Comparing CIREN and NASS Cases: A Multivariate Similarity Approach

Joel Stitzel, Ph. D.,
Patrick Kilgo, M. S.,
Brian Schmotzer, M. S.,
Clay Gabler, Ph. D.,
J. Wayne Meredith, M.D.

TOYOTA

Toyota – Wake Forest University
School of Medicine
CIREN Center

WAKE FOREST
UNIVERSITY
SCHOOL *of* MEDICINE
THE BOWMAN GRAY CAMPUS

Impetus: The Question

- CIREN - excellent source for detailed injury data
- NASS weighted sample of real world cases
- Question: How do we draw conclusions about injuries and safety using CIREN data, knowing that it accurately describes real world injuries? For:
  - Biomechanics Research
  - Safety Regulation
  - Automotive Industry

NASS/CDS vs. CIREN

- NASS/CDS
  - National Sample
  - Crash Data
  - Entry Criteria: Tow-away Crashes
  - 4000-5000 crashes/year
  - Large contingent of MAIS 1-2 injuries
  - Reliable Sampling Weights
- CIREN
  - Clinical Data
  - Crash Data
  - Entry Criteria: Level I Trauma Center
  - 300-400 crashes/year
  - Minimum MAIS3*
  - Detailed Injury Data
  - Exhaustive Outcomes

*CIREN inclusion criteria specify some exceptions to this rule.

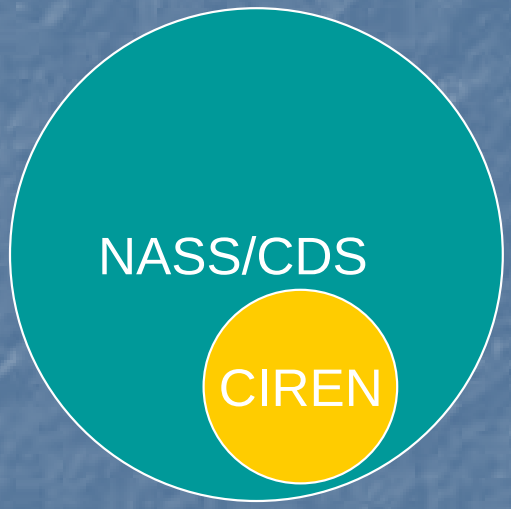
NASS/CDS vs. CIREN

- Why use NASS/CDS?
  - Allows National Estimates of Injury
  - Injury data not always complete or detailed (less clinical data)
  - Lots of data, sampling weights, well established, some injury data
- Why use CIREN?
  - Detailed Clinical Evaluation of Injuries
  - Cases reviewed by doctors, engineers, crash investigators, etc.
  - Not a population-based sample
  - Difficult to extrapolate to national outcome

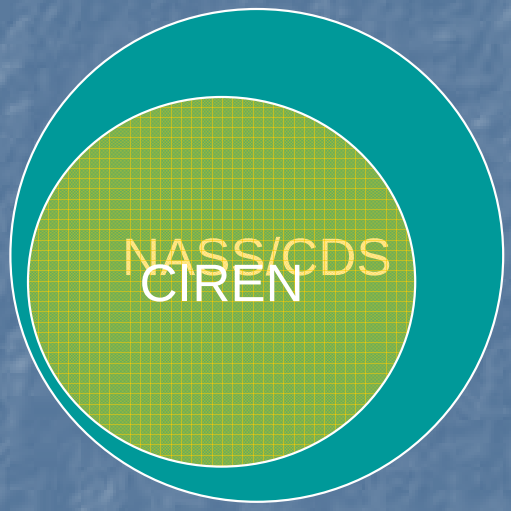
Objectives

- Quantify the similarity/difference between a given CIREN case (or a subset of CIREN cases) and the population of NASS cases
- Hypothetical questions:
  - Given a subset of NASS cases of interest:
    - How do we identify CIREN cases that “match it”?
    - What is the best way to make a comparison?
    - What variables in NASS or CIREN are important in describing the differences?

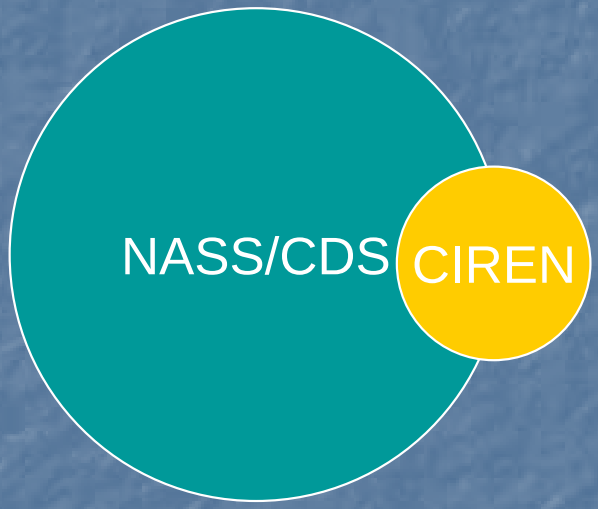
NASS/CDS vs. CIREN



Is CIREN a subset of NASS/CDS?



What is the nature and size of the overlap?



Does CIREN contain crash scenarios not in NASS/CDS?

Main Question

- How similar are individual CIREN cases to the average NASS case?
- By extension: how similar is a group of CIREN cases of interest, to a population of NASS cases of interest?

General Approach

- Compute a “similarity score” between CIREN cases and the average NASS case
- That is, the k-dimensional “distance” between the case types
  - Where k=number of variables common between the data sources

Methods – Basics of Mahalanobis Distance

- A multivariate measure of distance
- It puts all variables on the same scale – standardizes them
- Takes into account the correlation between the k variables

Mahalanobis Distance

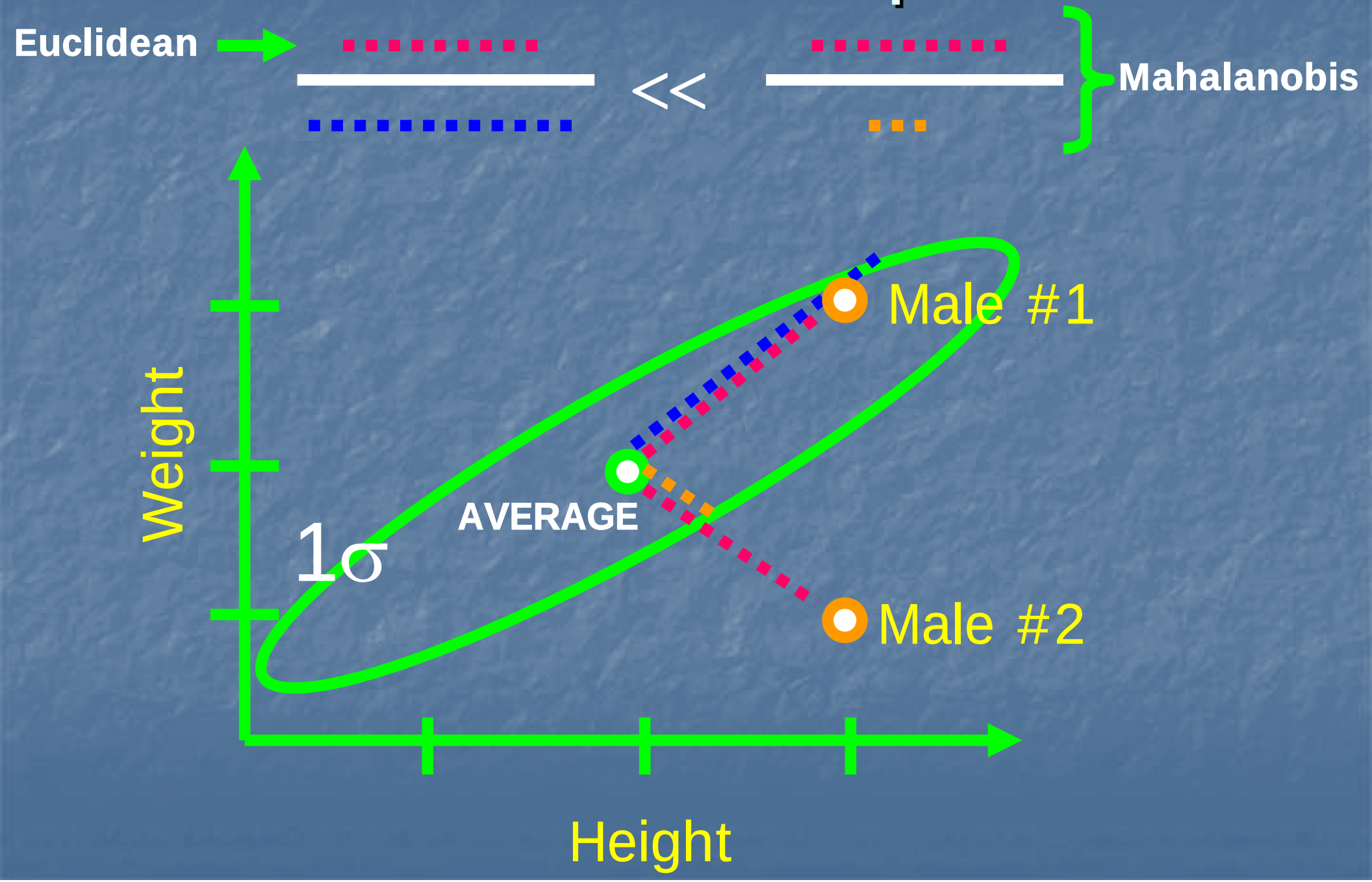
- Based on correlations between variables
- Different patterns can be identified and analyzed.
- Useful way of determining *similarity* of an unknown sample set to a known one.
- Differs from Euclidean distance in that it takes into account the correlations of the data set and is **scale-invariant**, i.e. not dependent on the scale of measurements

http://en.wikipedia.org/wiki/Mahalanobis_distance

Example: Case Study of Height/Weight of Two Men

- Male #1:
  - 6 inches above average height
  - 100 pounds **above** average weight
- Male #2
  - 6 inches above average height
  - 100 pounds **below** average weight
- By Euclidean Distance, these cases are equidistant from the average (centroid)
- By Mahalanobis Distance, Male #2 is MUCH further from the average

