LA TROBE UNIVERSITY

All kinds of clever

Modelling training loads and injury: methodological issues and improved strategies

David Carey
Kay Crossley, Rod Whiteley, Andrea Mosler, Kok-Leong Ong, Justin Crow, Meg Morris

@dlcarey88

Modelling continuous variables:

Modelling continuous variables: the dangers of DISCRETISATION

Discretisation = transforming continuous → discrete

Discretisation = transforming continuous → discrete

"...split by percentiles..."

"...split into equal groups..."

"...values 1SD above the mean were classified as high..."

"...median split..."

"...categorised based on z-score..."

What did we do?

Used the study of **training loads** and **injury** to illustrate the issues caused by discretisation

Acute:chronic workload ratio (ACWR) vs injury

Continuous variable

Binary outcome

Acute:chronic workload ratio (ACWR) vs injury

Continuous variable

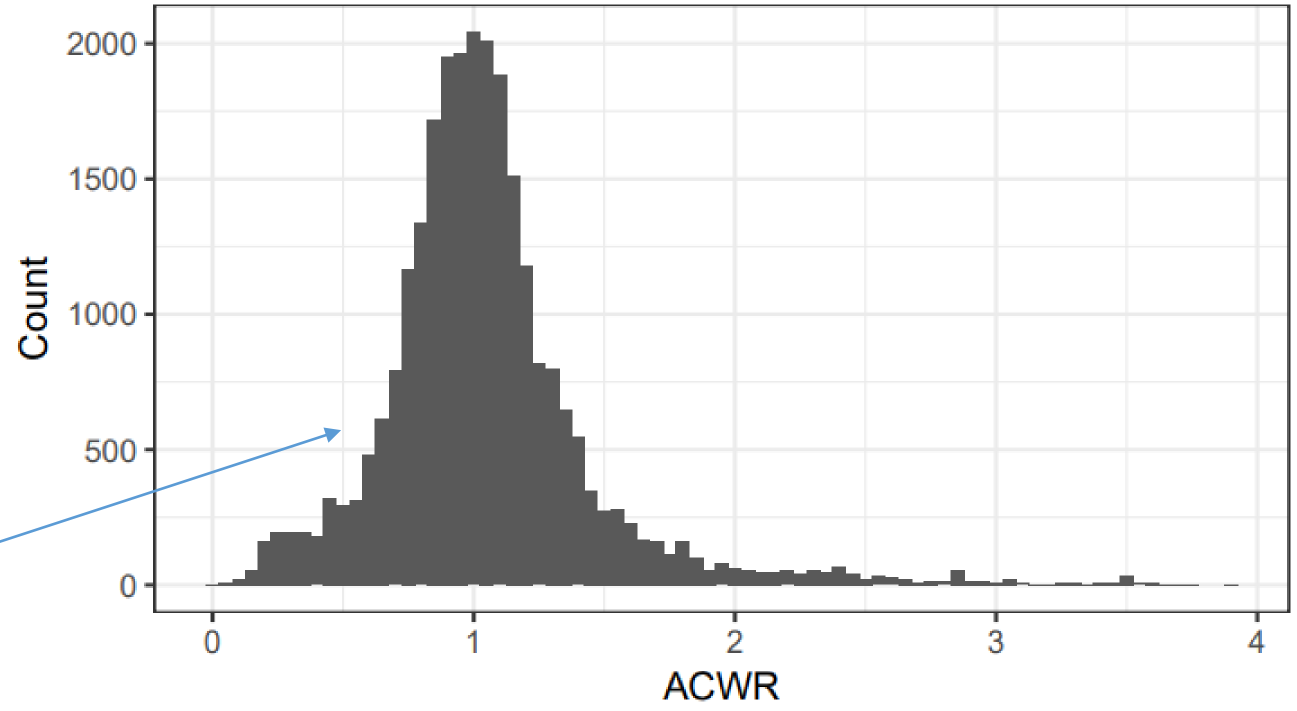
Binary outcome

Lots of previous studies looking at the **same** relationship

Lots of **different** modelling strategies

What we did:

- Got a large sample of workload data from **AFL** (n = 2,550) and **soccer** (n = 23,742)

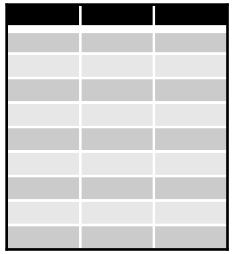


Representative distribution

What we did:

- Got a large sample of workload data from AFL (n = 2,550) and soccer (n = 23,742)

- Simulated a data set of 5000 observations by randomly drawing from the sample



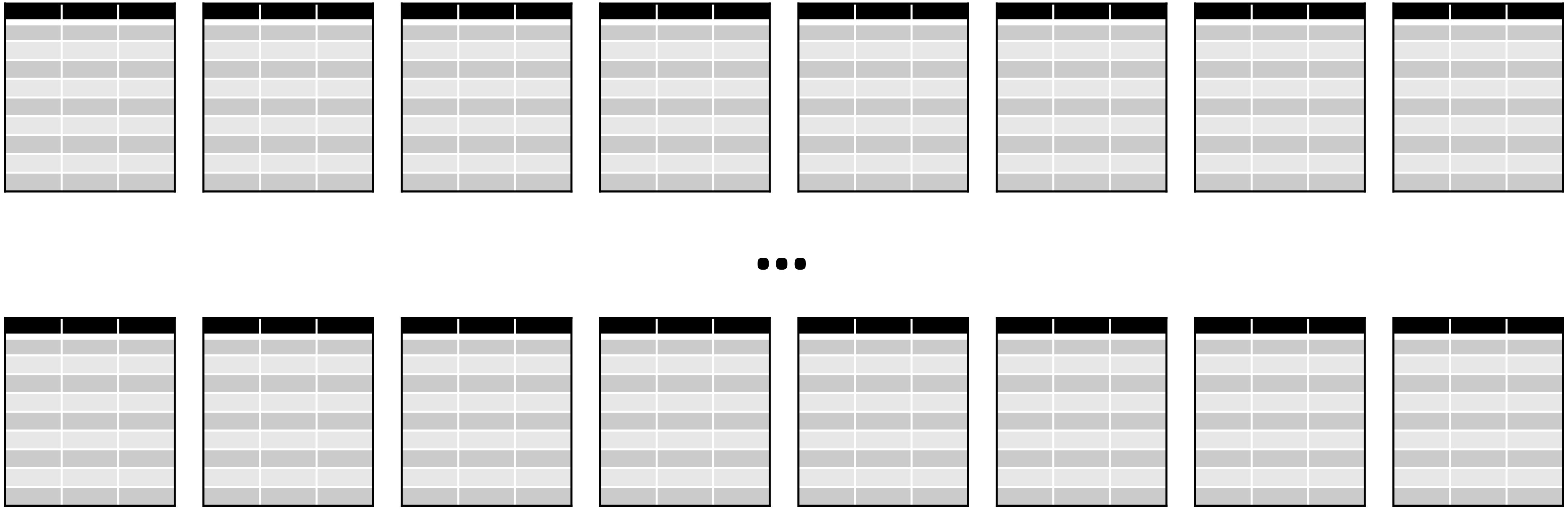
A hypothetical future study

What we did:

So we can also look at variability in results

- Got a large sample of workload data from AFL (n = 2,550) and soccer (n = 23,742)

- Simulated a data set of 5000 observations **(100 times)**



What we did:

- Got a large sample of workload data from AFL (n = 2,550) and soccer (n = 23,742)
- Simulated a data set of 5000 observations (100 times)
- Artificially inserted **injuries** in the data following a **known** injury risk shape

What we did:

- Got a large sample of workload data from AFL (n = 2,550) and soccer (n = 23,742)
- Simulated a data set of 5000 observations (100 times)
- Artificially inserted injuries in the data following a known injury risk shape
- Analysed the data using:
  - **3 x discretisation methods**
  - **2 x continuous methods**

What we did:

- Got a large sample of workload data from AFL (n = 2,550) and soccer (n = 23,742)
- Simulated a data set of 5000 observations (100 times)
- Artificially inserted injuries in the data following a known injury risk shape
- Analysed the data using:
  - **3 x discretisation methods**
  - **2 x continuous methods**

compared the results

Analysis methods

Analysis methods

Discrete

- D1: z-score categories
- D2: Percentiles
- D3: Arbitrary cut points

Analysis methods

Discrete

- D1: z-score categories
- D2: Percentiles
- D3: Arbitrary cut points

Continuous

- C1: Restricted cubic splines
- C2: Fractional polynomials

Analysis methods

Discrete

- D1: z-score categories
- D2: Percentiles
- D3: Arbitrary cut points

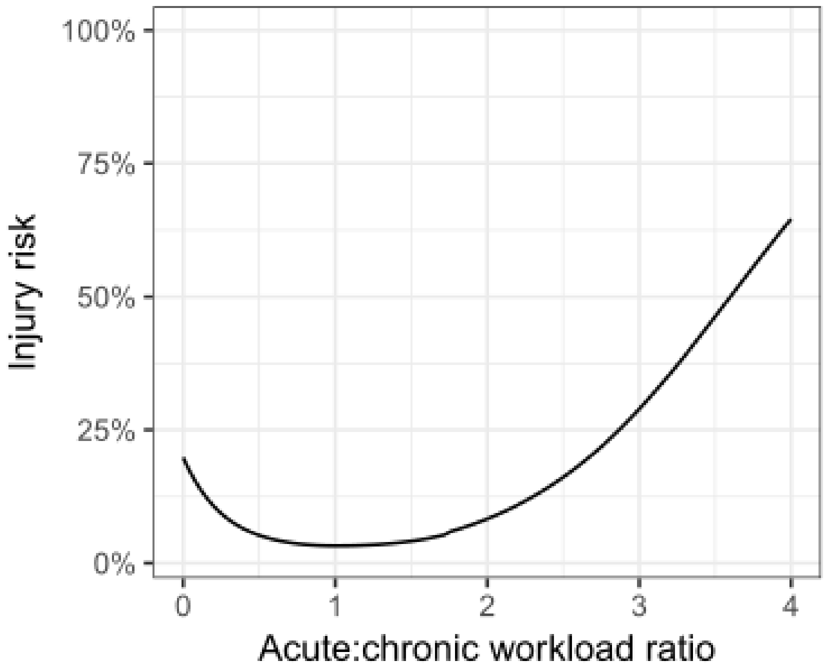
All have been used **multiple times** in existing literature

Continuous

- C1: Restricted cubic splines
- C2: Fractional polynomials

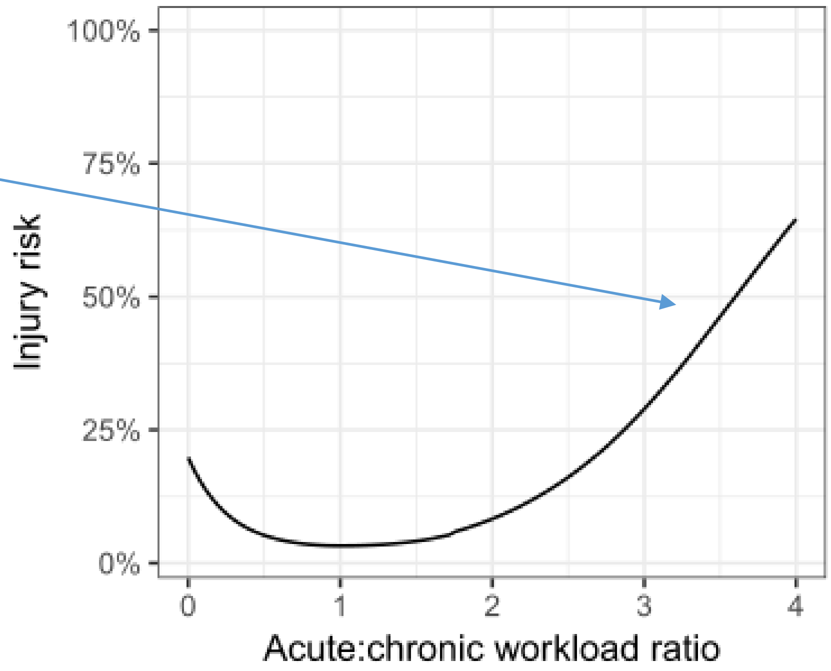
Have **not been used** in training load and injury studies

