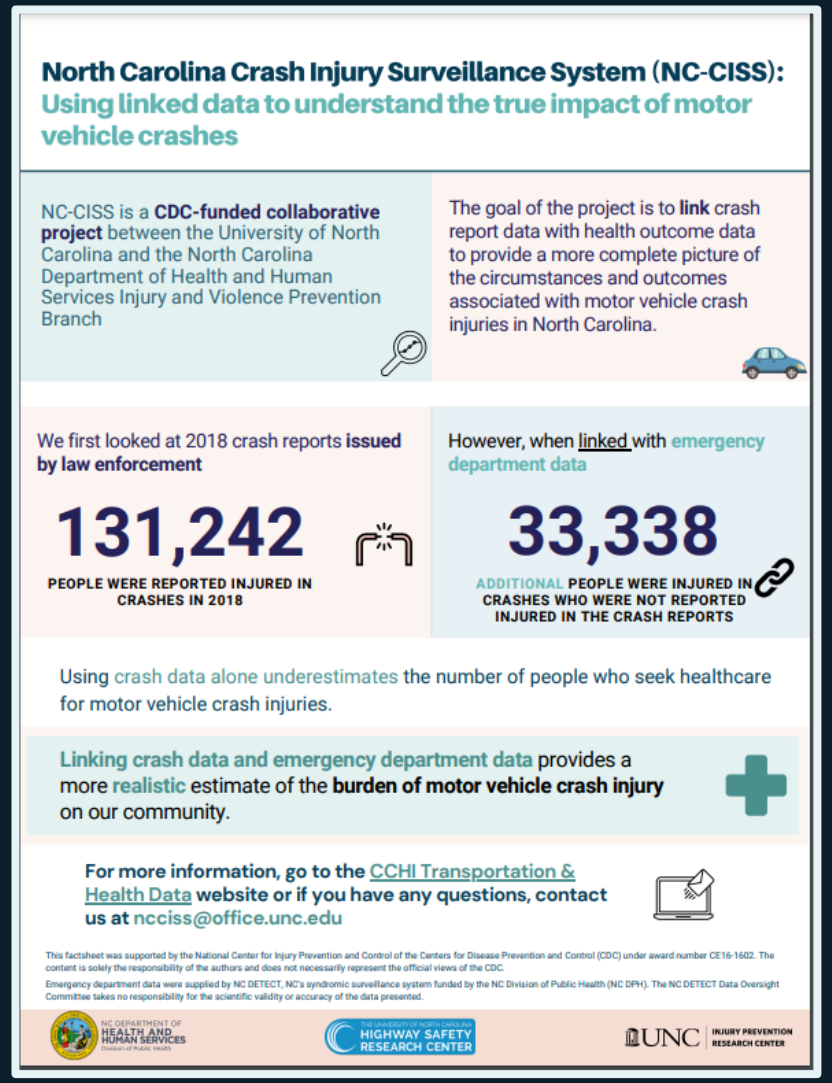
- Improve safety outcome analysis and evaluation,
- Expand research activities,
- And inform policy and safety programs.

Why Link Crash Data with Other Data Sources?

Most data sources are limited in scope; by linking multiple data sources, we create a much richer dataset that can then be used to answer important questions.

Part I: Why is Data Linkage Important?

What Did We Learn from Data Linkage?



North Carolina Crash Injury Surveillance System (NC-CISS):
Using linked data to understand the true impact of motor vehicle crashes

NC-CISS is a **CDC-funded collaborative project** between the University of North Carolina and the North Carolina Department of Health and Human Services Injury and Violence Prevention Branch

The goal of the project is to **link** crash report data with health outcome data to provide a more complete picture of the circumstances and outcomes associated with motor vehicle crash injuries in North Carolina.

We first looked at 2018 crash reports **issued by law enforcement**

131,242
PEOPLE WERE REPORTED INJURED IN CRASHES IN 2018

However, when linked with emergency department data

33,338
ADDITIONAL PEOPLE WERE INJURED IN CRASHES WHO WERE NOT REPORTED INJURED IN THE CRASH REPORTS

Using crash data alone underestimates the number of people who seek healthcare for motor vehicle crash injuries.

Linking crash data and emergency department data provides a more realistic estimate of the **burden of motor vehicle crash injury** on our community.

For more information, go to the CCHI Transportation & Health Data website or if you have any questions, contact us at ncciss@office.unc.edu