Previous Example:



Mahalanobis Distance and Similarity Scores

- Mahalanobis distance is an effective multivariate measure for how far points are apart in k-space in the context of the correlations between them
- So, we can use Mahalanobis distance as our Similarity Score

Methods: Data Gathering – Common Variables Between Datasets

- Total Delta V
- Occupant Age
- Weight (lbs)
- Height (ft)
- MAXAIS
- ISS
- Model year
- Gender
- Maximum Intrusion
- # of Lower Extr. Injuries
- # of Upper Extr. Injuries
- # of Head Injuries
- # of Chest Injuries

Methods: Our Sample

- NASS
  - All NASS cases meeting following criteria:
    - MAXAIS >= 3
    - 2001 to 2005
  - A subset of 1869 NASS cases (with MAXAIS >= 3)
- CIREN
  - All CIREN cases from 2001 to the present (2819)

Methods – Dealing with Missingness

- When there are large amounts of missingness that need to be dealt with.
- Possibilities:
  - Impute with averages
  - Estimate covariance structure with multiple imputation methods
  - Delete these observations

For this study, we simply imputed column-wise averages

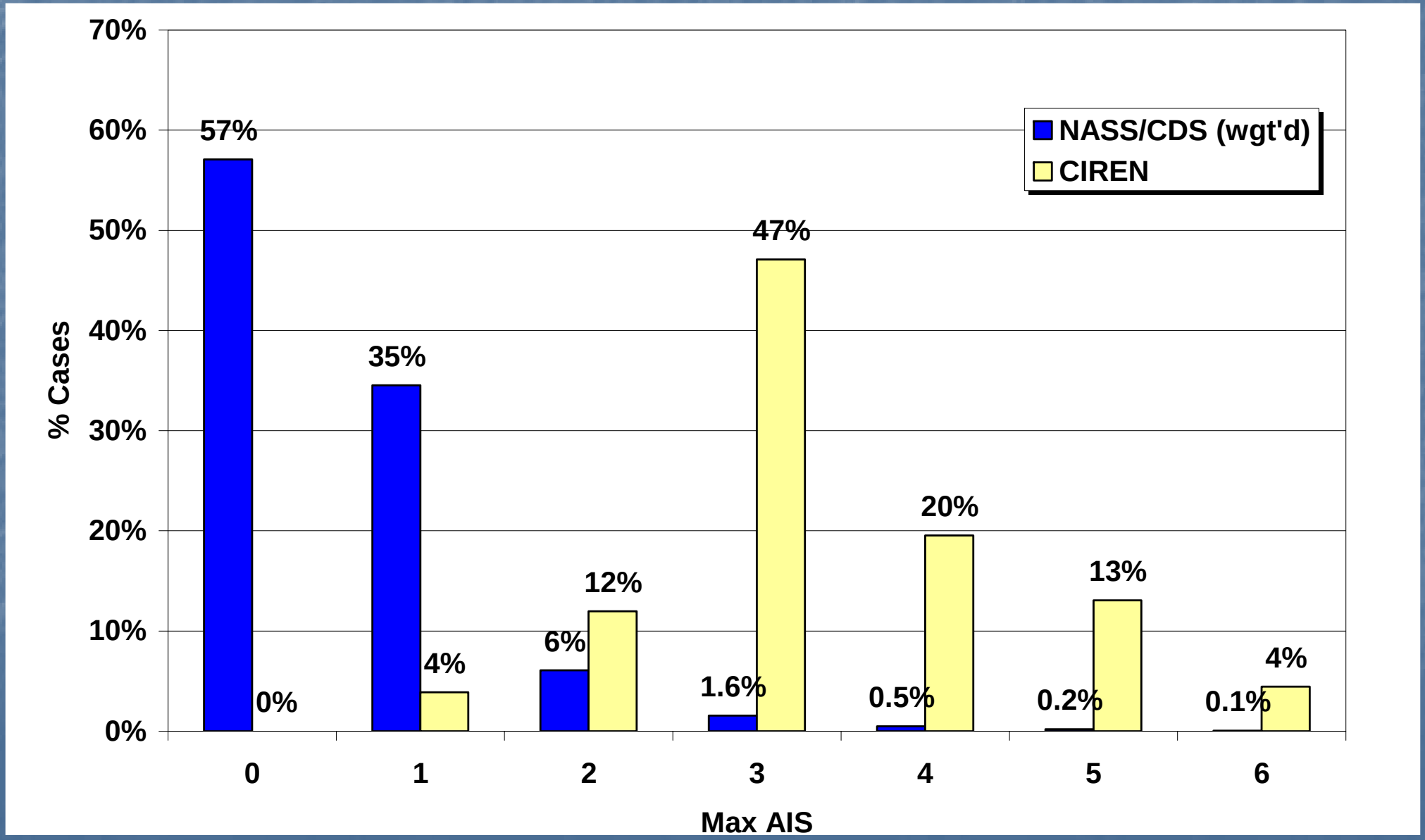
Next step

Methods: NASS Weights

- Used NASS sampling weight coefficient in calculating weighted average for NASS cases
- Used in weighting our "covariance" matrix
  - It's as if each row was repeated "k" times where "k" is the NASS weight for that row

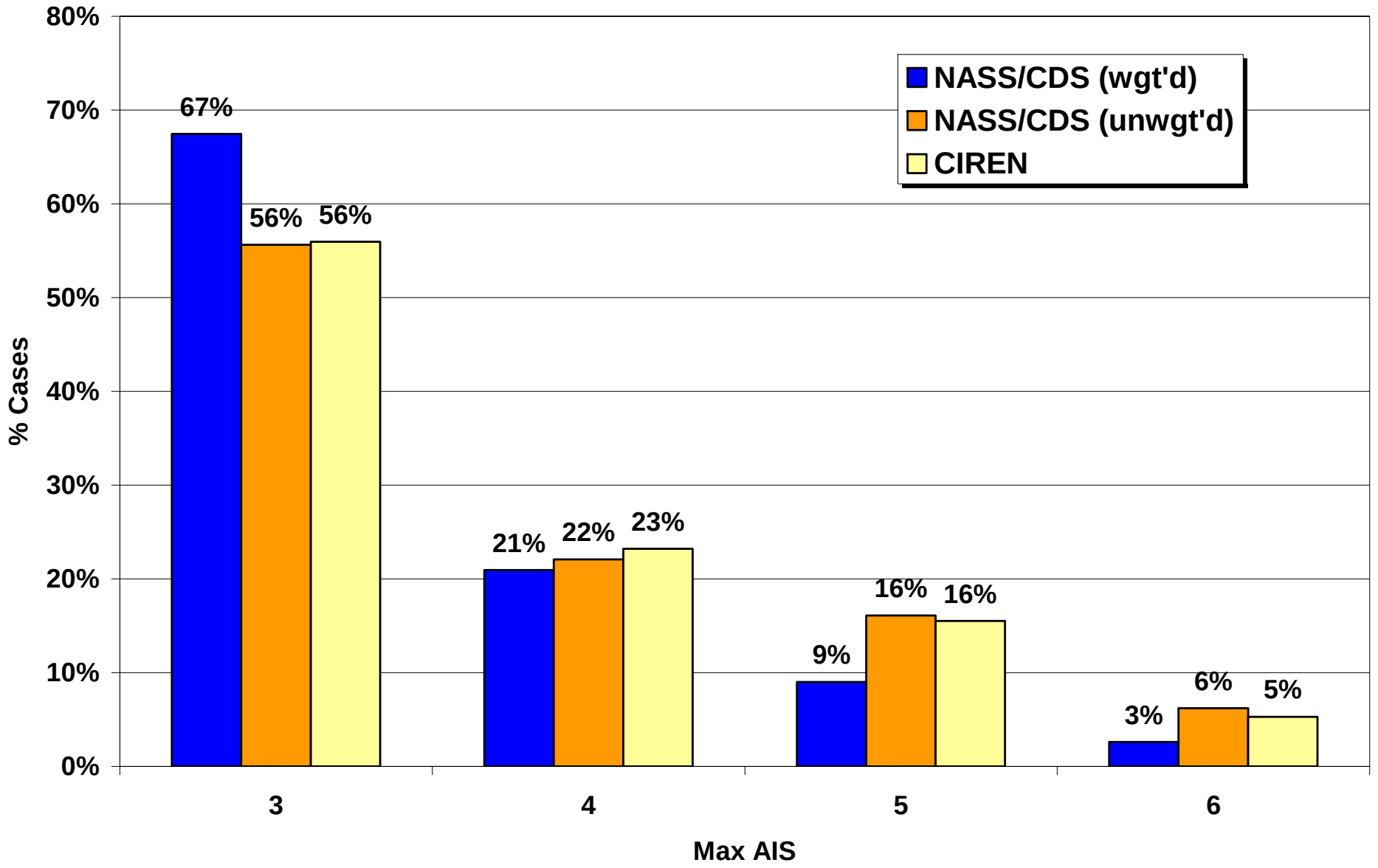
NASS/CDS is composed of lower severity cases

(Distribution of Cases by MAIS – NASS/CDS 2005 vs. CIREN)