Scenario 1: U-shaped risk



Scenario 1: U-shaped risk

Simulated **injuries** in all 100 data sets following this exact curve



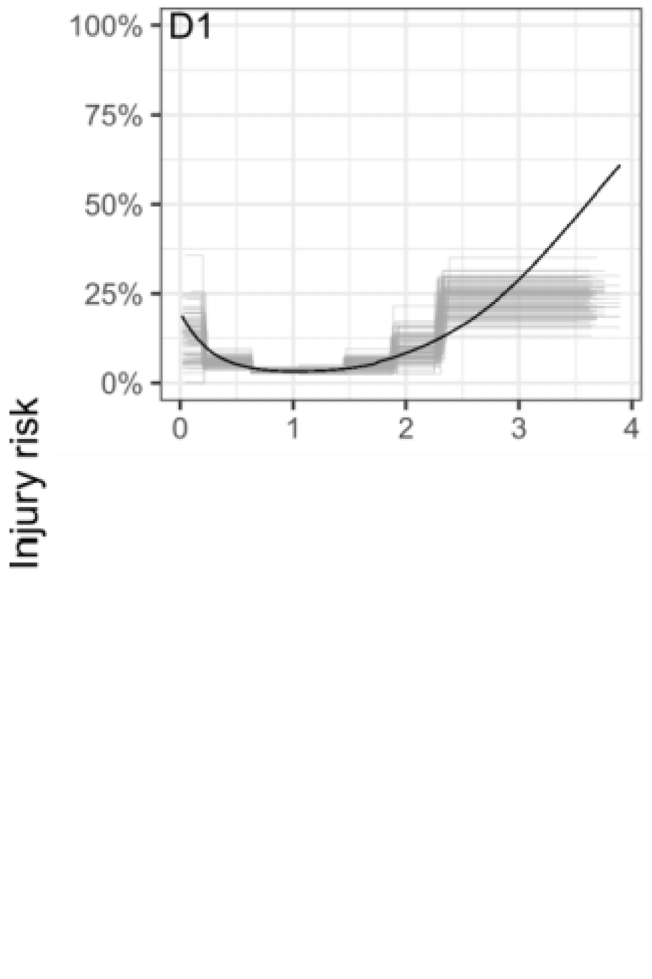
Results

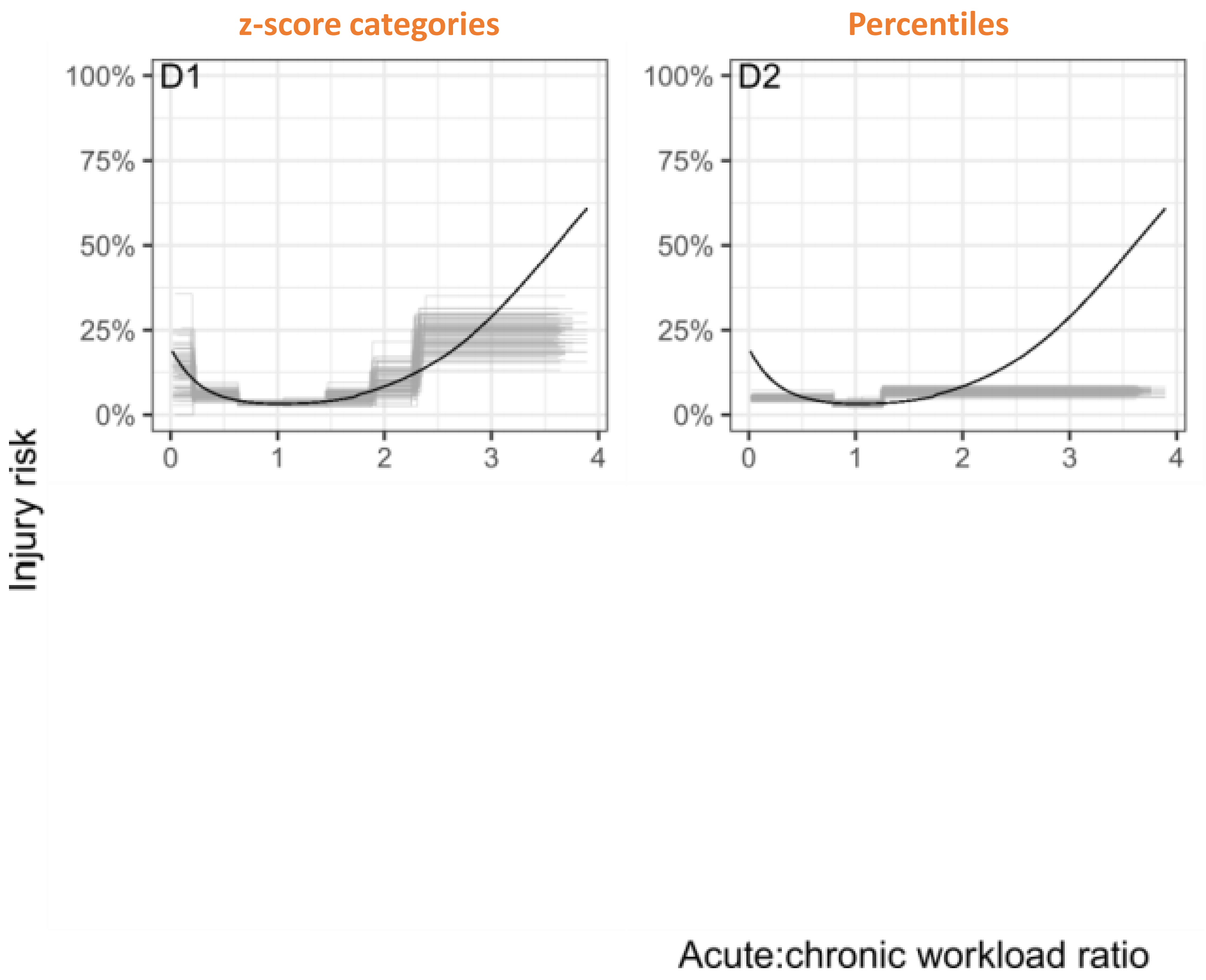
Injury risk

How well could each analysis method recover the **true relationship**?