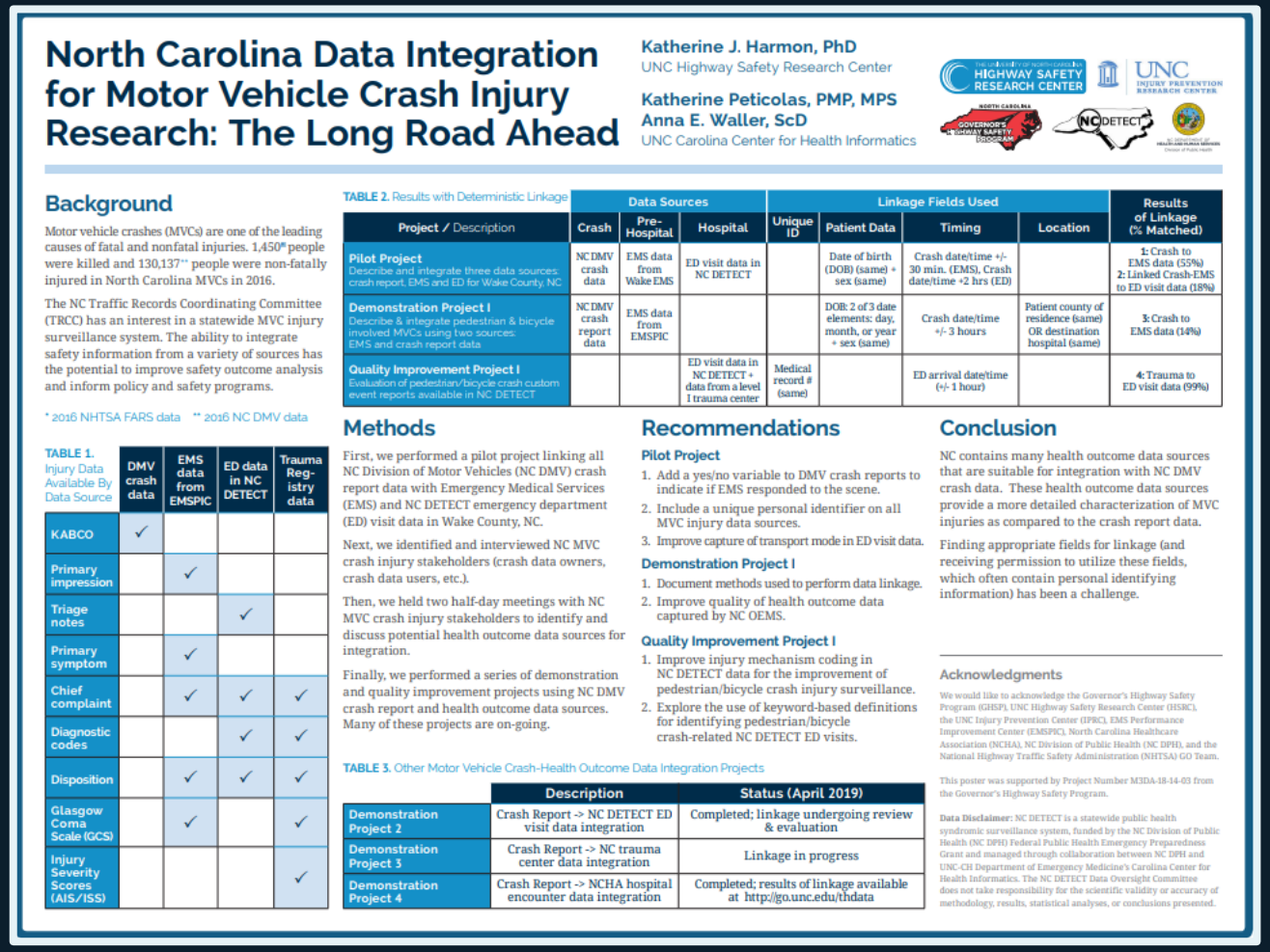
This factsheet was supported by the National Center for Injury Prevention and Control of the Centers for Disease Prevention and Control (CDC) under award number CE16-1602. The content is solely the responsibility of the authors and does not necessarily represent the official views of the CDC.

Emergency department data were supplied by NC DETECT, NC's syndromic surveillance system funded by the NC Division of Public Health (NC DPH). The NC DETECT Data Oversight Committee takes no responsibility for the scientific validity or accuracy of the data presented.

NC DEPARTMENT OF HEALTH AND HUMAN SERVICES · HIGHWAY SAFETY RESEARCH CENTER · UNC INJURY PREVENTION RESEARCH CENTER

TABLE 1. Injury Data Available By Data Source

	DMV crash data	EMS data from EMSPIC	ED data in NC DETECT	Trauma Registry data
KABCO	✓			
Primary impression		✓		
Triage notes			✓	
Primary symptom		✓		
Chief complaint		✓	✓	✓
Diagnostic codes			✓	✓
Disposition		✓	✓	✓
Glasgow Coma Scale (GCS)		✓		✓
Injury Severity Scores (AIS/ISS)				✓



North Carolina Data Integration for Motor Vehicle Crash Injury Research: The Long Road Ahead

Katherine J. Harmon, PhD
UNC Highway Safety Research Center
Katherine Peticolas, PMP, MPS
Anna E. Waller, ScD
UNC Carolina Center for Health Informatics

Background

Motor vehicle crashes (MVCs) are one of the leading causes of fatal and nonfatal injuries. 1,450* people were killed and 130,137** people were non-fatally injured in North Carolina MVCs in 2016.

The NC Traffic Records Coordinating Committee (TRCC) has an interest in a statewide MVC injury surveillance system. The ability to integrate safety information from a variety of sources has the potential to improve safety outcome analysis and inform policy and safety programs.

* 2016 NHTSA FARS data ** 2016 NC DMV data

TABLE 2. Results with Deterministic Linkage

Project / Description	Data Sources: Crash	Data Sources: Pre-Hospital	Data Sources: Hospital	Linkage Fields Used: Unique ID	Linkage Fields Used: Patient Data	Linkage Fields Used: Timing	Linkage Fields Used: Location	Results of Linkage (% Matched)
Pilot Project Describe and integrate three data sources: crash report, EMS and ED for Wake County, NC	NC DMV crash data	EMS data from Wake EMS	ED visit data in NC DETECT		Date of birth (DOB) (same) + sex (same)	Crash date/time +/- 30 min. (EMS), Crash date/time +2 hrs (ED)		1: Crash to EMS data (55%) 2: Linked Crash-EMS to ED visit data (18%)
Demonstration Project I Describe & integrate pedestrian & bicycle involved MVCs using two sources: EMS and crash report data	NC DMV crash report data	EMS data from EMSPIC			DOB: 2 of 3 date elements: day, month, or year + sex (same)	Crash date/time +/- 3 hours	Patient county of residence (same) OR destination hospital (same)	3: Crash to EMS data (14%)
Quality Improvement Project I Evaluation of pedestrian/bicycle crash custom event reports available in NC DETECT			ED visit data in NC DETECT + data from a level 1 trauma center	Medical record # (same)		ED arrival date/time (+/- 1 hour)		4: Trauma to ED visit data (99%)

Methods

First, we performed a pilot project linking all NC Division of Motor Vehicles (NC DMV) crash report data with Emergency Medical Services (EMS) and NC DETECT emergency department (ED) visit data in Wake County, NC.