NASS/CDS and CIREN roughly comparable for MAIS3+ Cases

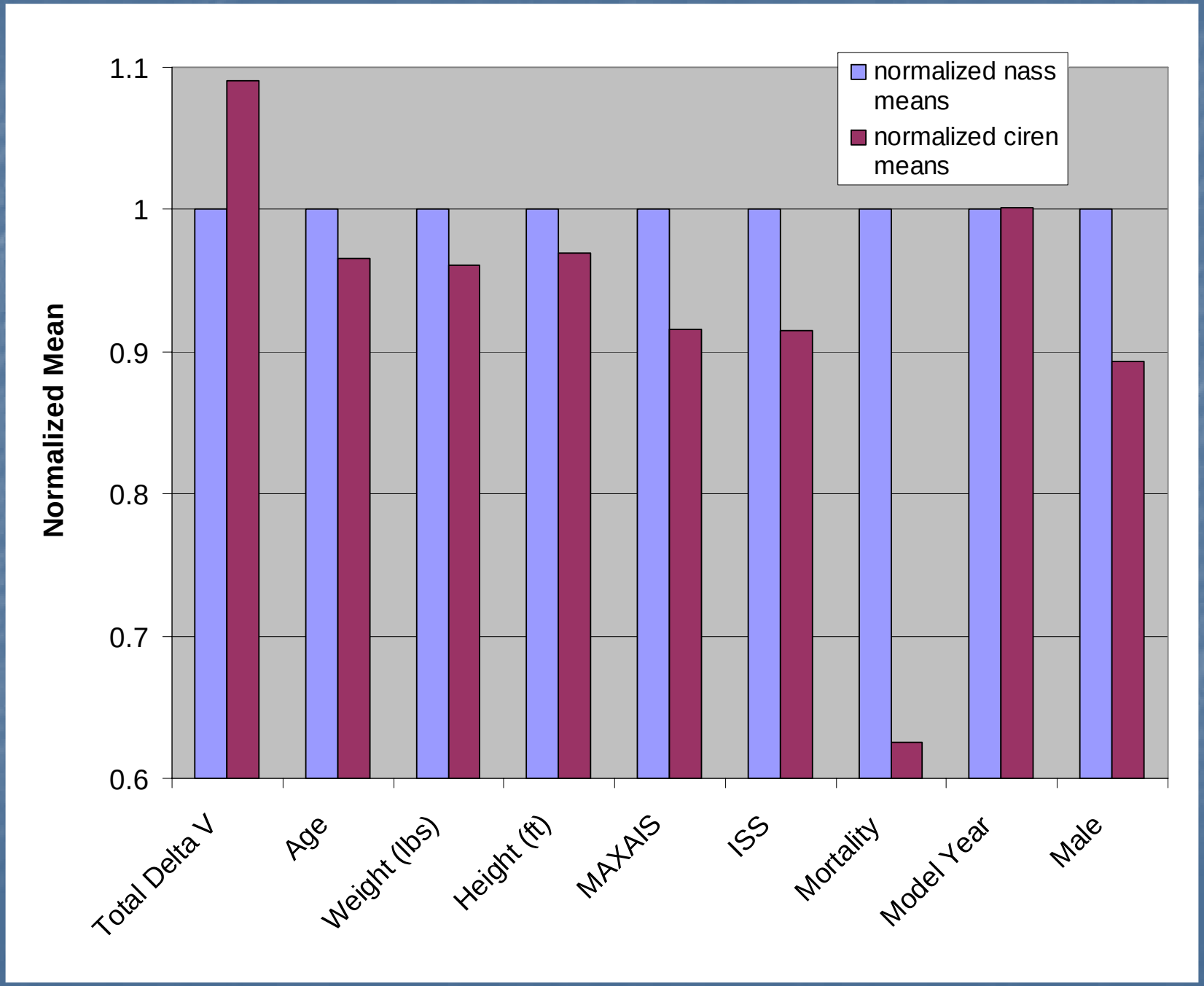
(Distribution of MAIS3+ Cases– NASS/CDS 2005 vs. CIREN)



Weighted NASS and CIREN Means

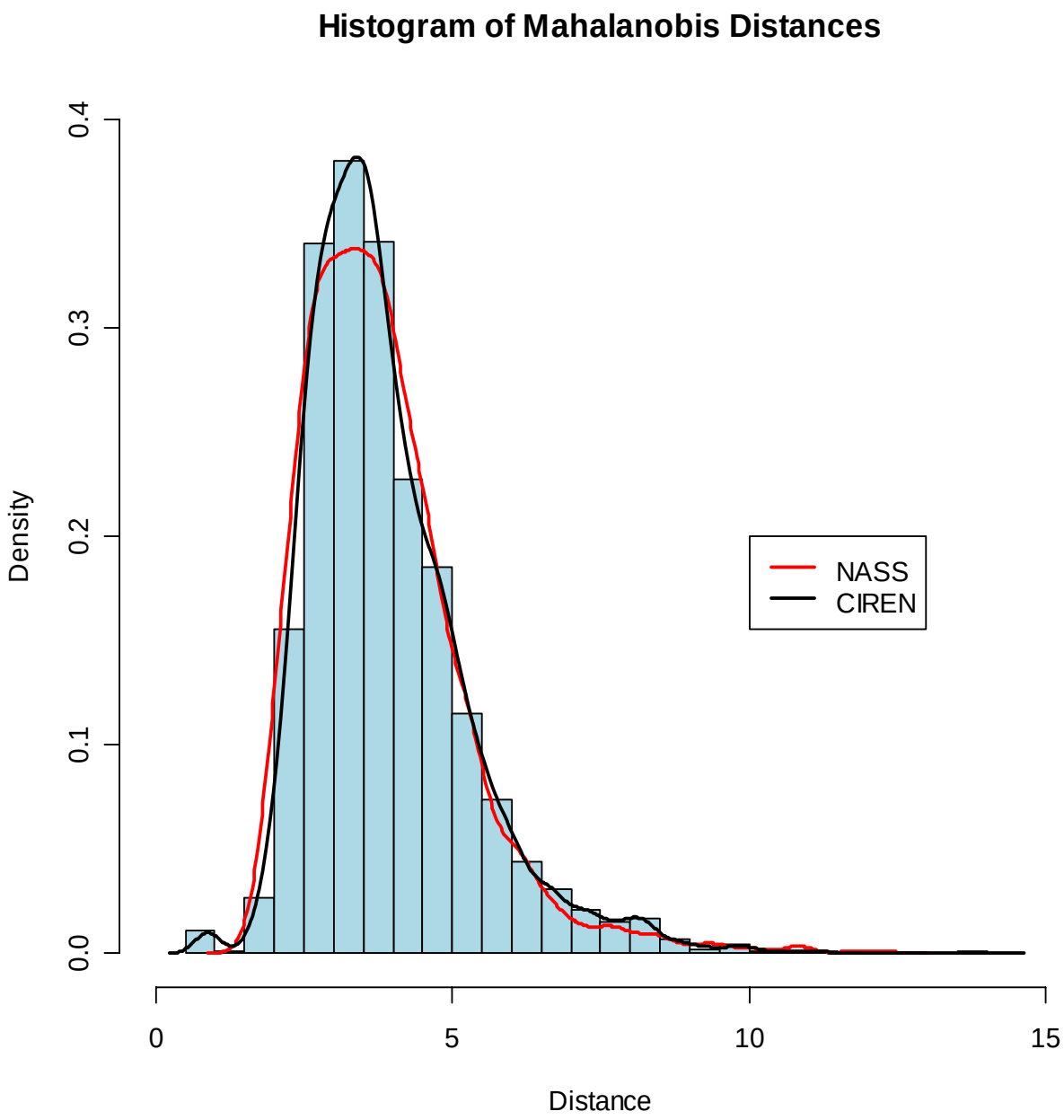
Variable	Weighted NASS Mean	CIREN Means
Total Delta V	37.5	40.9
Age	38.0	36.7
Weight (lbs)	167	160.5
Height (ft)	5.56	5.39
MAXAIS	3.7	3.39
ISS	23.6	21.6
Mortality	0.24	0.15
Model Year	1994	1997
Male	0.56	0.50