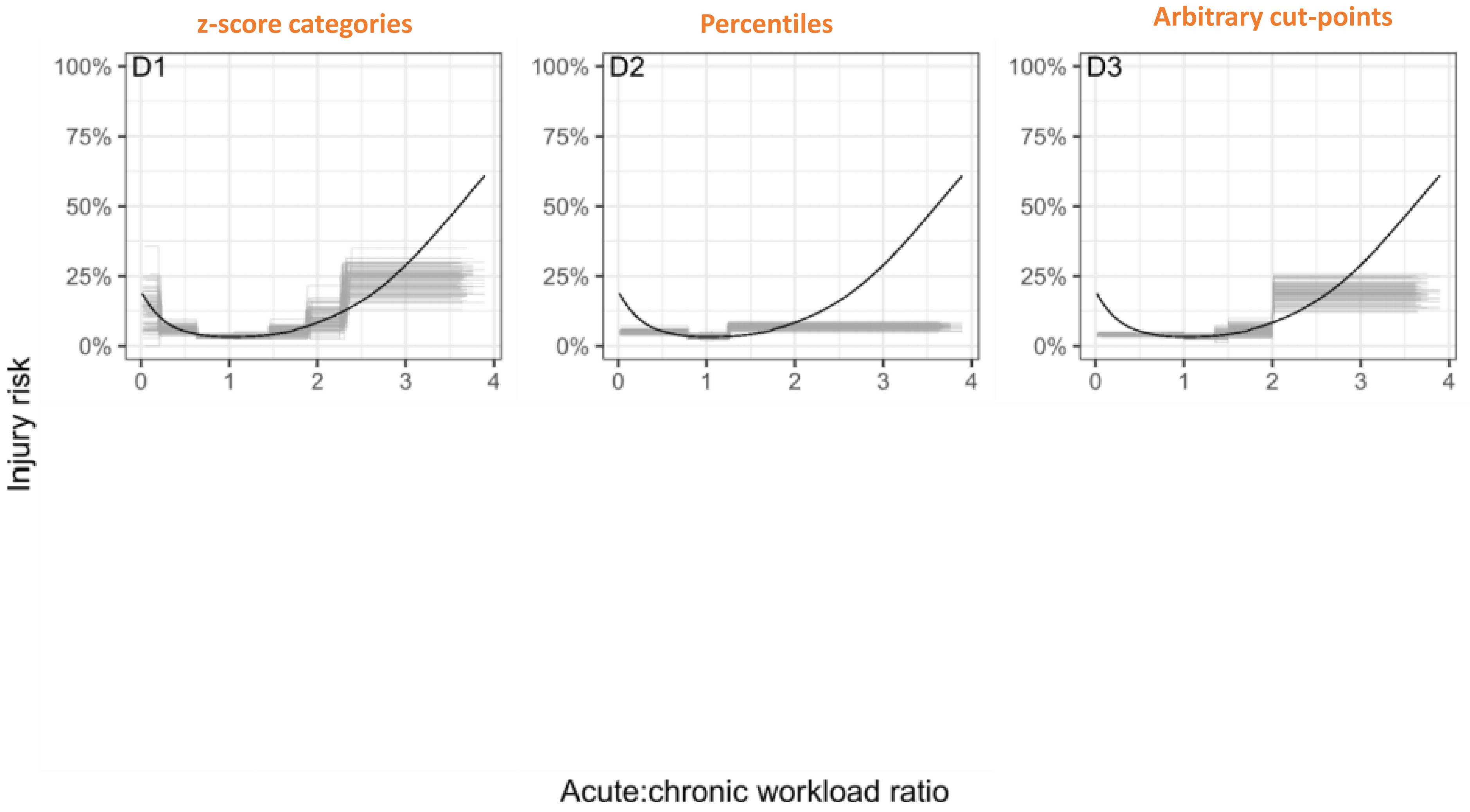
Acute:chronic workload ratio

z-score categories





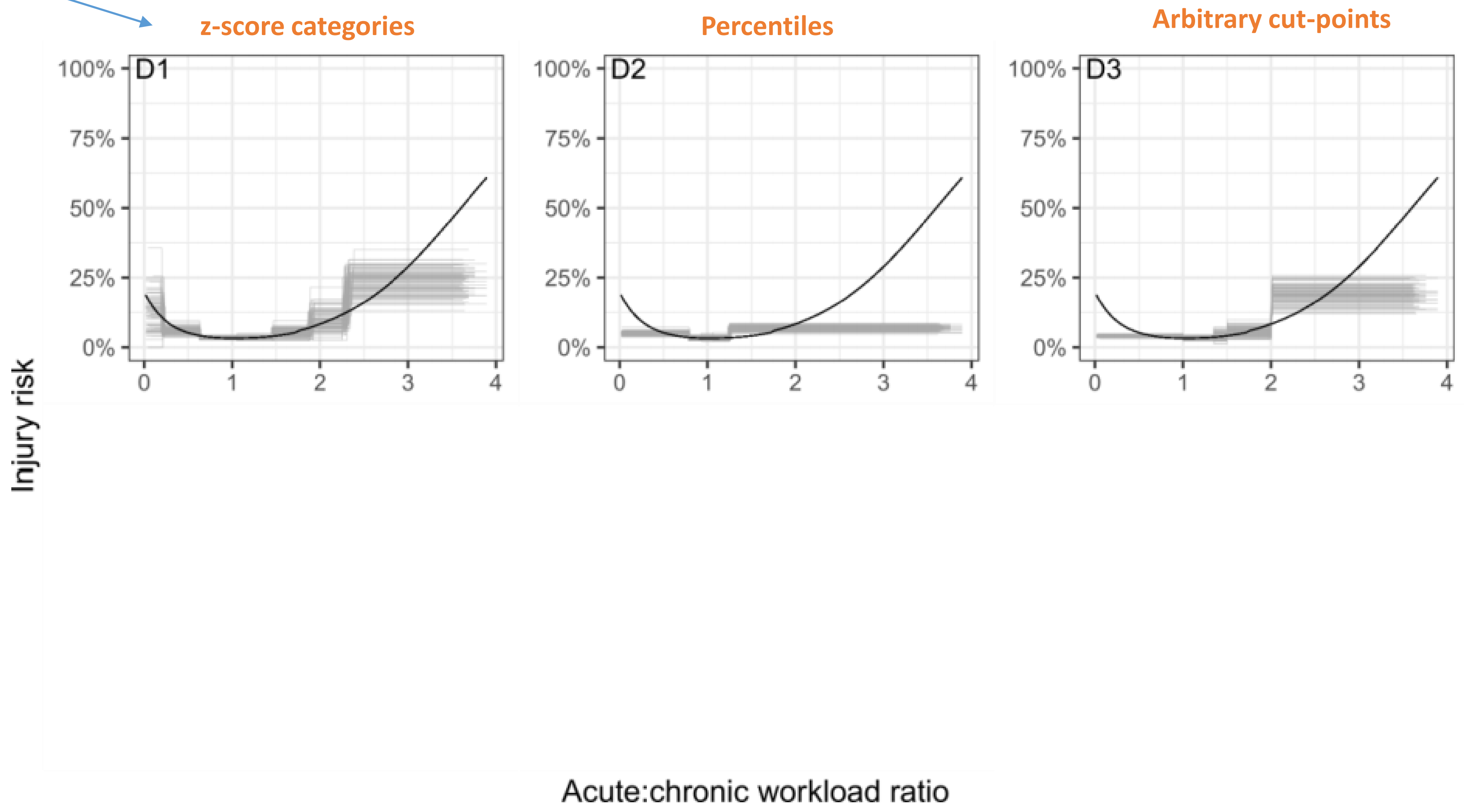
z-score categories
Percentiles
D1
D2
100%
75%
50%
25%
0%
0
1
2
3
4
Injury risk
Acute:chronic workload ratio

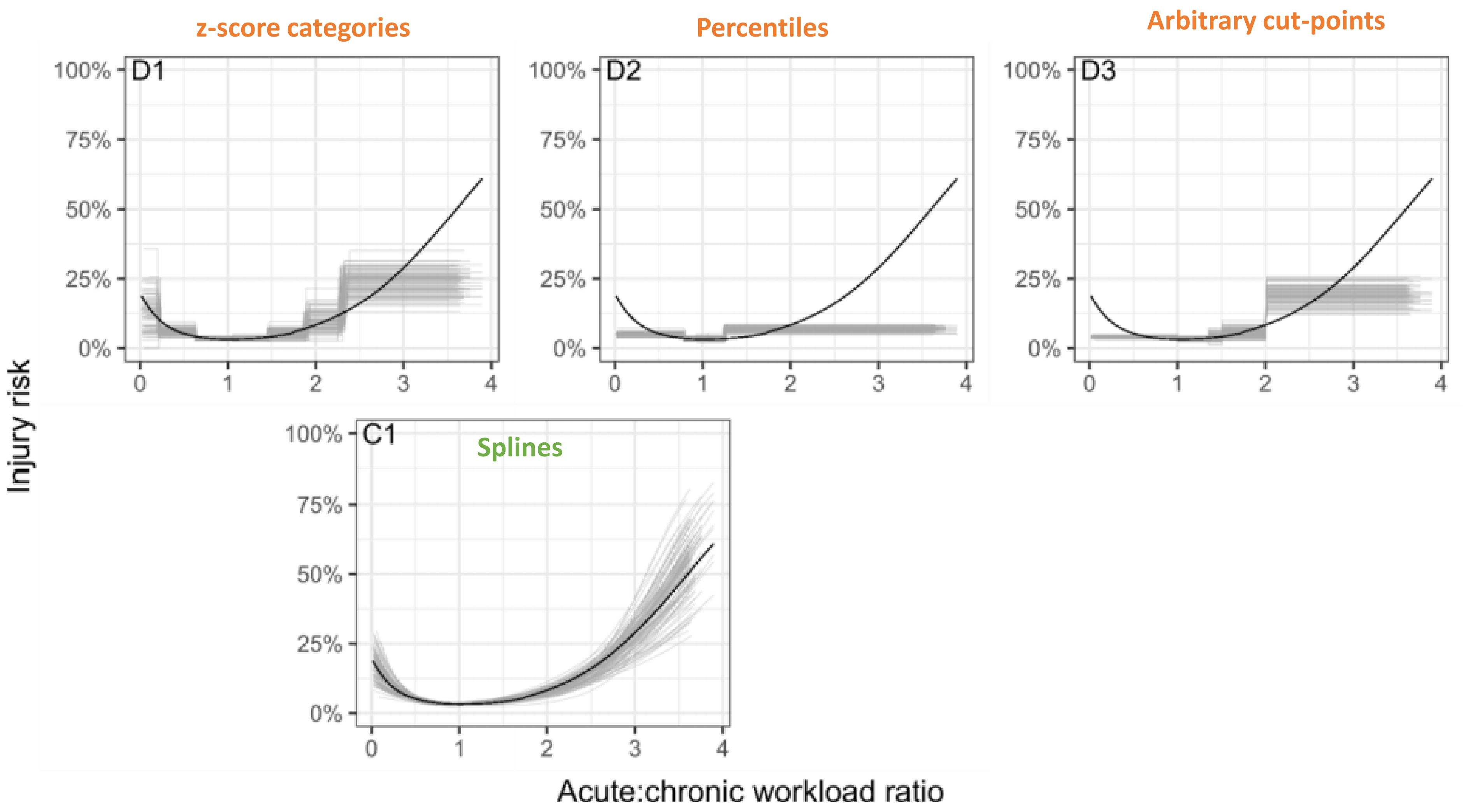


z-score categories
Percentiles
Arbitrary cut-points
D1
D2
D3
100%
75%
50%
25%
0%
0
1
2
3
4
Injury risk
Acute:chronic workload ratio

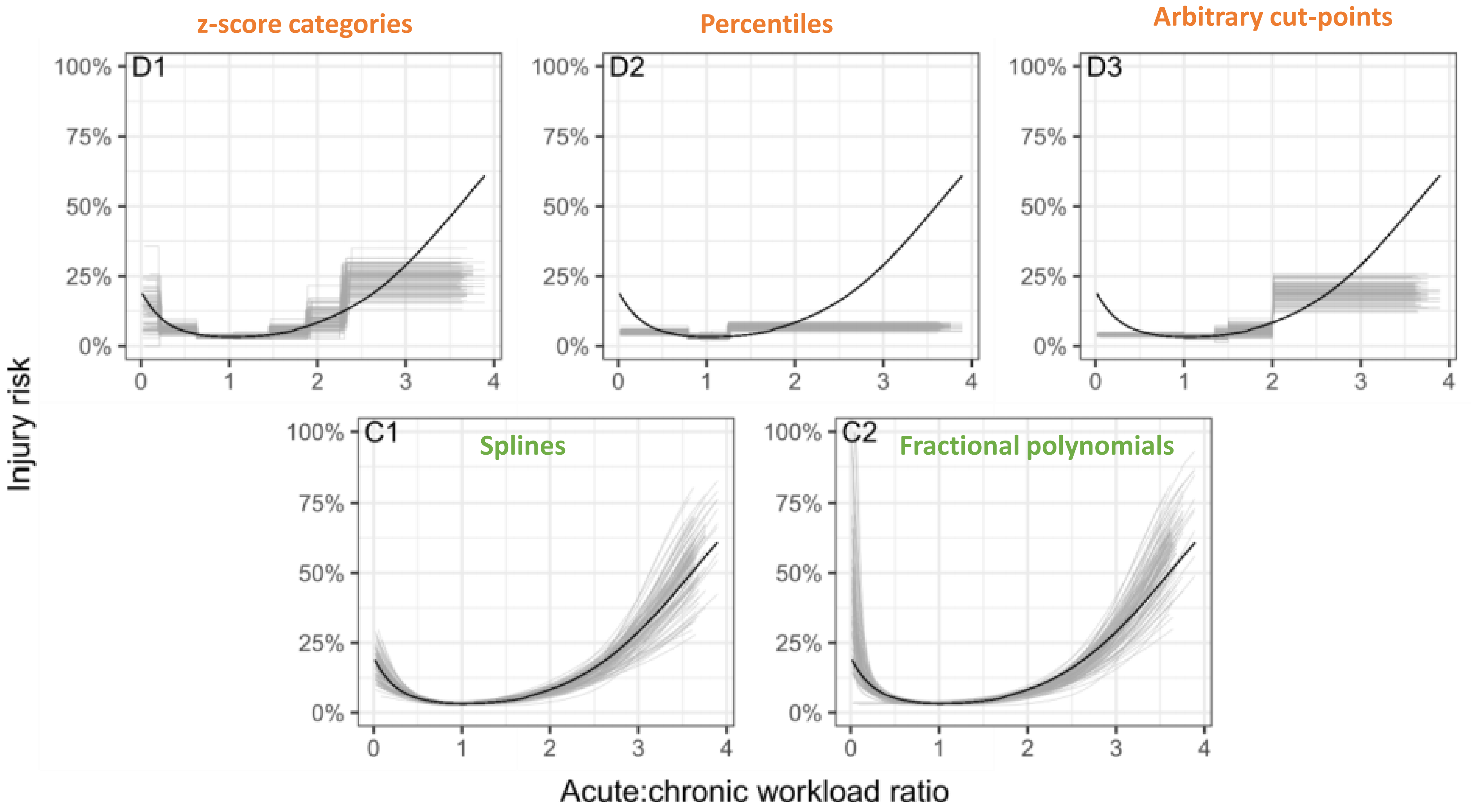
Discretisation forces the models to try and fit an **unrealistic** step profile