Next, we identified and interviewed NC MVC crash injury stakeholders (crash data owners, crash data users, etc.).

Then, we held two half-day meetings with NC MVC crash injury stakeholders to identify and discuss potential health outcome data sources for integration.

Finally, we performed a series of demonstration and quality improvement projects using NC DMV crash report and health outcome data sources. Many of these projects are on-going.

Recommendations

Pilot Project
1. Add a yes/no variable to DMV crash reports to indicate if EMS responded to the scene.
2. Include a unique personal identifier on all MVC injury data sources.
3. Improve capture of transport mode in ED visit data.

Demonstration Project I
1. Document methods used to perform data linkage.
2. Improve quality of health outcome data captured by NC OEMS.

Quality Improvement Project I
1. Improve injury mechanism coding in NC DETECT data for the improvement of pedestrian/bicycle crash injury surveillance.
2. Explore the use of keyword-based definitions for identifying pedestrian/bicycle crash-related NC DETECT ED visits.

Conclusion

NC contains many health outcome data sources that are suitable for integration with NC DMV crash data. These health outcome data sources provide a more detailed characterization of MVC injuries as compared to the crash report data.

Finding appropriate fields for linkage (and receiving permission to utilize these fields, which often contain personal identifying information) has been a challenge.

TABLE 3. Other Motor Vehicle Crash-Health Outcome Data Integration Projects

	Description	Status (April 2019)
Demonstration Project 2	Crash Report -> NC DETECT ED visit data integration	Completed; linkage undergoing review & evaluation
Demonstration Project 3	Crash Report -> NC trauma center data integration	Linkage in progress
Demonstration Project 4	Crash Report -> NCHA hospital encounter data integration	Completed; results of linkage available at http://go.unc.edu/thdata

Acknowledgments

We would like to acknowledge the Governor's Highway Safety Program (GHSP), UNC Highway Safety Research Center (HSRC), the UNC Injury Prevention Center (IPRC), EMS Performance Improvement Center (EMSPIC), North Carolina Healthcare Association (NCHA), NC Division of Public Health (NC DPH), and the National Highway Traffic Safety Administration (NHTSA) GO Team.

This poster was supported by Project Number M3DA-18-14-03 from the Governor's Highway Safety Program.

Data 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.

https://cchi.web.unc.edu/transportation-health-data/

Collaboration Is Essential

- Project Staff
  - Investigators
  - Program managers
  - Statisticians
- Data owners
- Data users
  - State/Local departments of transportation
  - State/Local health departments
  - Investigators
- Community and advocacy groups
- Funders

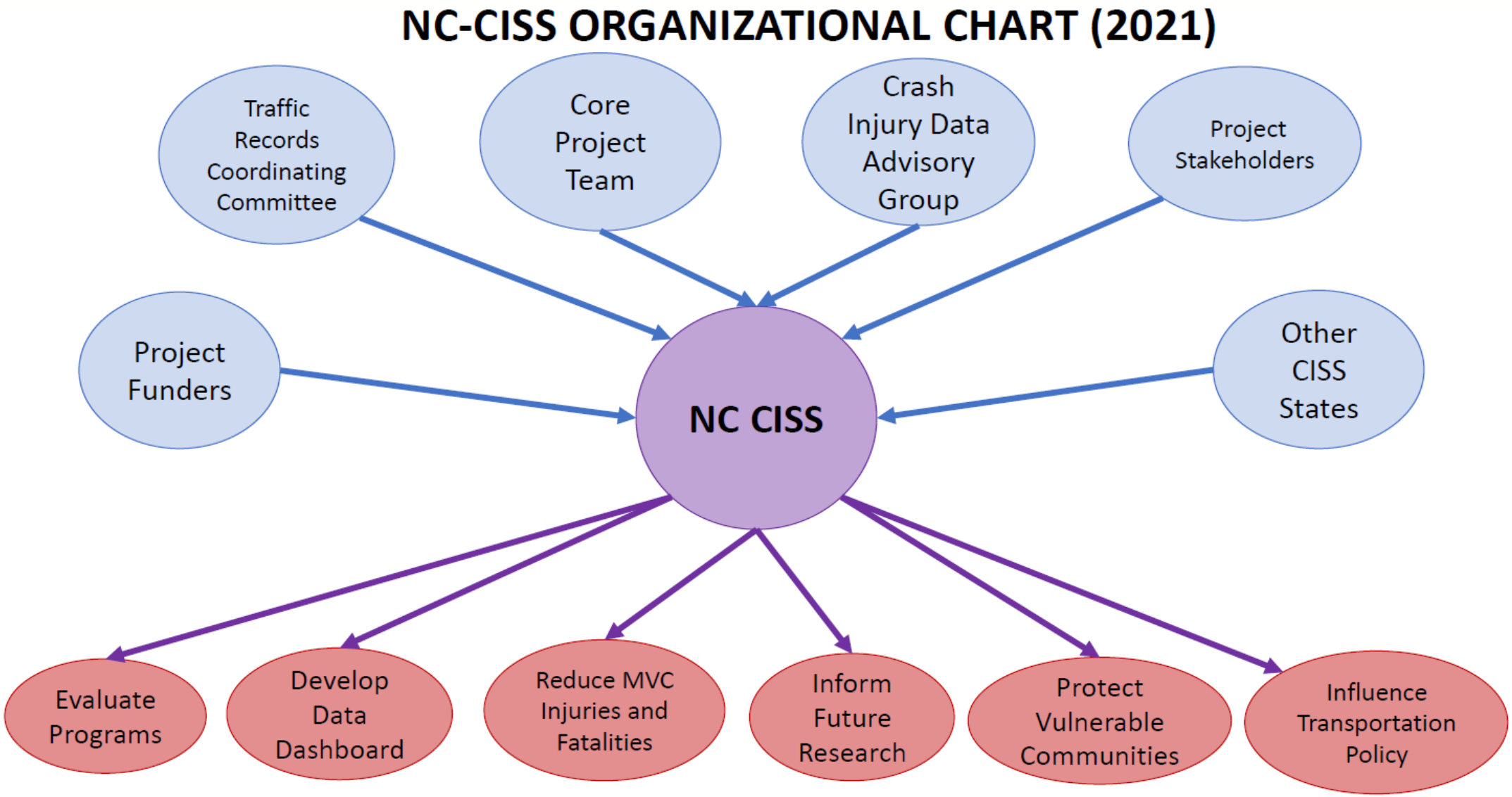
Centers for Disease Control and Prevention