Normalized CIREN to NASS Means

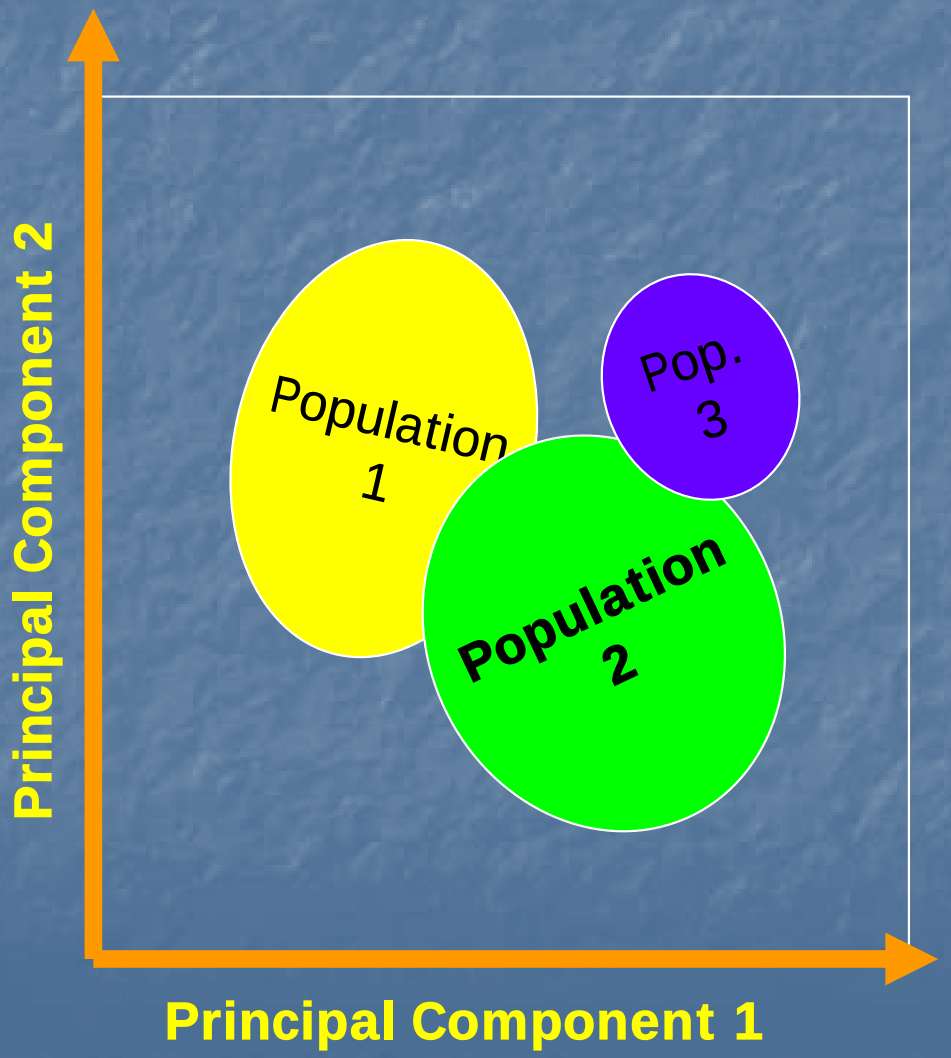


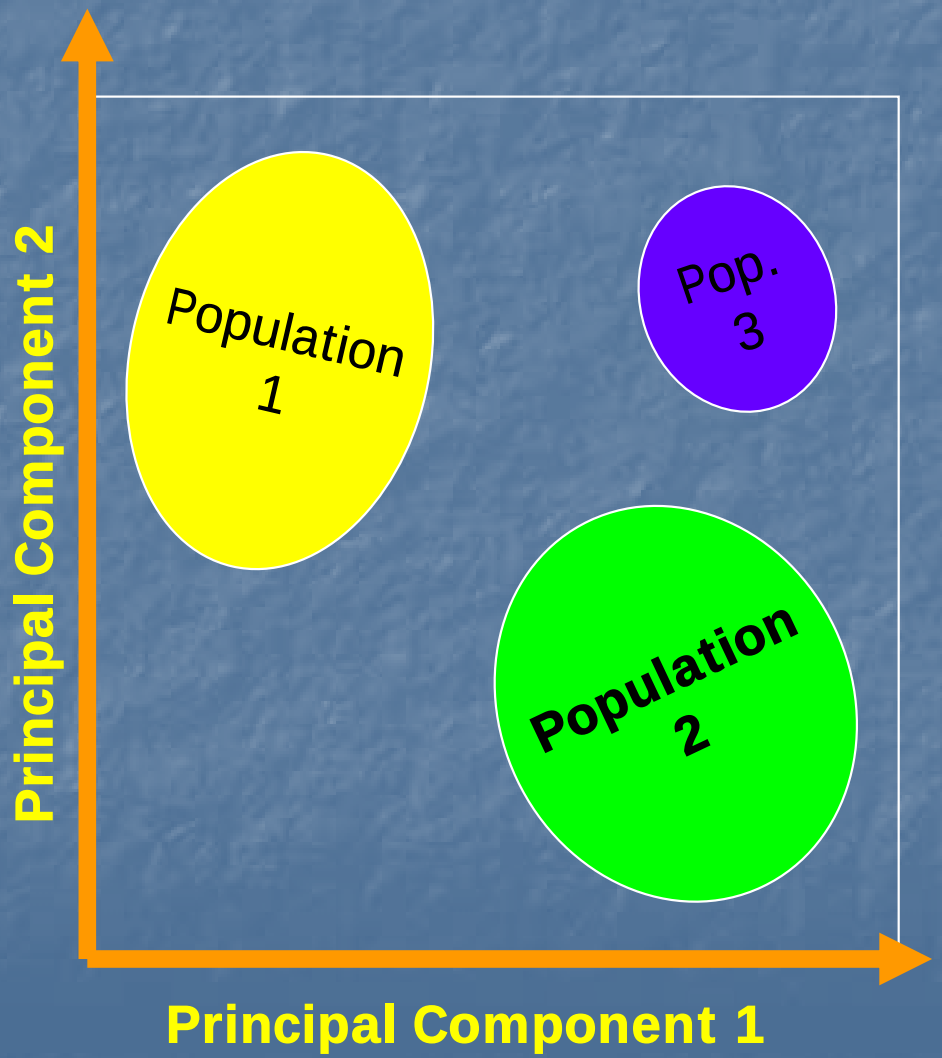
Results

- Distance Distribution is roughly the same.
- CIREN has a greater proportion of its cases clustered about the mean distance than NASS does
- “CIREN is more like NASS than NASS”



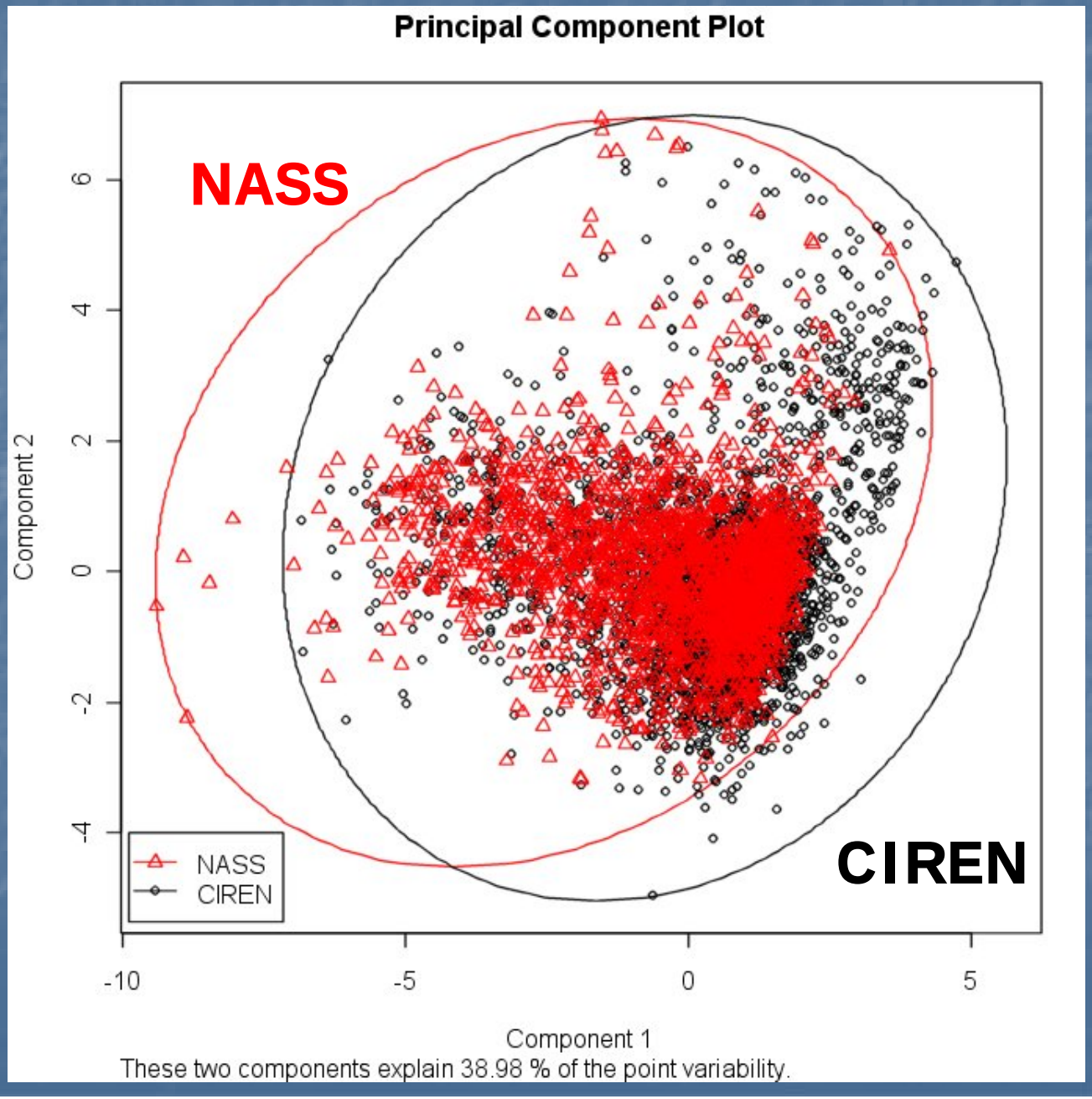
Potential outcomes of principal components analysis