All of these 3 have been used in previous studies

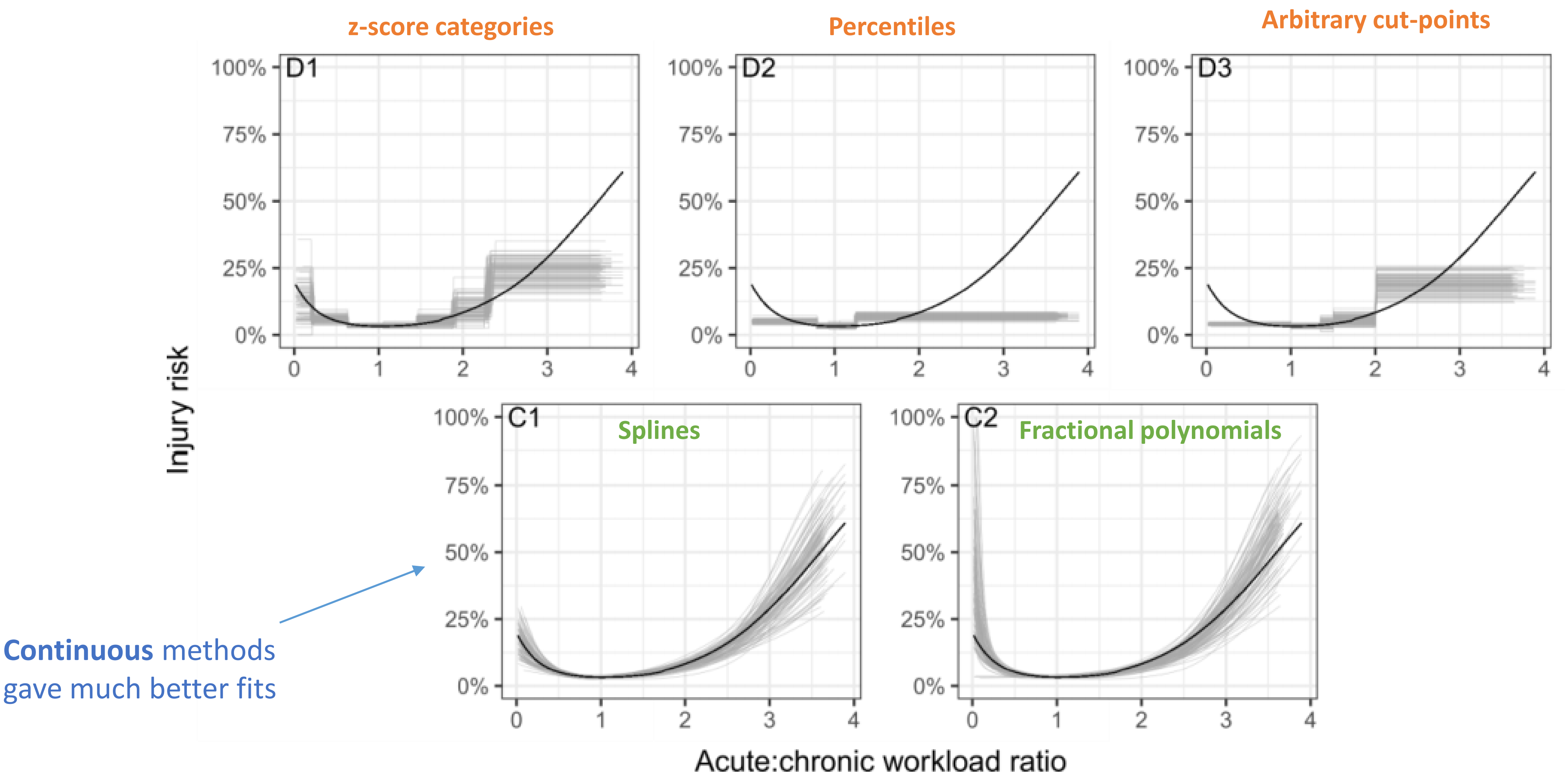




z-score categories
Percentiles
Arbitrary cut-points
D1
D2
D3
C1
Splines
100%
75%
50%
25%
0%
0
1
2
3
4
Injury risk
Acute:chronic workload ratio



z-score categories
Percentiles
Arbitrary cut-points
D1
D2
D3
C1
C2
Splines
Fractional polynomials
100%
75%
50%
25%
0%
0
1
2
3
4
Injury risk
Acute:chronic workload ratio


z-score categories
Percentiles
Arbitrary cut-points
D1
D2
D3
C1
Splines
C2
Fractional polynomials
100%
75%
50%
25%
0%
0
1
2
3
4
Injury risk
Acute:chronic workload ratio
Continuous methods gave much better fits

Take home message 1

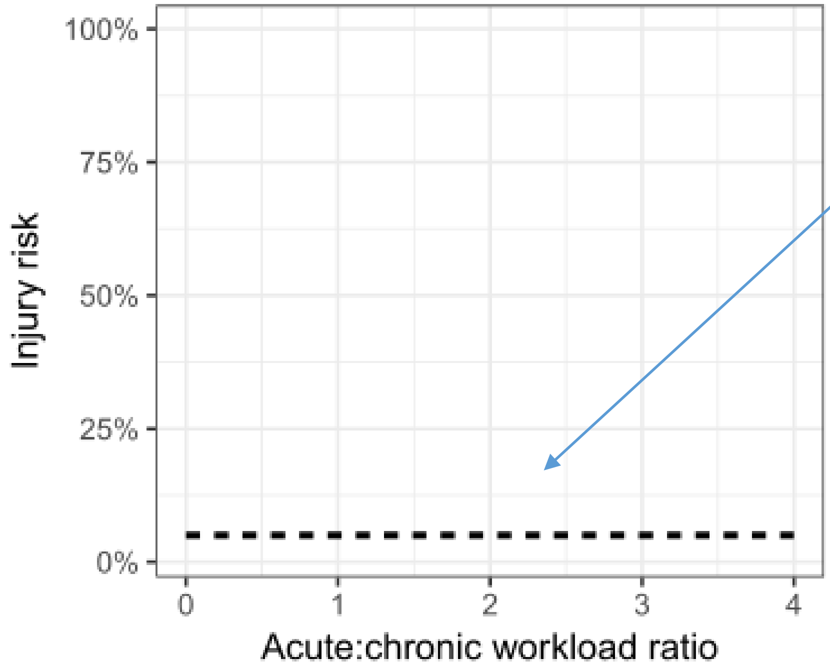
Discretisation can **hide** the real relationships in your data

Take home message 1

Discretisation can **hide** the real relationships in your data

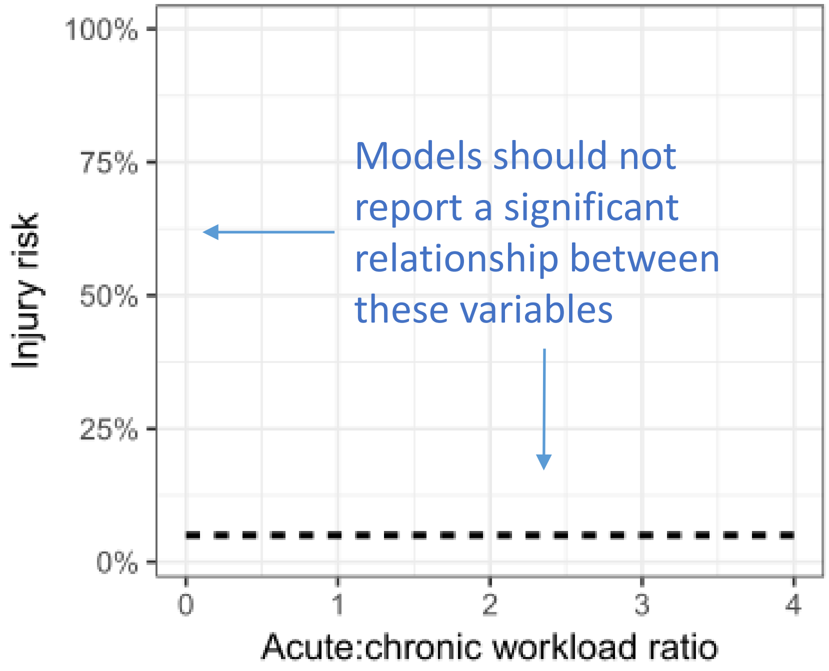
[don’t waste all of your hard earned data by chopping up your variables]

Scenario 2: Flat risk



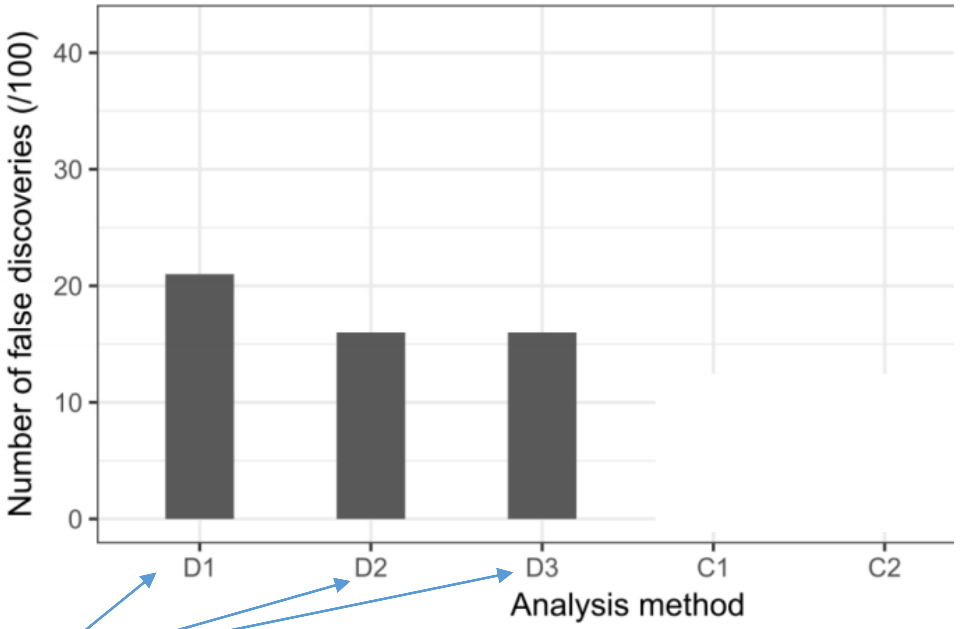
Represents scenario where workload has **no influence** on injury risk

Scenario 2: Flat risk

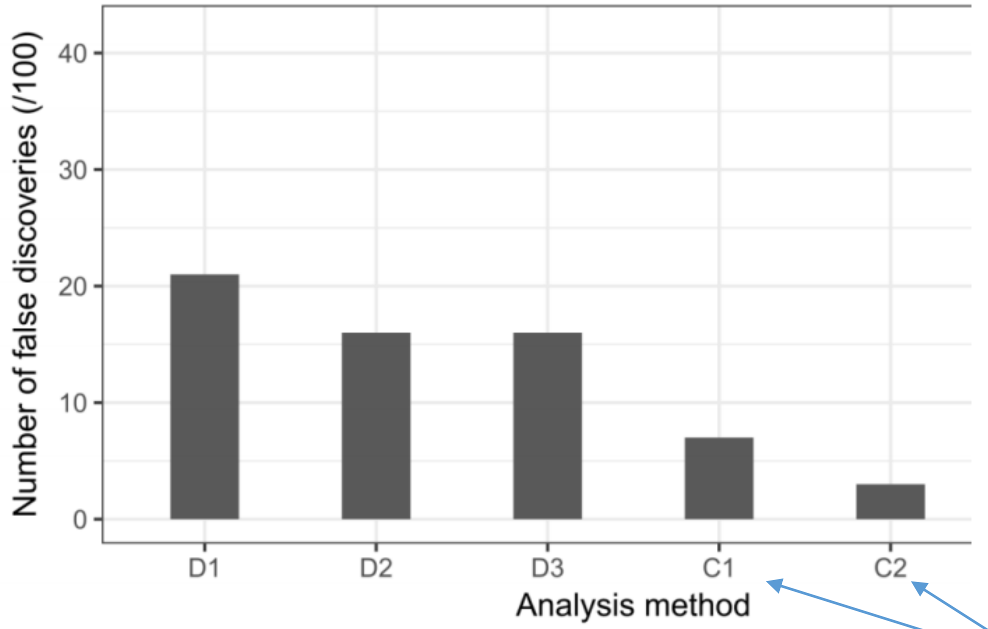


Results

What fraction of the 100 simulated studies find a significant result?