North Carolina Department of Health and Human Services
- Injury and Violence Prevention Branch
- State Center for Health Statistics
- Communicable Disease Branch

North Carolina Department of Transportation
- Governor's Highway Safety Program
- Traffic Records Coordinating Committee

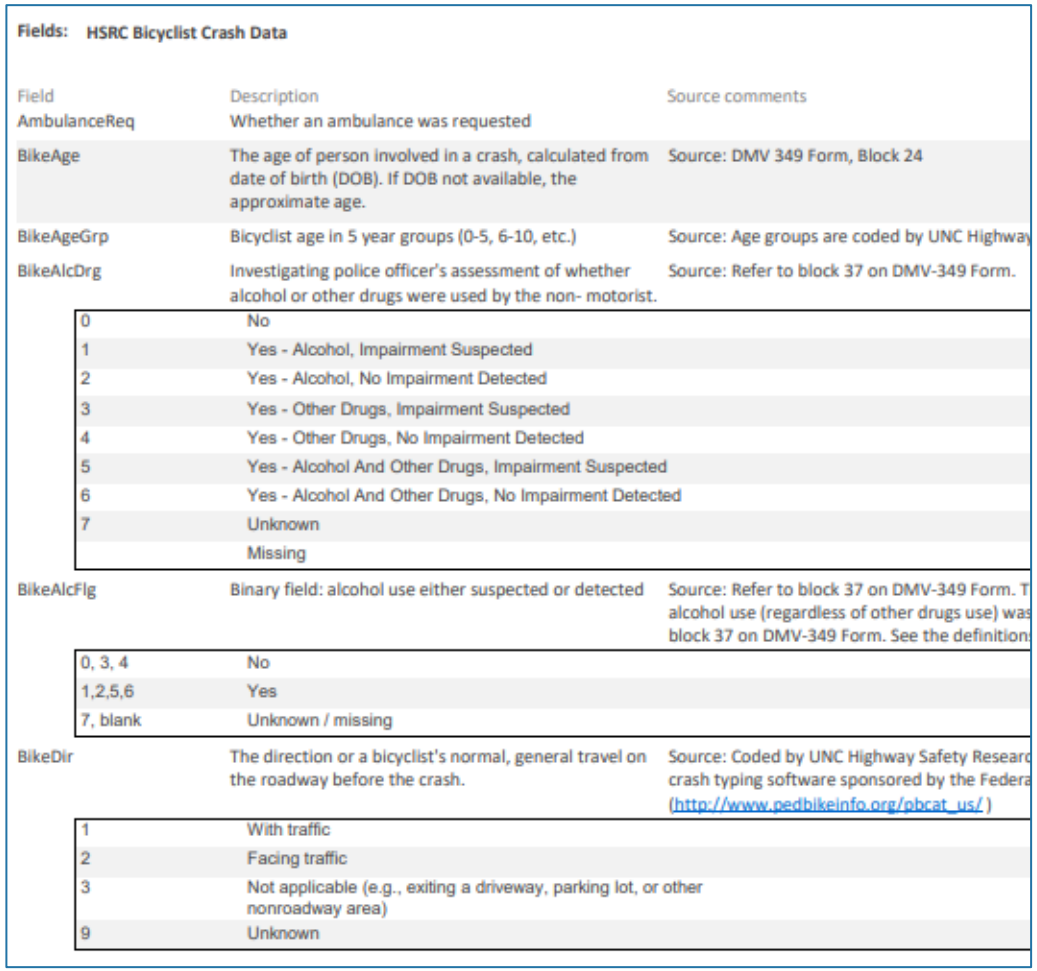
University of North Carolina – Chapel Hill
- Carolina Center for Health Informatics
- Highway Safety Research Center
- Injury Prevention Research Center

Building Relationships Across Organizations



Data Quality – MVC Data Sources Identified & Documented

- Crash Data (NC DMV)
- Pedestrian/Bicycle Crash Data (HSRC/NCDOT)
- EMS data (NC OEMS)
- Emergency Department Data (NC DETECT)
- Hospital Discharge Data (SCHS)
- Hospital Claims Data (UNC Sheps)
- North Carolina Trauma Registry (NCTR)
- BCBS/Medicaid Claims Data (UNC Sheps)
- Death Registration Data (SCHS)
- Medical Examiners Reports (OCME)
- Fatality Analysis Reporting System (NHTSA)
- Highway Safety Information System (FHWA)



Fields: HSRC Bicyclist Crash Data

Field	Description	Source comments
AmbulanceReq	Whether an ambulance was requested	
BikeAge	The age of person involved in a crash, calculated from date of birth (DOB). If DOB not available, the approximate age.	Source: DMV 349 Form, Block 24
BikeAgeGrp	Bicyclist age in 5 year groups (0-5, 6-10, etc.)	Source: Age groups are coded by UNC Highway
BikeAlcDrg	Investigating police officer's assessment of whether alcohol or other drugs were used by the non- motorist.	Source: Refer to block 37 on DMV-349 Form.