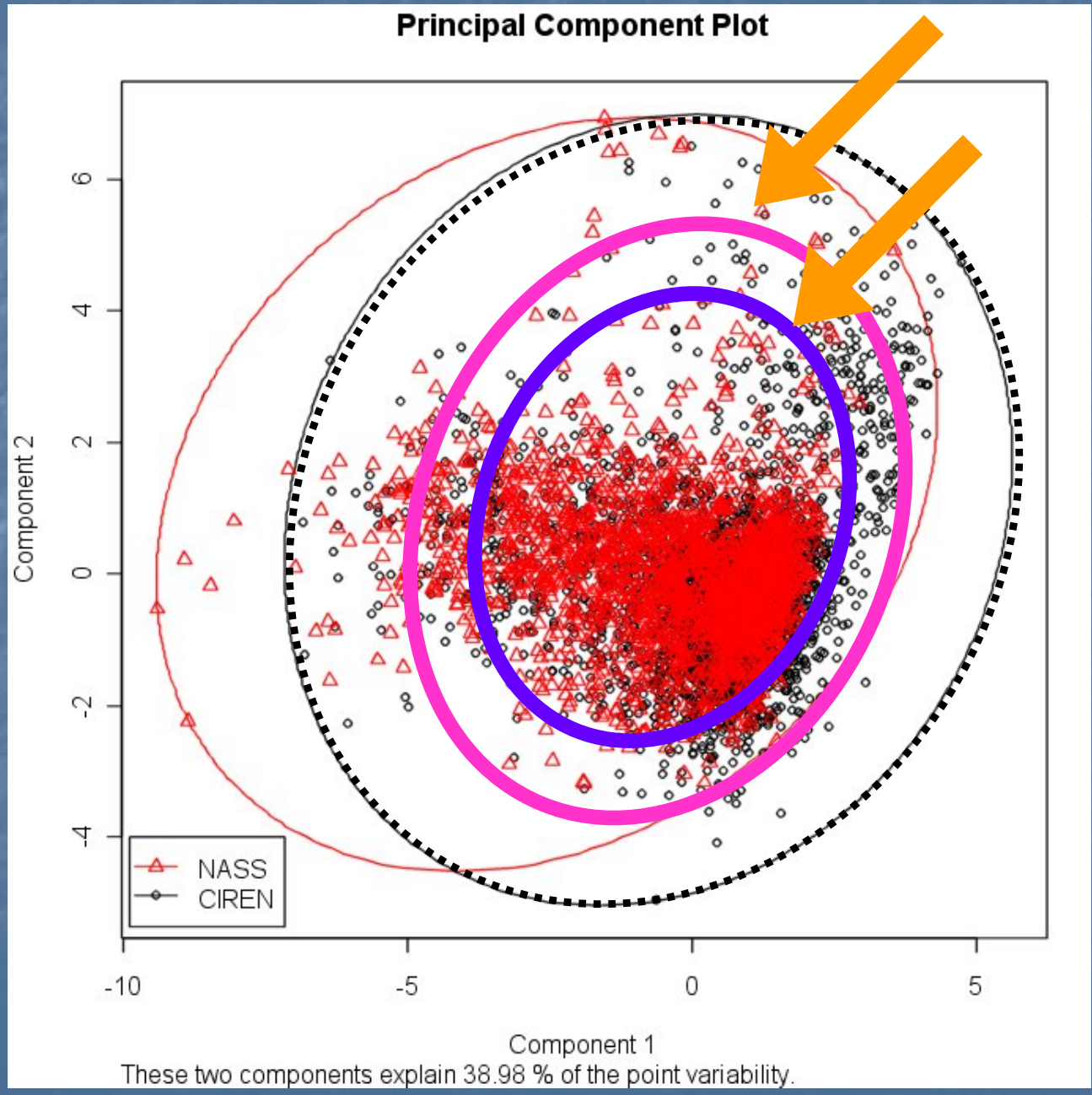
Principal Components Analysis

- There are differences between NASS and CIREN
- But more important part is **overlap, they are largely the same.**
- CIREN cases are driven to the right by some low MAIS

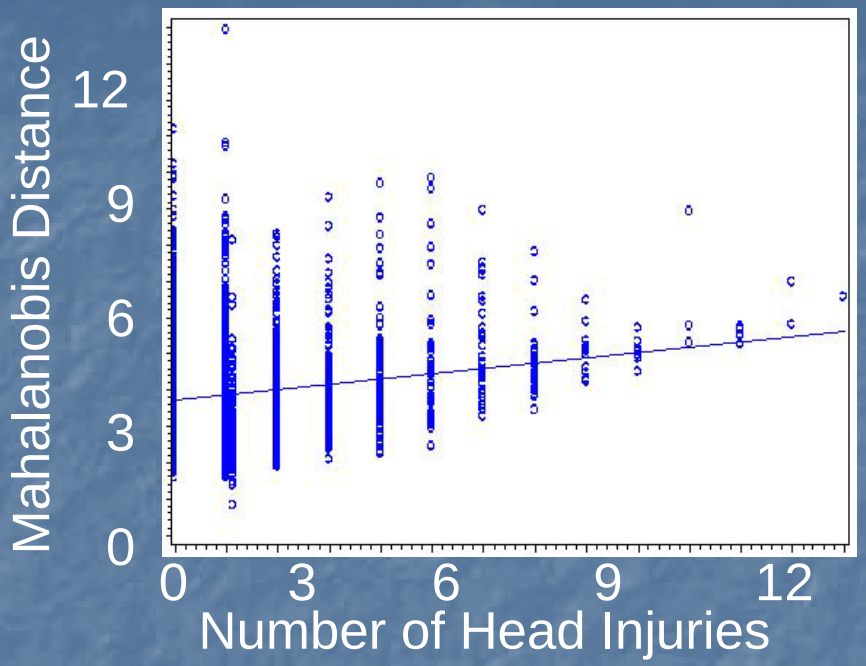


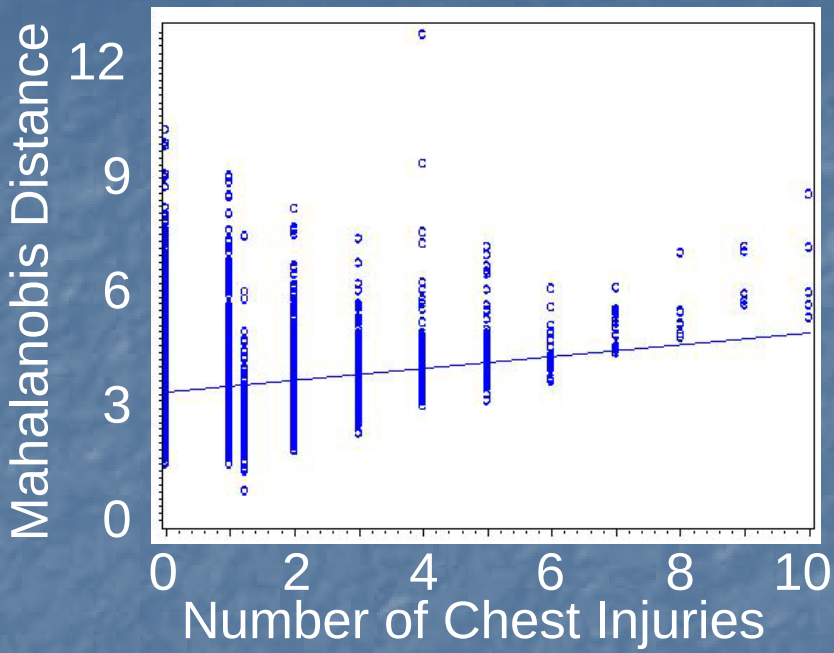
Principal Components Analysis

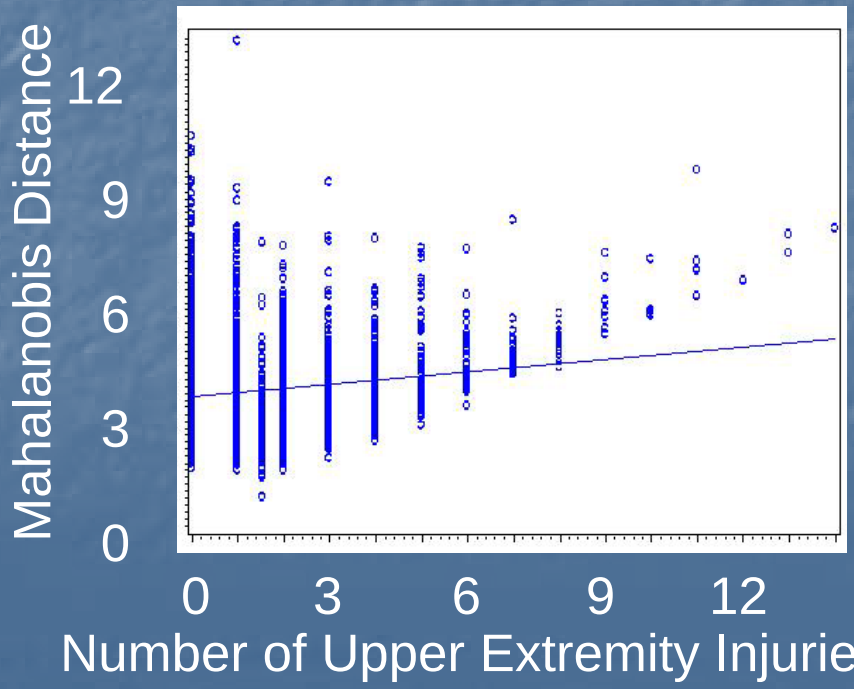
- Could **subset** CIREN or NASS to minimize Mahalanobis distance depending on interest.
- This is shown here in 2-d by minimizing variation in 1st 2 principal components.