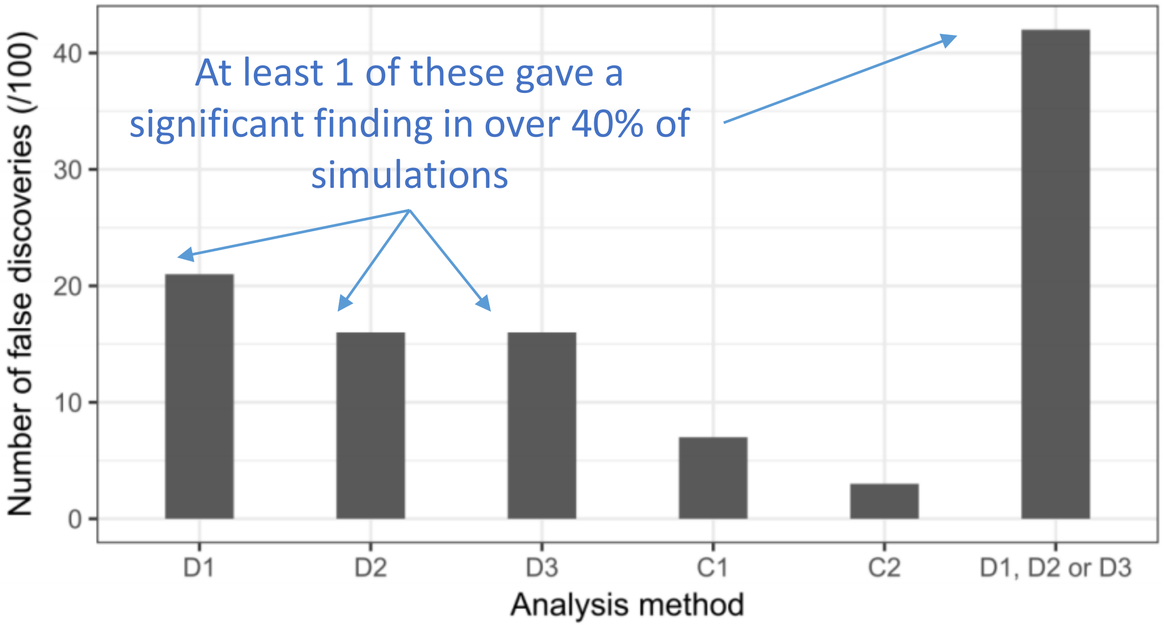


Discrete models had **high false discovery rates** (10-20%)



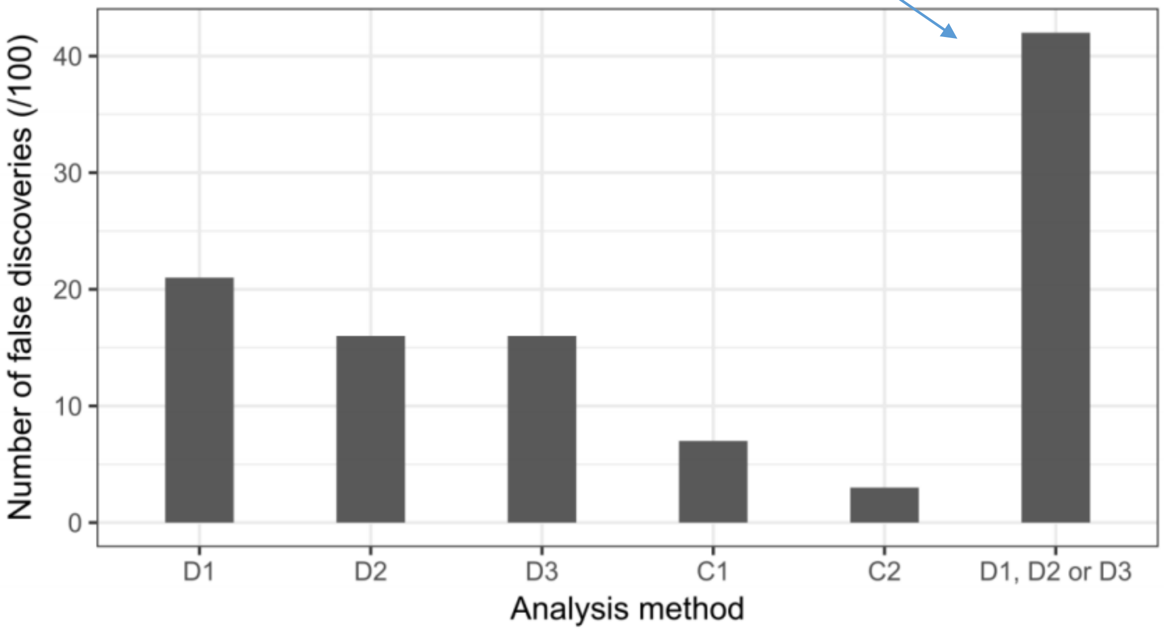
Continuous models were better

(remember around 5% is expected)



At least 1 of these gave a significant finding in over 40% of simulations
Number of false discoveries (/100)
40
30
20
10
0
D1
D2
D3
C1
C2
D1, D2 or D3
Analysis method

If you try a few binning methods I think you are nearly **guaranteed** of getting a significant result (even if there is **explicitly nothing**)



Take home message 2

Discretisation can **increase** the
false positive rate

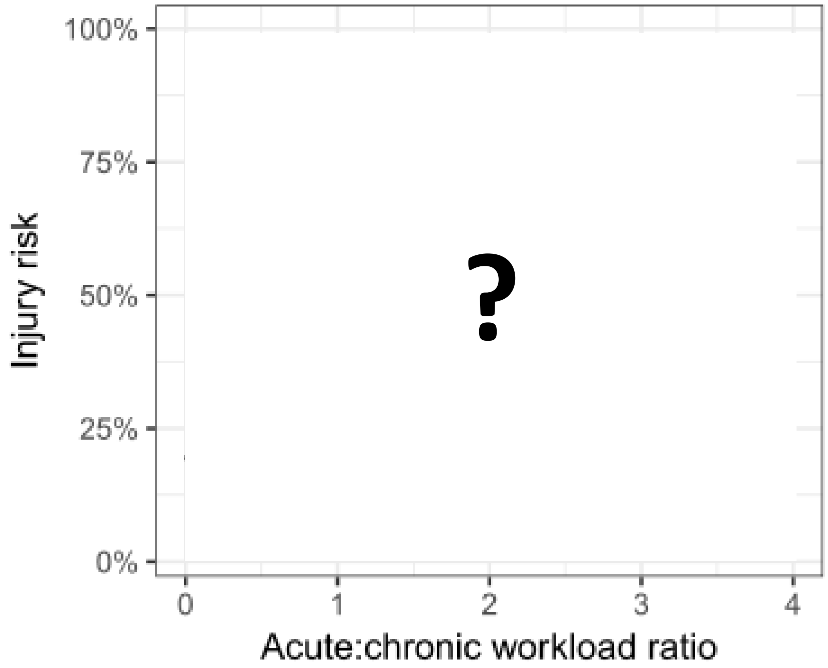
Take home message 2

Discretisation can **increase** the
false positive rate

[don’t fool yourself by chopping up your variables]

But in practice we don't know the true risk shape

(how can we tell which model is best?)



Typical evaluation metrics

- Sensitivity
- Specificity
- Likelihood ratio
- ROC curves and AUC

Typical evaluation metrics

- Sensitivity
- Specificity
- Likelihood ratio
- ROC curves and AUC

These all rely on **discretisation** of **probabilities**

Probabilities are ***continuous***

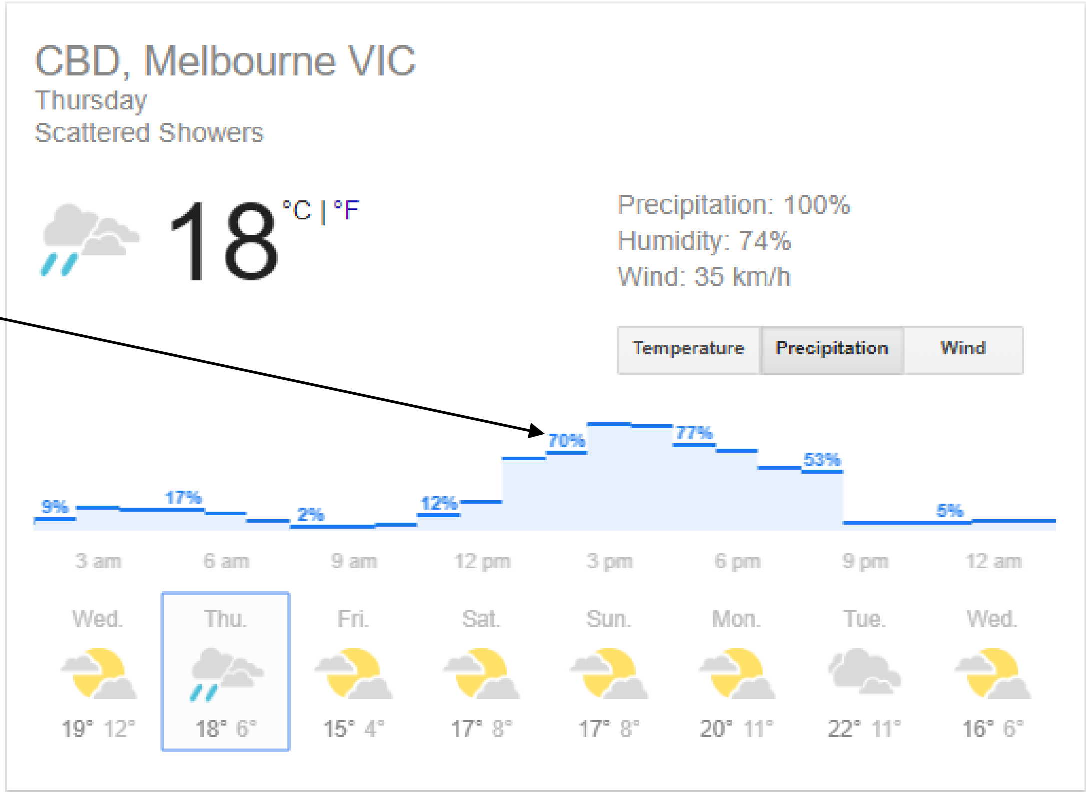
Typical evaluation metrics

Instead of discrete thresholds –
we should be looking at **calibration** and employing **probabilistic reasoning**

Instead of discrete thresholds –
we should be looking at **calibration** and employing **probabilistic reasoning**

What is this? I want a decision rule

They give you a [continuous] probability and let you decide how **much risk you're willing to accept**