0	No
1	Yes - Alcohol, Impairment Suspected
2	Yes - Alcohol, No Impairment Detected
3	Yes - Other Drugs, Impairment Suspected
4	Yes - Other Drugs, No Impairment Detected
5	Yes - Alcohol And Other Drugs, Impairment Suspected
6	Yes - Alcohol And Other Drugs, No Impairment Detected
7	Unknown
	Missing

BikeAlcFlg	Binary field: alcohol use either suspected or detected	Source: Refer to block 37 on DMV-349 Form. T alcohol use (regardless of other drugs use) was block 37 on DMV-349 Form. See the definition

0, 3, 4	No
1,2,5,6	Yes
7, blank	Unknown / missing

BikeDir	The direction or a bicyclist's normal, general travel on the roadway before the crash.	Source: Coded by UNC Highway Safety Researc crash typing software sponsored by the Federa (http://www.pedbikeinfo.org/pbcat_us/)

1	With traffic
2	Facing traffic
3	Not applicable (e.g., exiting a driveway, parking lot, or other nonroadway area)
9	Unknown

https://cchi.web.unc.edu/data-sources-for-motor-vehicle-crash-injury-in-north-carolina

Linkage Methods

We investigated four different linkage methodologies but focused on deterministic linkage.

Linkage methods	Description
Hierarchical deterministic linkage w/ fuzzy matching	Matches records using a set of pre-defined shared identifiers over multiple passes or "cascades"; allows some flexing with matching variables (age +/- 1 year)
Recursive partitioning trees	Matches records using a calculated 'distance' between linkage variables
Probabilistic linkage	Matches records based on a pre-assigned probability that the match is correct (e.g. Linksolv)
Hand review	Matches records through manual review

Strengths:

- Easy to explain to a multi-disciplinary audience,
- High quality results,
- Fast,
- And replicable in many applications.

Challenge:

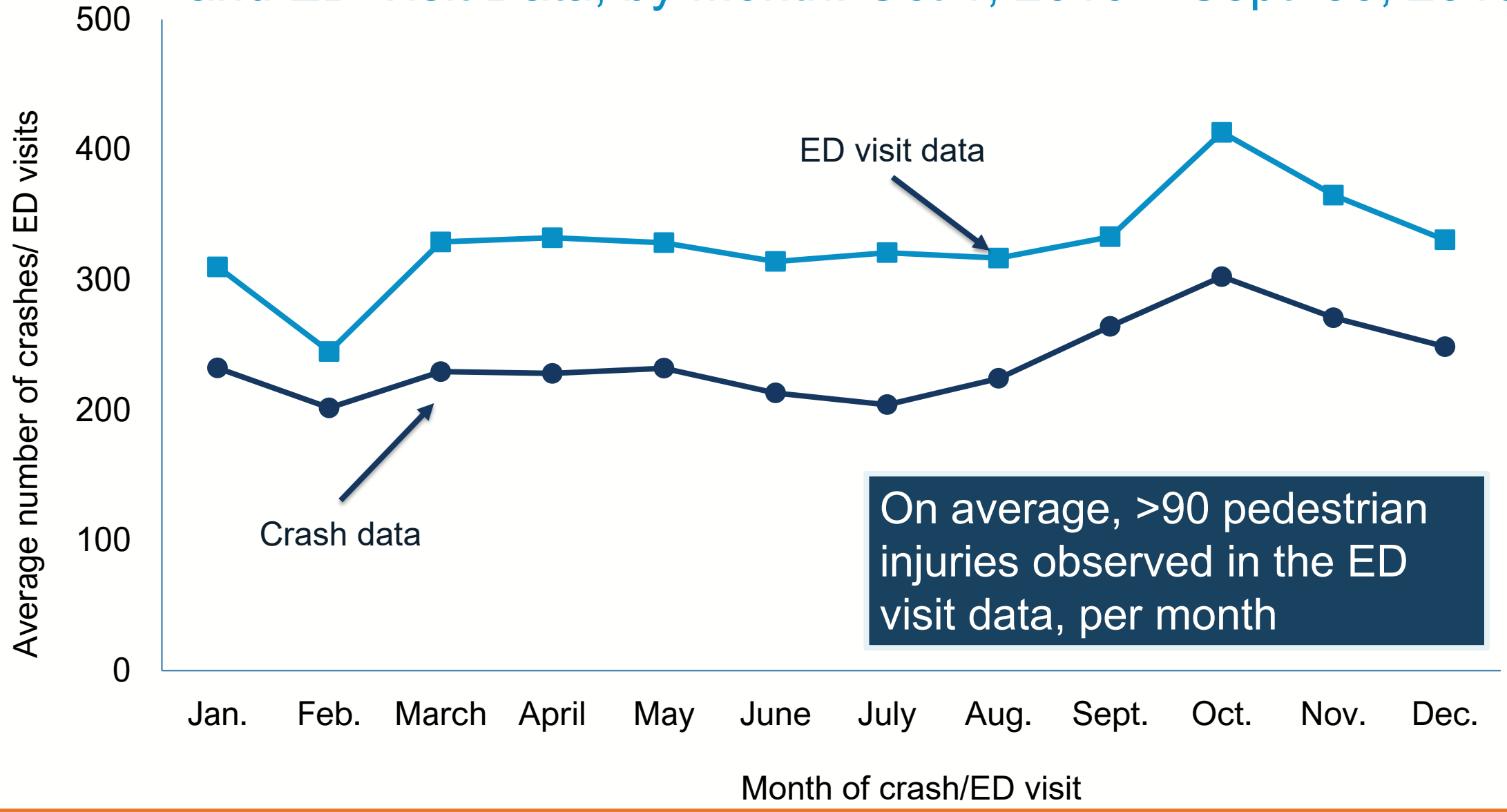
A sufficient & representative match rate.