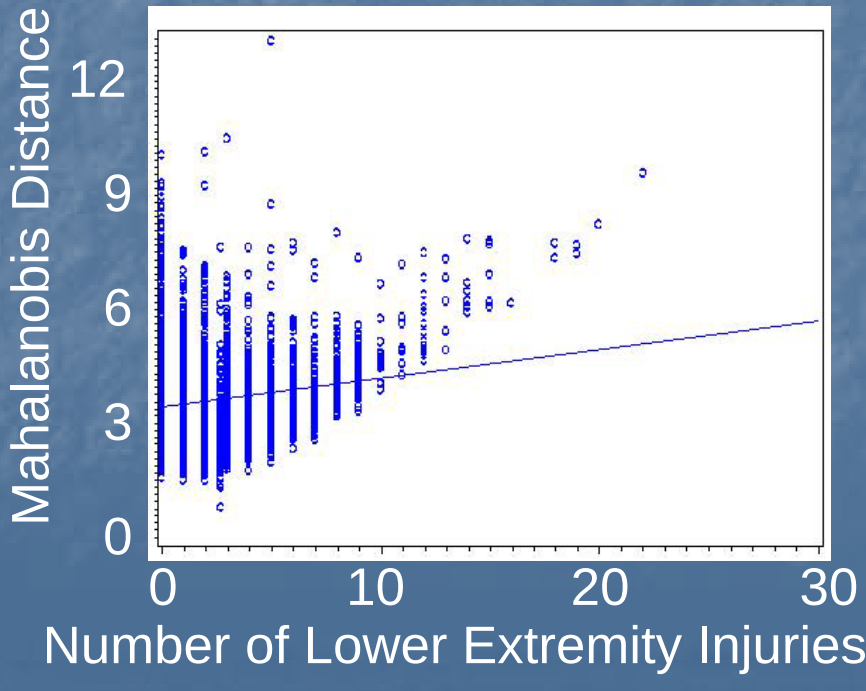


Distance vs. # Injuries

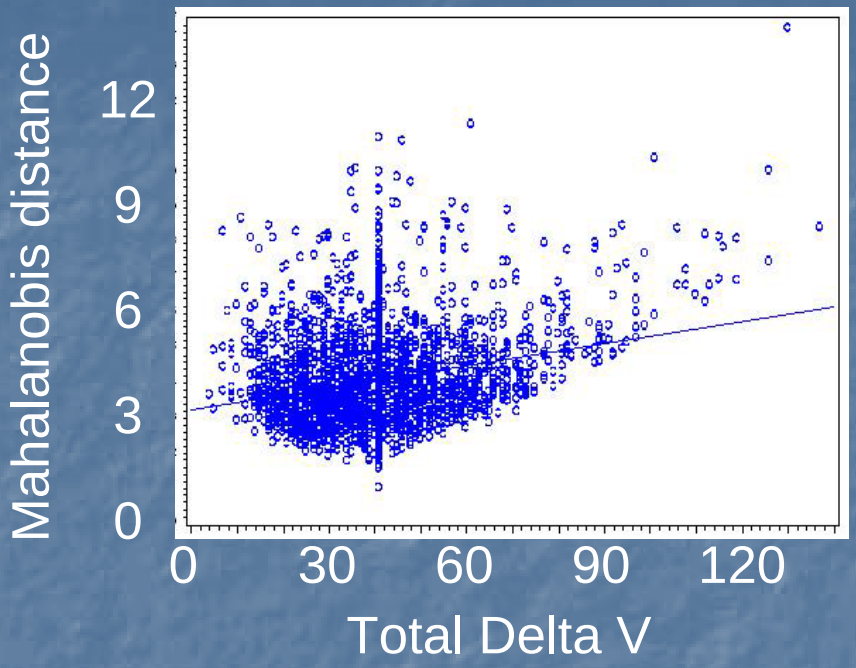


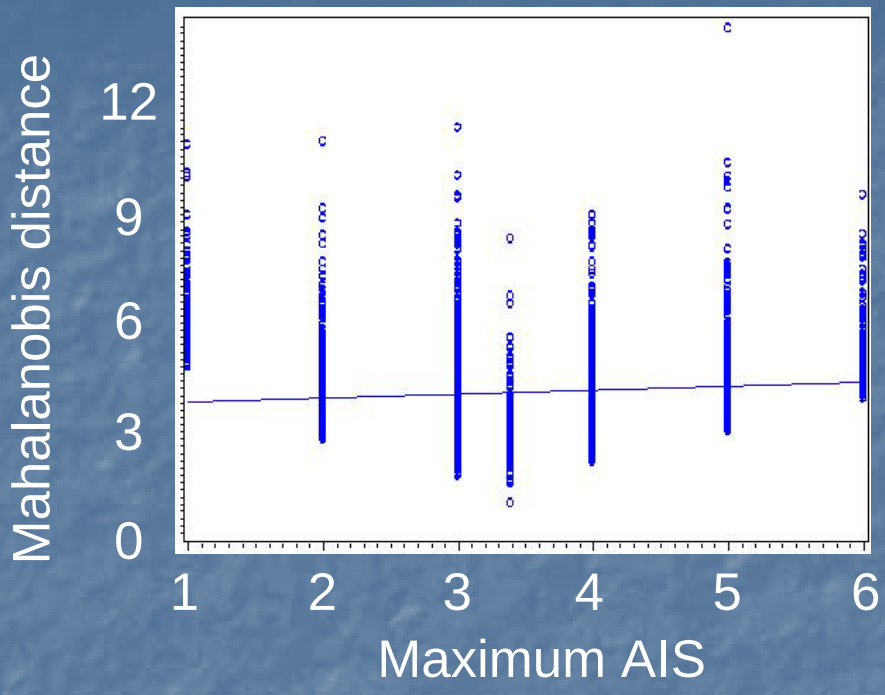


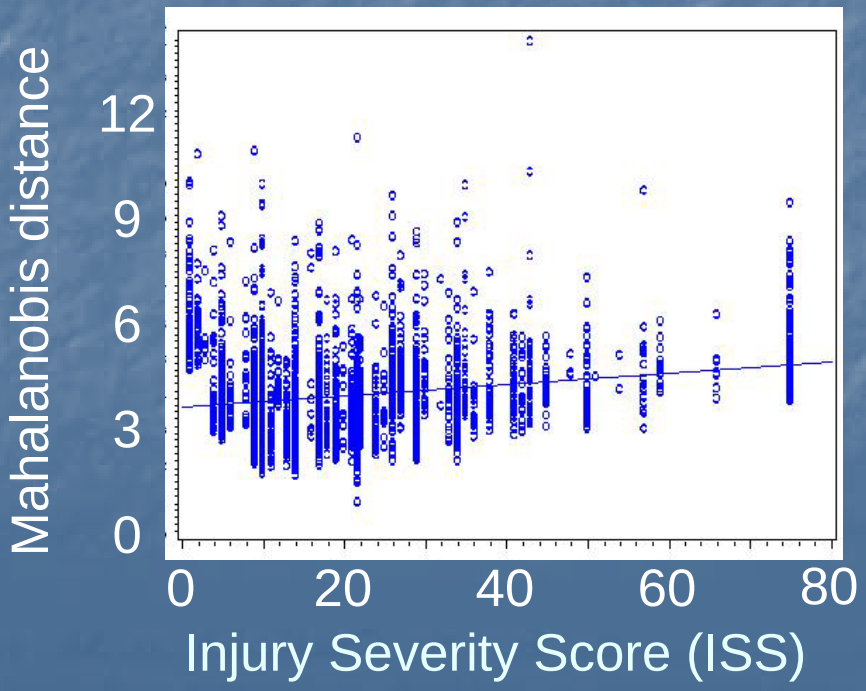


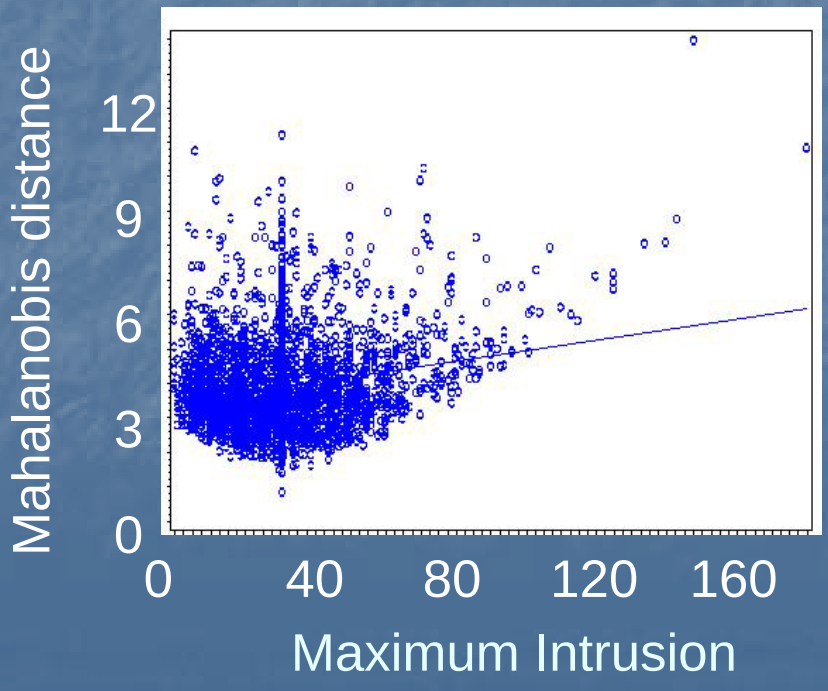


Distance vs. Crash and Injury Characteristics

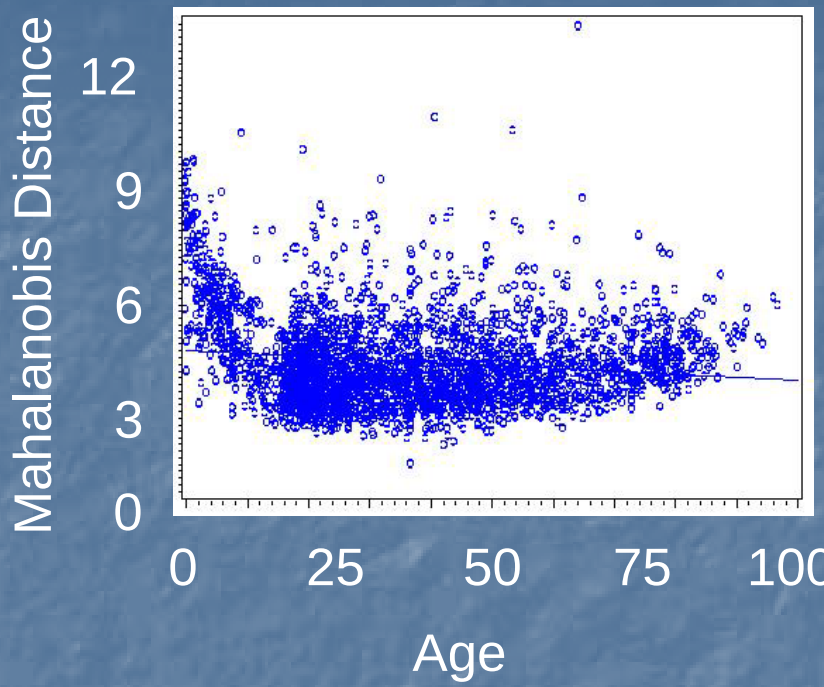


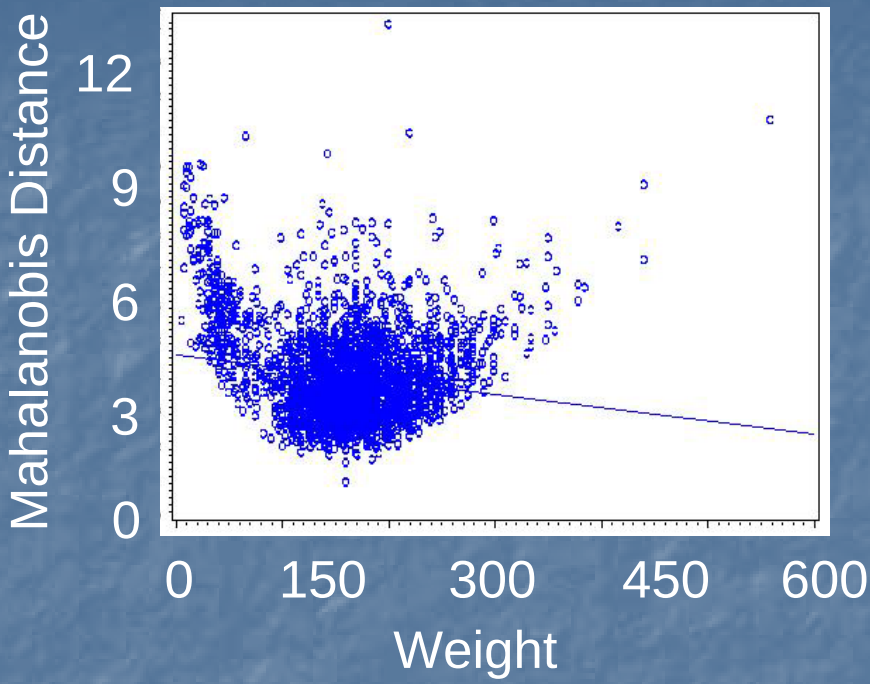


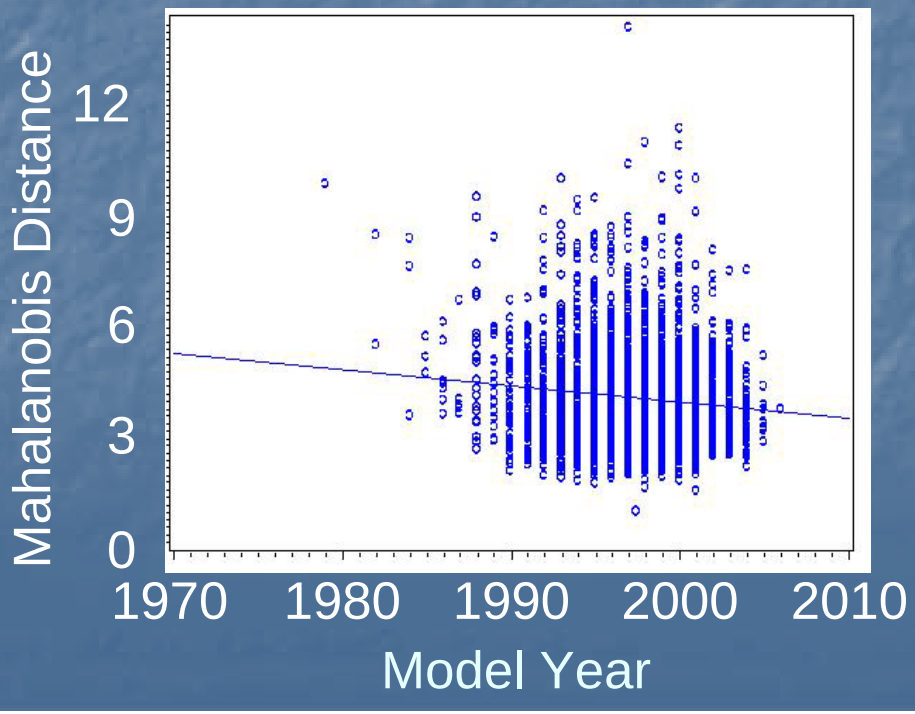


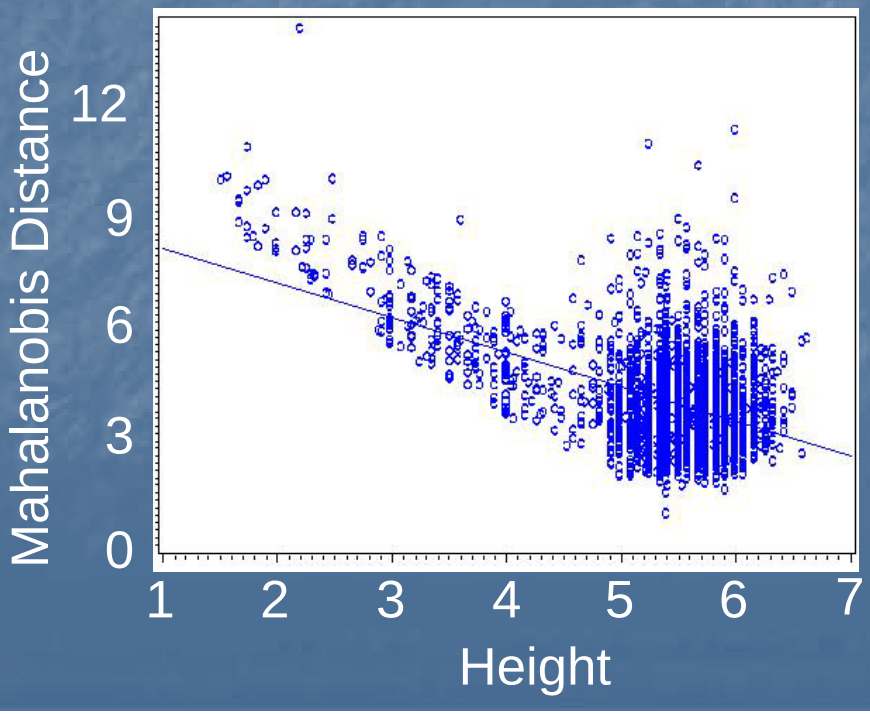


Distance vs. Model Year and Anthropometric Variables





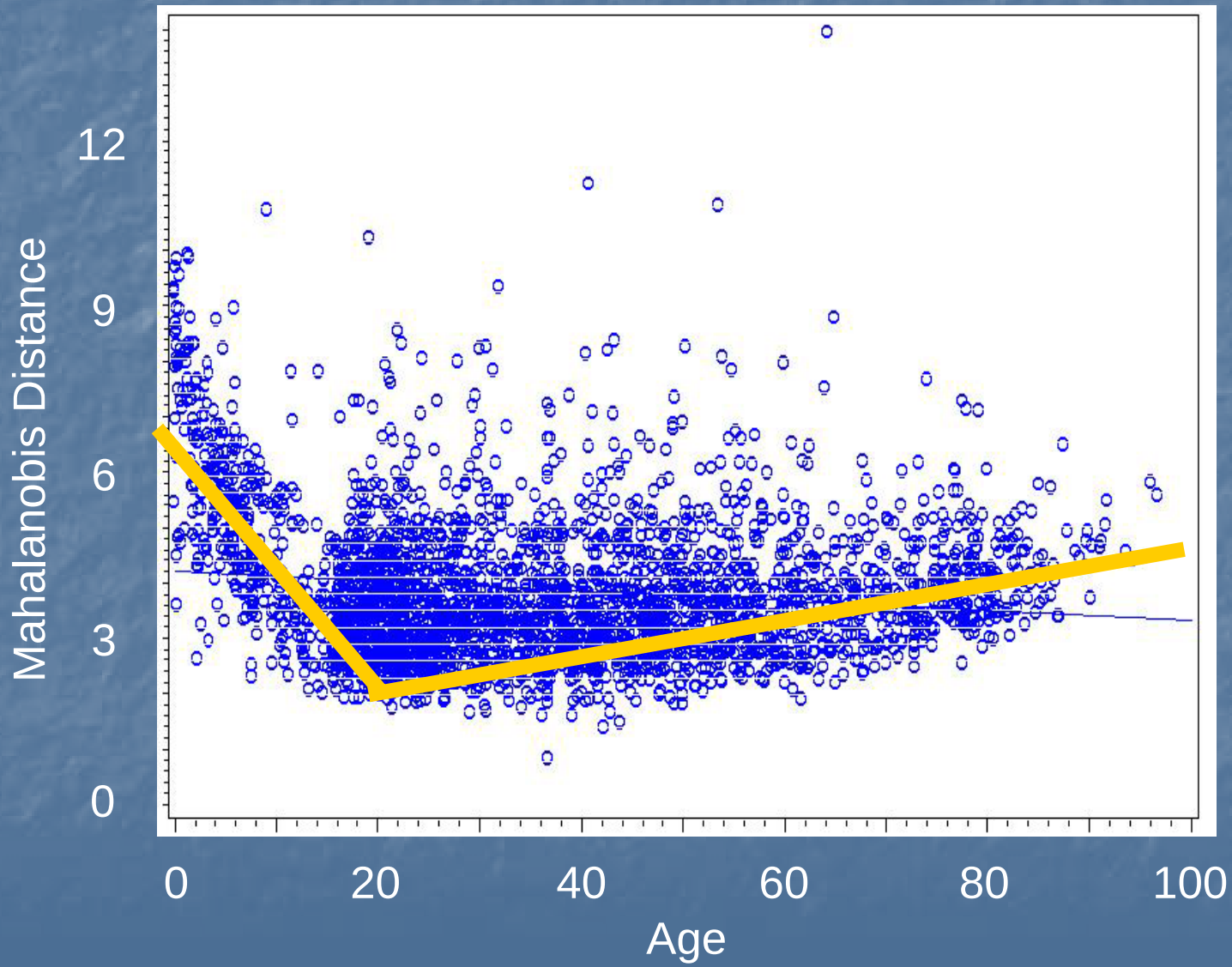




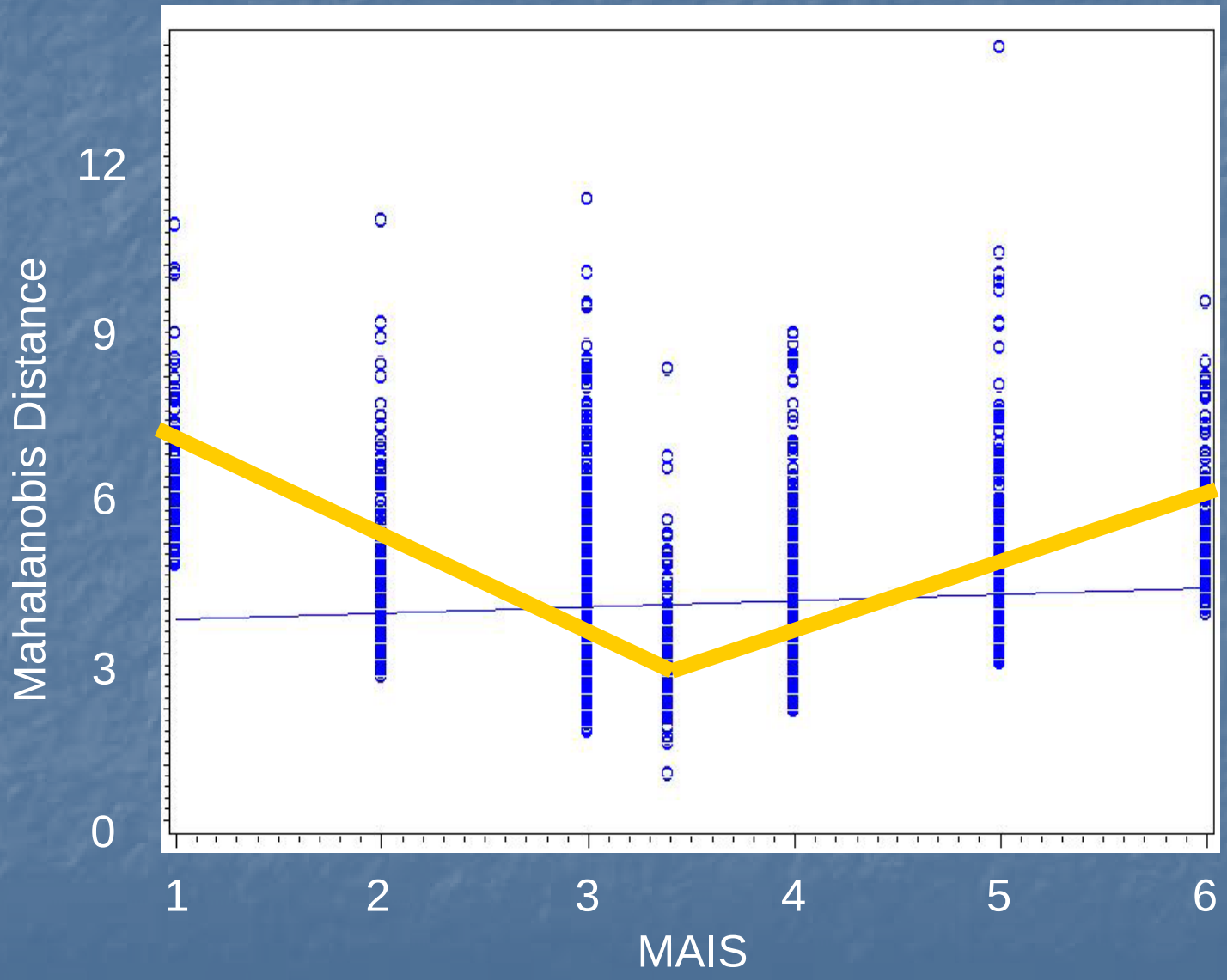
Caveats

- Other regressions could be performed
- Shape of data suggests maybe they should be performed