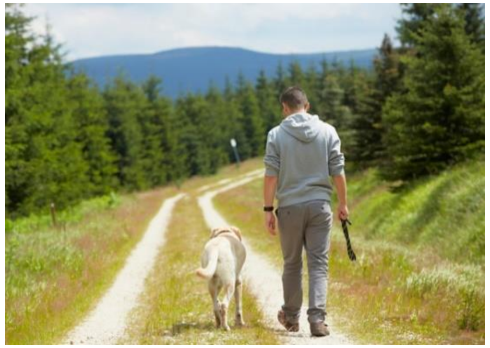


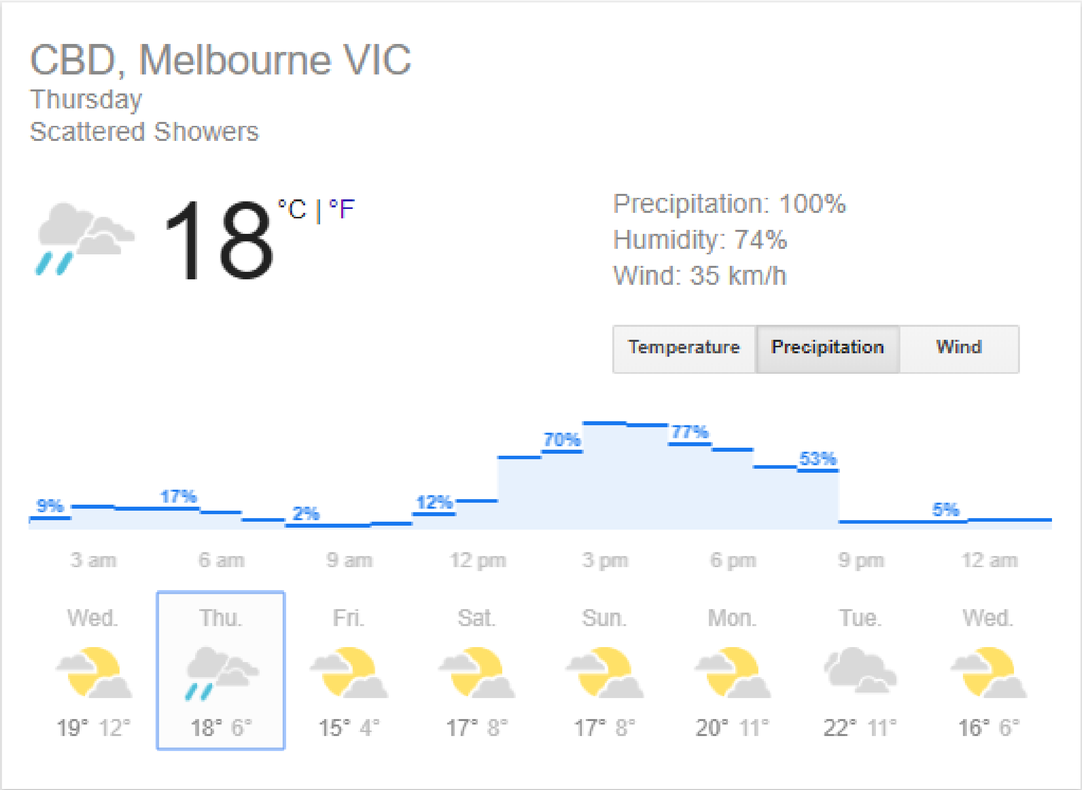
?


CBD, Melbourne VIC
Thursday
Scattered Showers
18 °C | °F
Precipitation: 100%
Humidity: 74%
Wind: 35 km/h
Temperature
Precipitation
Wind
9%
17%
2%
12%
70%
77%
53%
5%
3 am
6 am
9 am
12 pm
3 pm
6 pm
9 pm
12 am
Wed. 19° 12°
Thu. 18° 6°
Fri. 15° 4°
Sat. 17° 8°
Sun. 17° 8°
Mon. 20° 11°
Tue. 22° 11°
Wed. 16° 6°


10

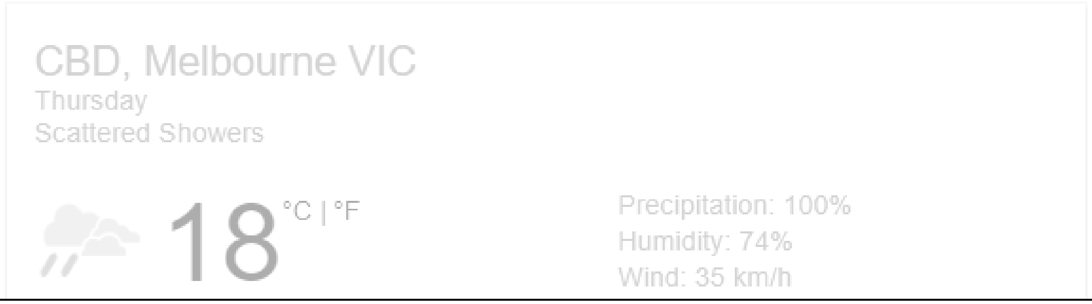
?




CBD, Melbourne VIC
Thursday
Scattered Showers
18 °C | °F
Precipitation: 100%
Humidity: 74%
Wind: 35 km/h
Temperature
Precipitation
Wind
9%
17%
2%
12%
70%
77%
53%
5%
3 am
6 am
9 am
12 pm
3 pm
6 pm
9 pm
12 am
Wed.
19° 12°
Thu.
18° 6°
Fri.
15° 4°
Sat.
17° 8°
Sun.
17° 8°
Mon.
20° 11°
Tue.
22° 11°
Wed.
16° 6°

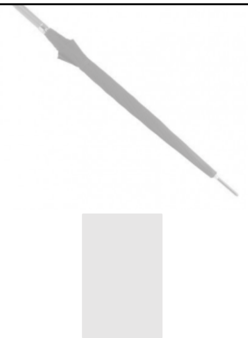

PHYSIO

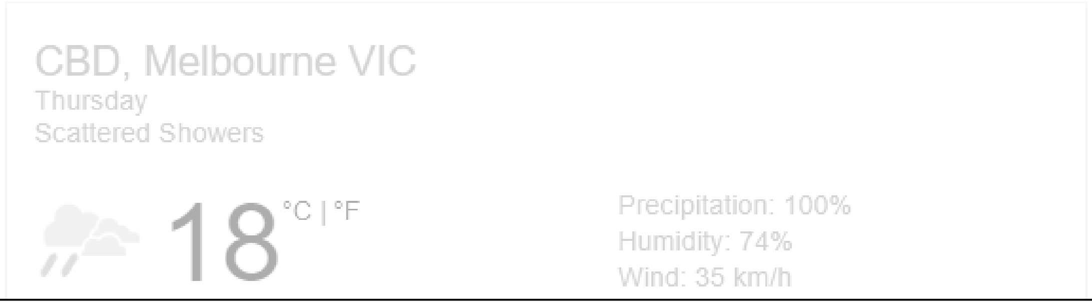
?



Are the probabilities **well calibrated**?

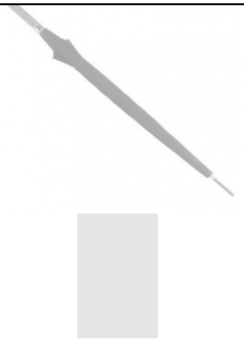
Does it **rain** on approx. 20% of the days the weather model predicts 20% chance of **rain**.





Are the probabilities **well calibrated**?

Did **injuries** occur on approx. 20% of the days the **injury** model predicts 20% chance of **injury**.



What happens if we evaluate models with different metrics?

Area under ROC vs **Brier score**

[discrete thresholds] [calibration]

Number of times selected as best model (/100 simulated studies)

	AUC
D1	
D2	
D3	
C1	
C2	

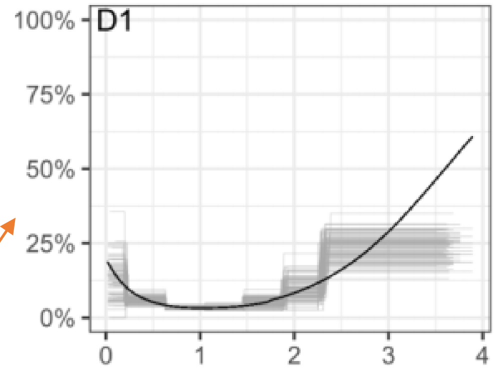
Number of times selected as best model (/100 simulated studies)

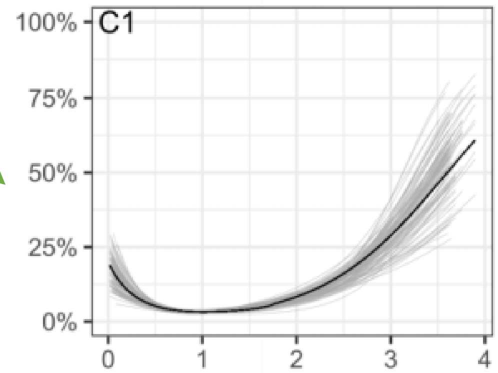
Evaluating using **ROC curves** leads to picking discrete models as best in **38/100** simulations

	AUC
D1	28
D2	2
D3	8
C1	35
C2	27

Number of times selected as best model (/100 simulated studies)	
	AUC
D1	28
D2	2
D3	8
C1	35
C2	27

the **ROC curves** think this is doing better than this





They are assuming a **discrete** Y/N decision is needed

Number of times selected as best model (/100 simulated studies)

	AUC	Brier
D1	28	6
D2	2	0
D3	8	0
C1	35	80
C2	27	14

Probabilistic (continuous) scoring rules hardly ever rank the discrete models as better

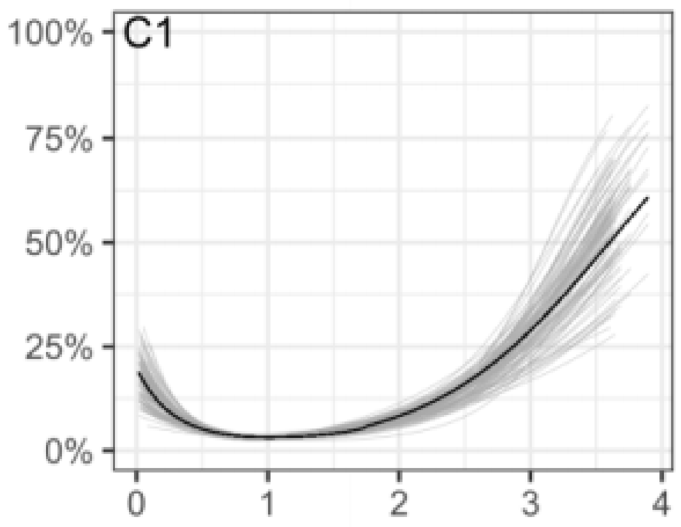
Take home message 3

Avoid discrete scoring metrics

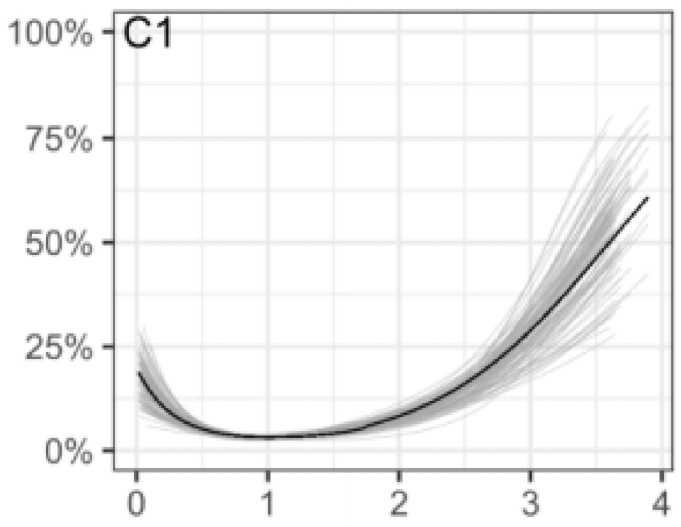
(AUC, Sensitivity, Specificity, Youden Index, …)

for risk probability models

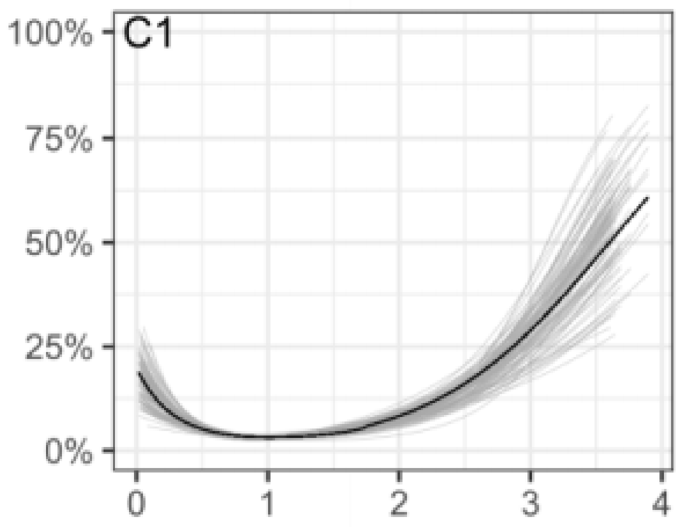
This is important – we may be missing a lot!