https://cchi.web.unc.edu/wp-content/uploads/sites/2506/2021/03/NC-CISSFinalReport_20210316.pdf

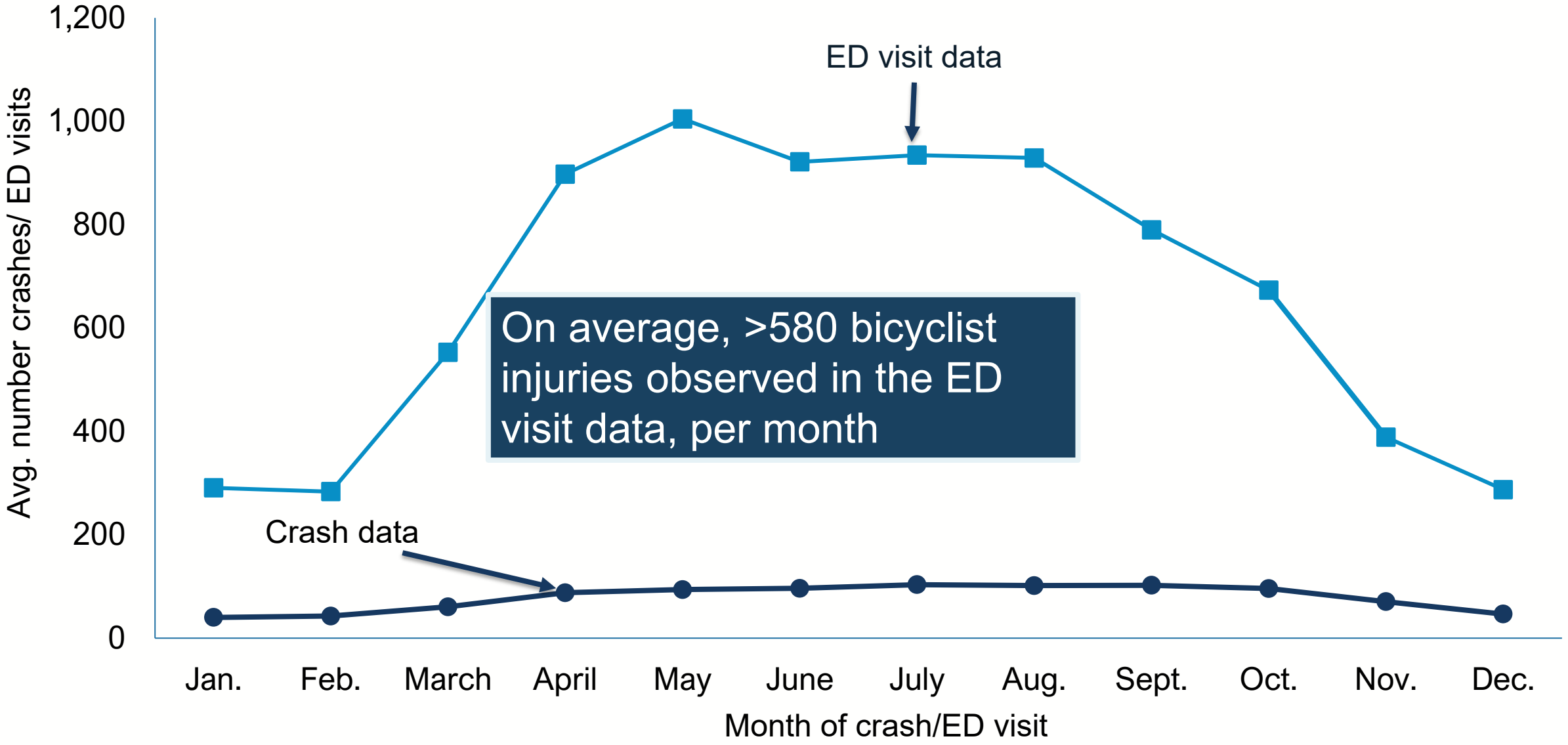
NC Transportation Safety & Public Health Data Dashboard

https://cchi.web.unc.edu/nc-transportation-safety-public-health-data-dashboard/

Average Number of Pedestrian Injuries Reported in NC crash and ED Visit Data, by Month: Oct 1, 2010 – Sept. 30, 2015