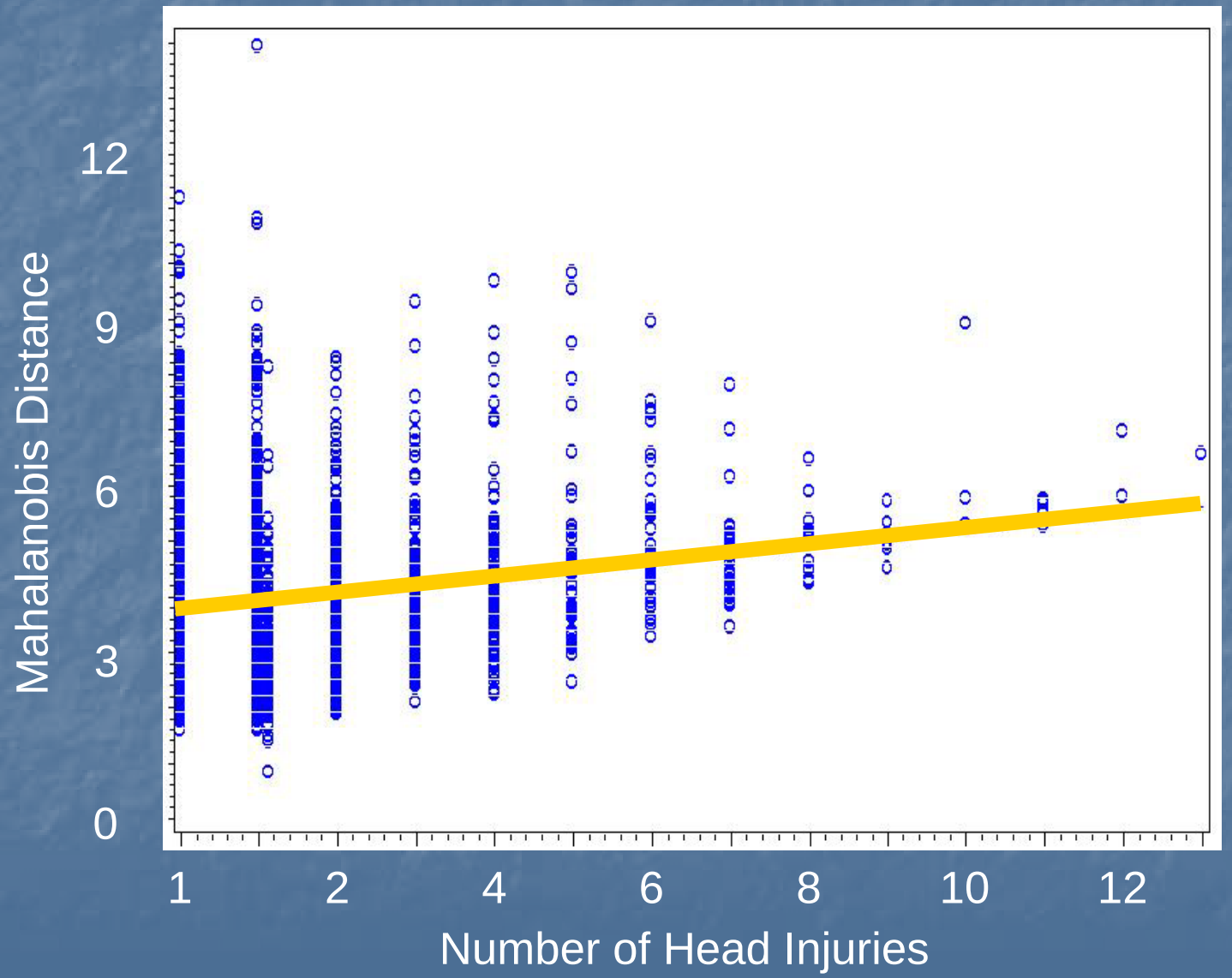
Distance vs. Age



Distance vs. Max AIS



Distance vs. # Head Injuries



Results – Subset of CIREN cases "Furthest" from weighted average NASS case

Delta V	Head	Chest	Up_ex	Low_ex	Age	Weight	Height	Gender	Max_int	Max_AIS	ISS	Dead	Model Year
130	1	4	1	5	64.3	201	2.2	1	148	5	43	1	1997
61	0	0	0	3	40.8	560	6	1	31	3	22	0	2000
41	1	0	0	2	53.5	220	5.25	1	180	2	9	0	1998
46	1	0	0	0	9	66	1.74	1	6	1	2	0	2000
101	0	4	11	22	19.3	143	5.68	1	71	5	43	1	1997
36	0	0	0	0	1.3	24	1.57	0	13	1	1	0	2000
126	0	1	3	2	1.4	26	2.49	1	70	3	10	0	1999
35	0	0	0	0	0.2	13	1.9	1	31	1	1	0	2001
41	5	1	0	0	0.2	11	1.51	1	12	5	35	0	1993
45	4	1	1	0	0.1	11	1.84	1	50	5	57	0	1979

Conclusions

- CIREN crashes are similar to NASS crashes
  - For severe injuries (MAIS 3)
- Valuable approach to improve understanding and use of CIREN
- Could refine method to be able to assign a NASS similarity score to each CIREN crash
- Or NASS crash similarity or anthropomorphic similarity score or injury similarity score
- Good method to quality control data in CIREN and NASS

Thank you

- Acknowledgements
  - Toyota
  - Carol Flannagan and Jonathan Rupp, UMTRI
  - Analytical Committee

- Disclaimer: the published material represents the position of the authors and not necessarily that of Toyota or NHTSA
- For more information contact:

  - Dr. Joel Stitzel (jstitzel@wfubmc.edu)