- ✓ Near perfect calibration
- ✓ Fits the signal in the data
- ✓ Could be used to manage injury risk



- ✓ Near perfect calibration
- ✓ Fits the signal in the data
- ✓ Could be used to manage injury risk
- **Mean AUC = 0.61**



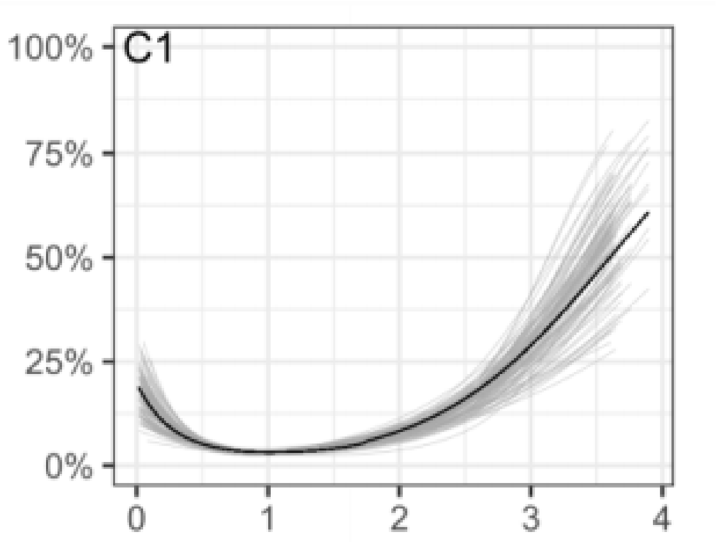
✓ Near perfect calibration
✓ Fits the signal in the data
✓ Could be used to manage injury risk
- **Mean AUC = 0.61**

Predictive Modelling of Training Loads and Injury in Australian Football

Carey, D. L.[1,4], Ong, K.[2], Whiteley, R.[3], Crossley, K. M.[1], Crow, J.[3, 1], Morris, M. E.[1]

Predictive performance was only marginally better than chance for models of non-contact and non-contact time-loss injuries ***(AUC<0.65)***

Injury prediction models built using training load data from a single club showed ***poor ability to predict injuries*** *when tested on previously unseen data*

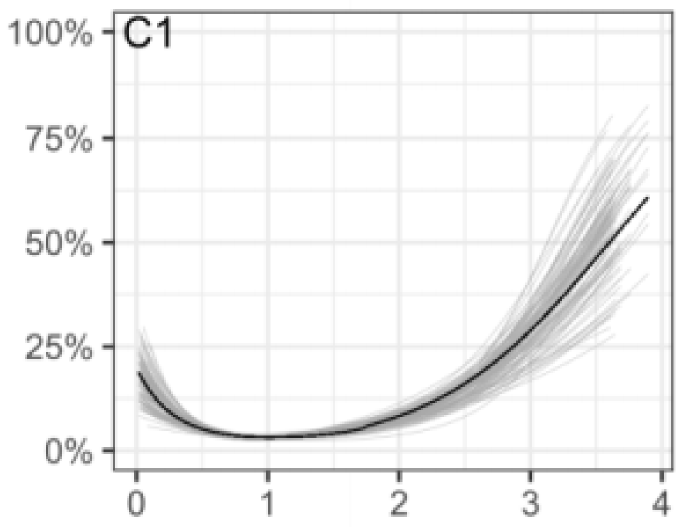


- ✓ Near perfect calibration
- ✓ Fits the signal in the data
- ✓ Could be used to manage injury risk
- **Mean AUC = 0.61**

SCIENCE AND MEDICINE IN FOOTBALL, 2018
VOL. 2, NO. 2, 108–114
https://doi.org/10.1080/24733938.2018.1429014
Routledge
Taylor & Francis Group
ARTICLE
Check for updates

Despite association, the acute:chronic work load ratio does not predict non-contact injury in elite footballers

The ROC curve (Figure 1), the values AUC (90% CI) and the J for each load marker (Table 2) showed ***poor predictive ability*** *of injury* ***(AUC: 0.55–0.60)***

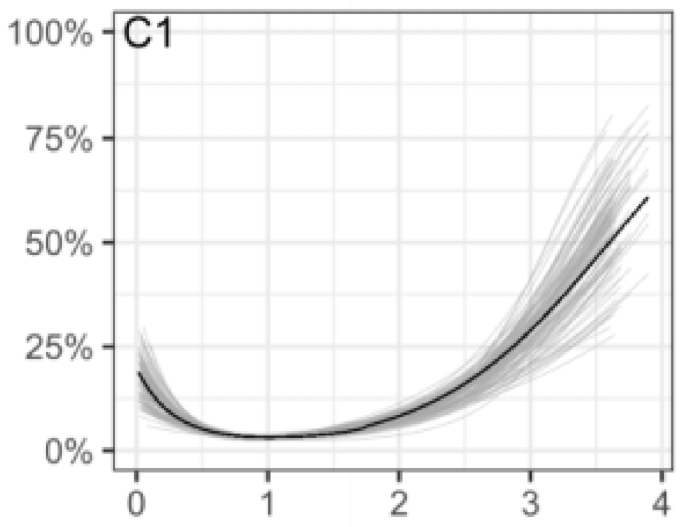


✓ Near perfect calibration
✓ Fits the signal in the data
✓ Could be used to manage injury risk
- **Mean AUC = 0.61**

Workload and non-contact injury incidence in elite football players competing in European leagues

*The **AUC** were **0.56** (4-weeks absolute workload), **0.56** (3-weeks), **0.54** (2-weeks) and **0.53** (1-week), respectively*

No A:C workload combination was appropriate to predict injury



- ✓ Near perfect calibration
- ✓ Fits the signal in the data
- ✓ Could be used to manage injury risk
- **Mean AUC = 0.61**

Section: Original Investigation

Article Title: Greater Association of Relative Thresholds Than Absolute Thresholds With Noncontact Lower-Body Injury in Professional Australian Rules Footballers: Implications For Sprint Monitoring

*Model **accuracy** for all workload thresholds and training variables were **classed as low (AUC = 0.48-0.61)**.*

All **TRUE**

- Training load – injury models typically have low AUC
  - Bad at predicting yes/no injury
- But that's **not** what we should be focussing on

Predictive Modelling of Training Loads and Injury in Australian Football

Carey, D. L.[1,4], Ong, K.[2], Whiteley, R.[3], Crossley, K. M.[1], Crow, J.[3,1], Morris, M. E.[1]

SCIENCE AND MEDICINE IN FOOTBALL, 2018

Routledge Taylor & Francis Group

Check for updates

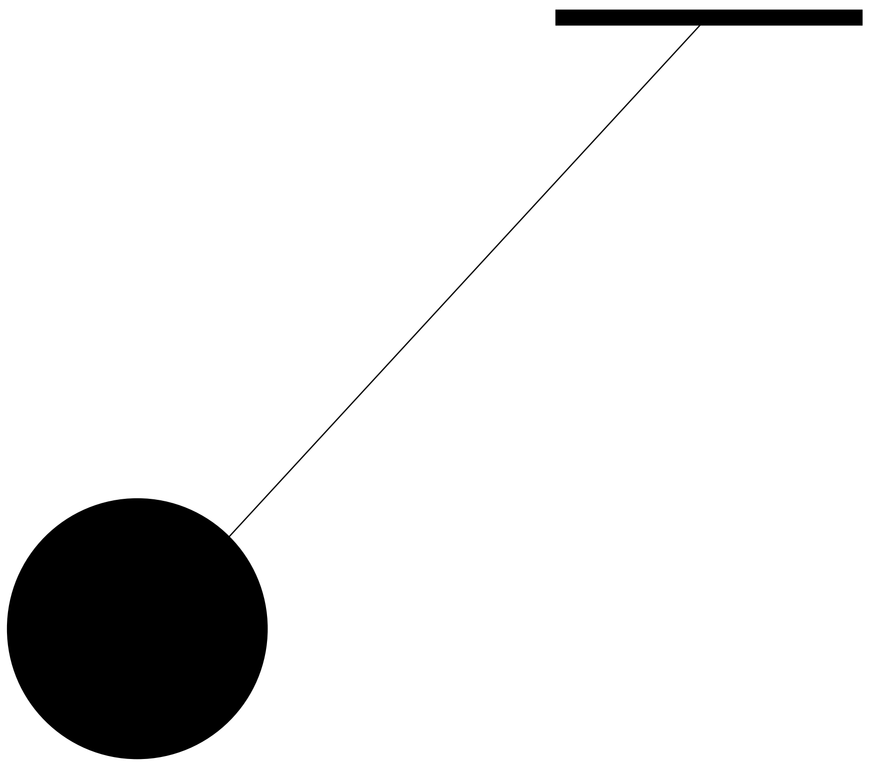
Despite association, the acute:chronic work load ratio does not predict non-contact injury in elite footballers

Section: Original Investigation

Article Title: Greater Association of Relative Thresholds Than Absolute Thresholds With Noncontact Lower-Body Injury in Professional Australian Rules Footballers: Implications For Sprint Monitoring