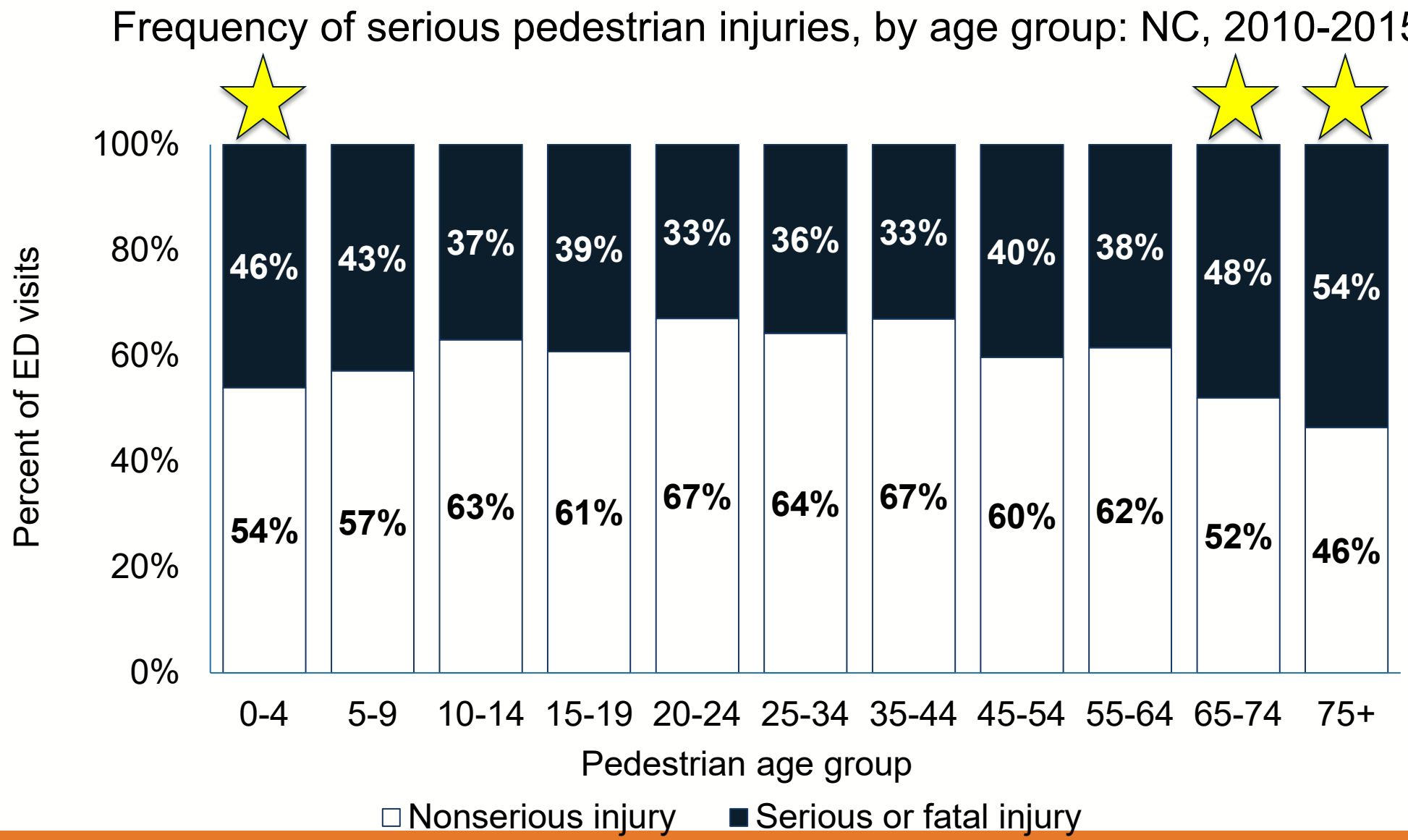
Average Number of Bicyclist Injuries Reported in NC crash and ED Visit Data, by Month: Oct 1, 2010 – Sept. 30, 2015



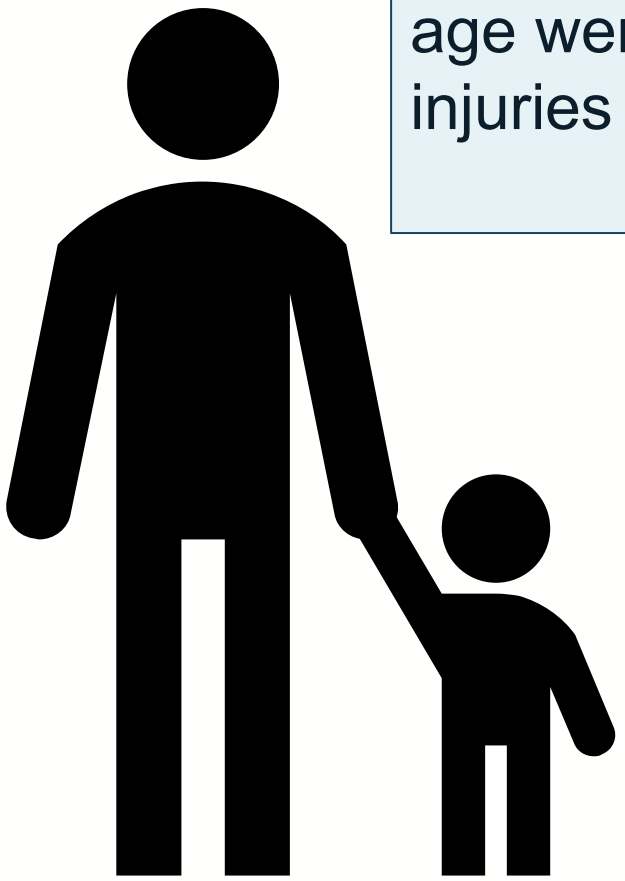
KABCO Does Not Always Provide an Accurate Assessment of Pedestrian Injury Severity

Police assigned injury severity (KABCO)	Serious or fatal injury (based on clinical assessment) N (%)	Non-serious injury (based on clinical assessment) N (%)
K: Killed	206 (100%)	0 (0%)
A: Disabling injury	437 (89%)	53 (11%)
B: Evident injury	1,431 (50%)	1,440 (50%)
C: Possible injury	488 (16%)	2,523 (84%)
O: No injury	20 (12%)	141 (88%)
Total	**2,582 (38%)**	**4,157 (62%)**

Pedestrian Injury Severity was Highest for Children and Older Adults



One Possible Explanation for the Higher Injury Severity Observed among NC Child Pedestrians was the Frequency of Head Injuries



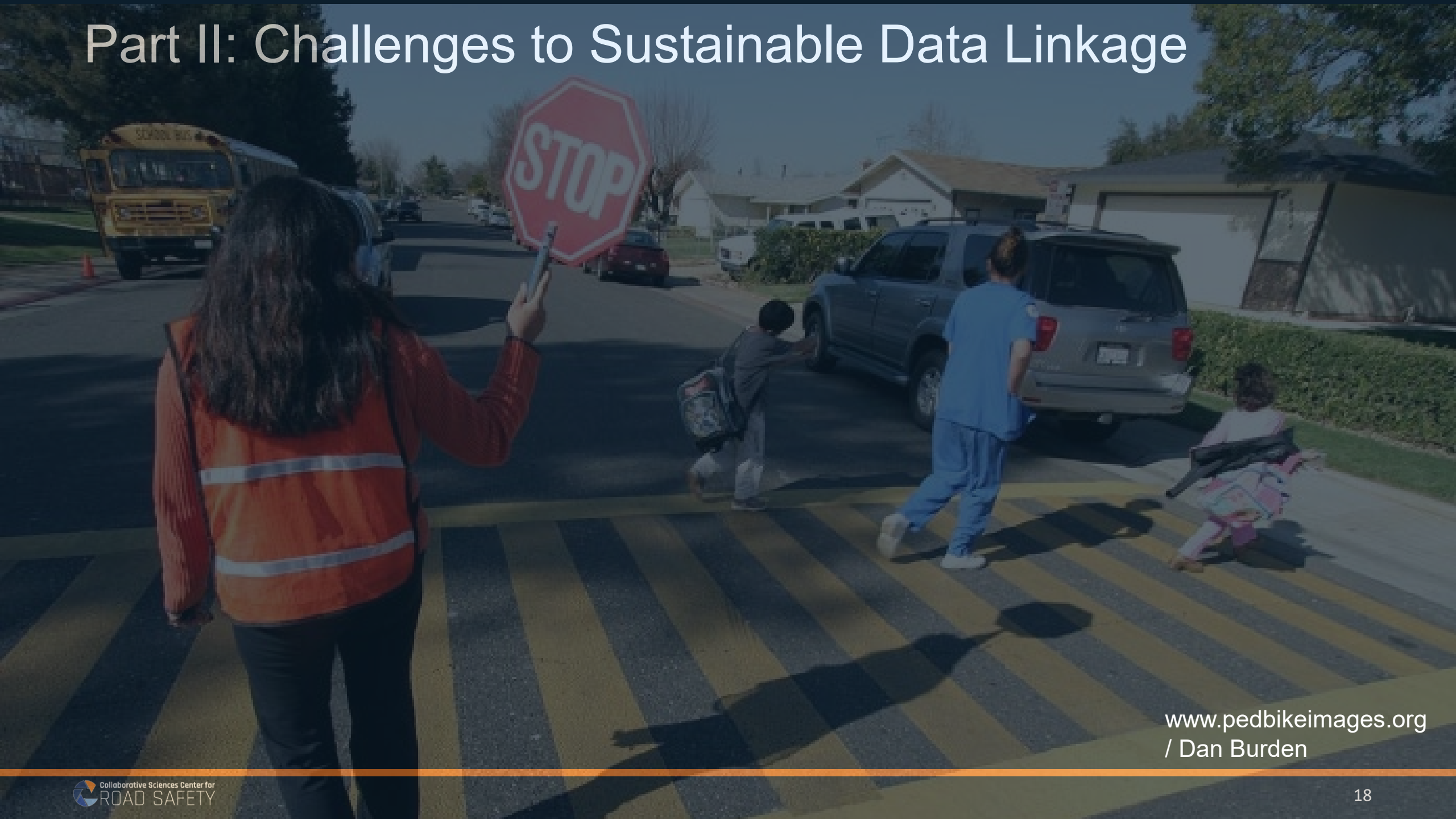
24% of adults 25-44 years of age were diagnosed with head injuries

39% of children 0-14 years of age were diagnosed with head injuries

Male pedestrians (31%) and pedestrians struck at estimated speeds **>35 MPH (40%)** were also more likely to be diagnosed with head injuries, as compared to females (24%) and pedestrians struck at speeds ≤35 MPH (25%)

https://www.roadsafety.unc.edu/research/projects/2019r22/

Part II: Challenges to Sustainable Data Linkage



www.pedbikeimages.org / Dan Burden

Challenges – Stakeholder Support

- ***Obtaining stakeholder support*** is rewarding but not always successful.
  - Accept that you will win some battles and lose others.
- ***Maintaining stakeholder engagement*** can also be challenging.
- While having a broad base of support has clear strengths, when you have many stakeholders, there may be a reluctance among stakeholders to feel ***ownership*** of the system.

Challenges - Funding

- Despite efforts to build a sustainable data linkage system, we have been ***unable to secure continued funding***.
- Large data linkage systems are not conducive to annual, low-budget, proof-of-concept proposals. Support is needed for:
  - Staff with expertise in transportation safety, public health, statistics, and program management;
  - System maintenance and ongoing linkage efforts;
  - Computers and software;
  - The examination of specific research question,;
  - And the dissemination of data and research products.
- To date, most states have been ***unable (or unwilling) to pay*** for ongoing, sustainable data linkage.

Challenges - Technical

- ***Data acquisition*** is a time-consuming process (weeks to months).
- Cleaning and linking large datasets is ***methodologically complex***.
  - Requires considerable technical skills.
  - Linking more than one year of data can take hours.
  - A single change to a dataset structure can result in onerous changes to linkage methodology.
- Linking to ***protected health information*** impacts the analysis and dissemination of data.
  - Health data are typically aggregated to protect patients' anonymity.

Despite Challenges, Data Linkage is Worth It

- Data linkage provides an opportunity to:
  - Build relationships;
  - Improve data quality;
  - Develop methodology and technical capacity;
  - Characterize the burden of pedestrian/bicycle crash injury across multiple data sources;
  - Evaluate pedestrian/bicycle safety decisions, programs/interventions, and policies;
  - Educate decision-makers and the public;
  - And address equity issues around safety and access.

Questions?

Contact information:

Anna Waller, ScD

Research Professor

Director, Carolina Center for Health Informatics

Department of Emergency Medicine

awaller@med.unc.edu

Contact information:

Katie Harmon, PhD

Research Associate

UNC HSRC

919.962.0745

harmon@hsrc.unc.edu