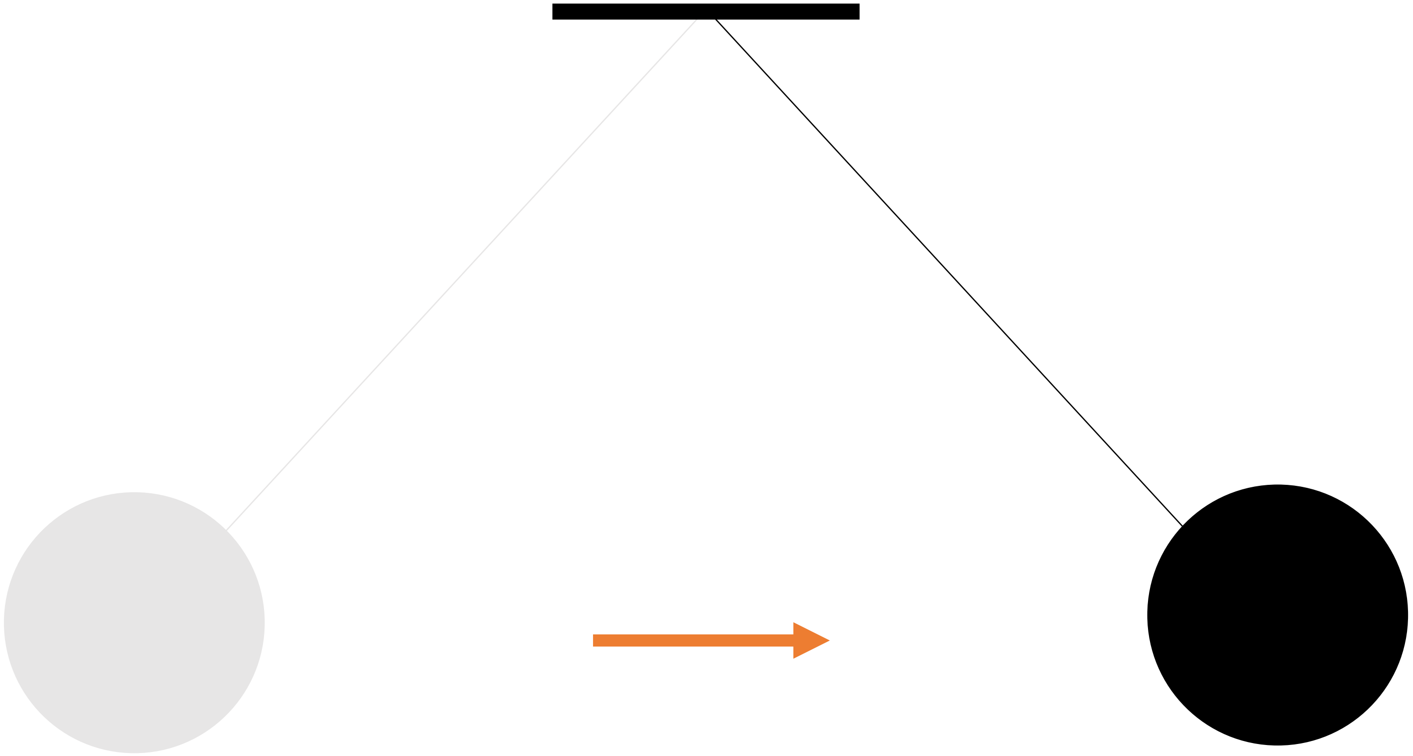
Workload and non-contact injury incidence in elite football players competing in European leagues

Research Pendulum



- Too many researcher degrees of freedom
- Very high risk of false positive results
- Inflated claims of prediction

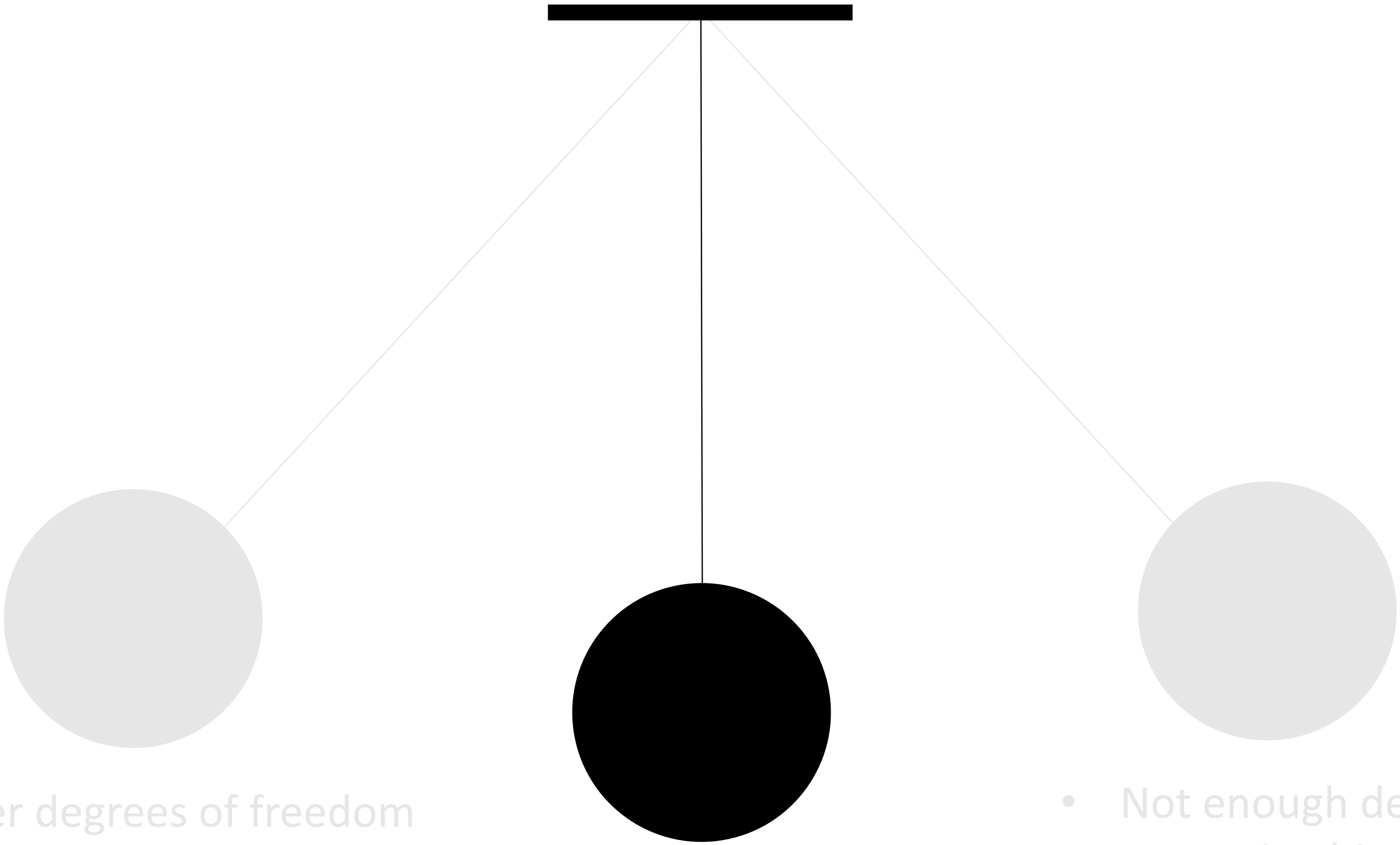
Research Pendulum



- To many researcher degrees of freedom
- Very high risk of false positive results
- Inflated claims of prediction

- Not enough degrees of freedom
- Evaluating binary classification performance (Sens, Spec, ROC)
- Showing what the model can't do

Research Pendulum



- To many researcher degrees of freedom
- Very high risk of false positive
- Inflated claims of prediction

- Be careful with choice of metrics (ACWR issues)
- Don’t discretise
- Don’t assume linear
- Don’t test for binary prediction
- Are the probability estimates useful?
- <u>Simplify</u>

- Not enough degrees of freedom
- ting binary classification
- mance (Sens, Spec, ROC)
- ıg what the model can’t do

ELSEVIER

Journal of Clinical Epidemiology

Volume 110, June 2019, Pages 12-22

Review

A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models

Evangelia Christodoulou [a], Jie Ma [b], Gary S. Collins [b, c], Ewout W. Steyerberg [d], Jan Y. Verbakel [a, e, f], Ben Van Calster [a, d]

To summarise:

If you **discretise**:

If you **discretise**:

- Increase risk of finding nothing when there is something there ($\uparrow$ false negatives)

- Increase risk of finding something if there is nothing there ($\uparrow$ false positives)

- Risk choosing the wrong model

Supplementary message:

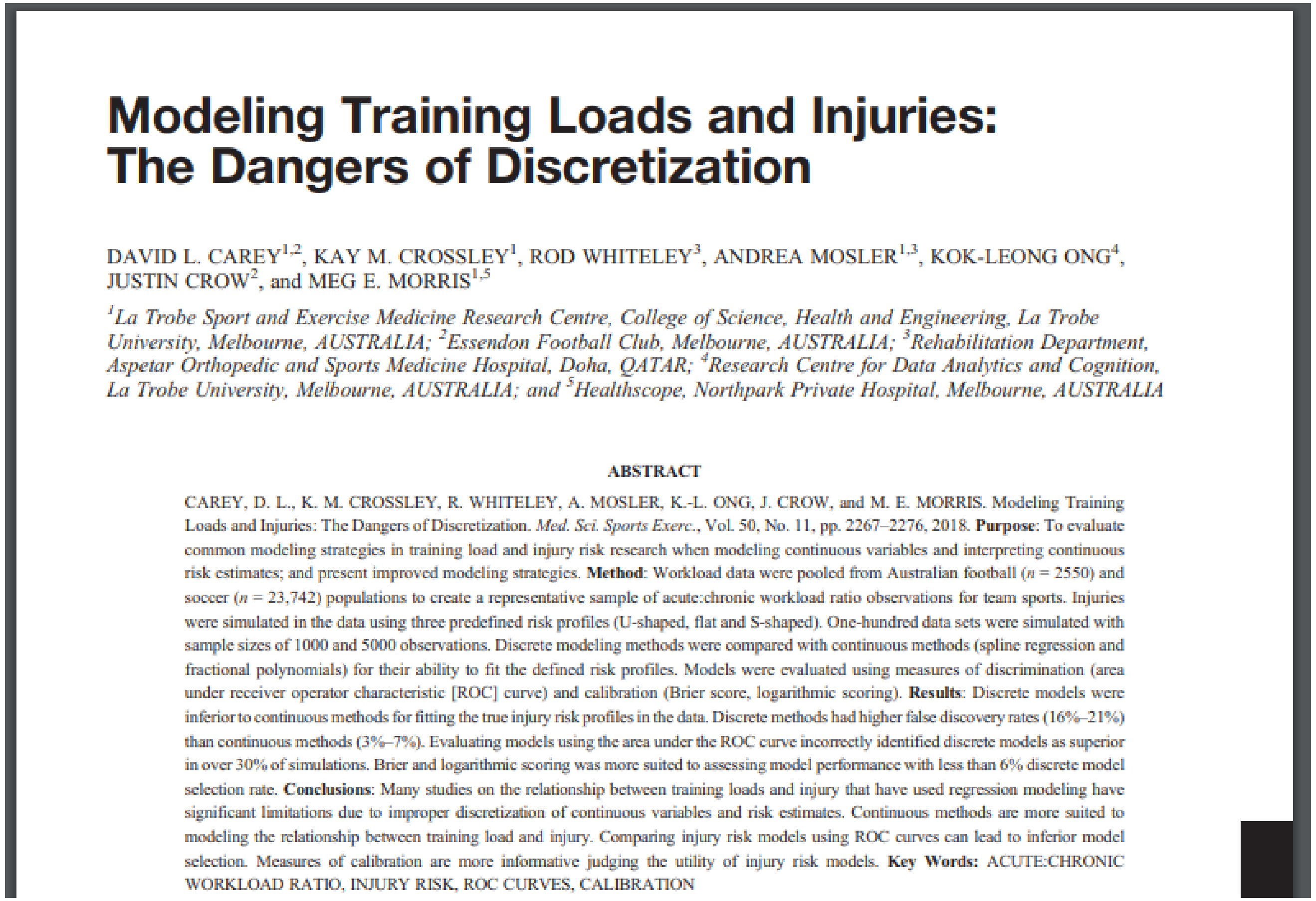
These findings apply to **all continuous variables**

[nothing special about training load]

Length, strength, weight, height, time, speed, angle, …

Interested? @dlcarey88

- Paper in MSSE
- Also examines issues with **repeated measures**
- Supplementary **R code** online



Modeling Training Loads and Injuries: The Dangers of Discretization

DAVID L. CAREY[1,2], KAY M. CROSSLEY[1], ROD WHITELEY[3], ANDREA MOSLER[1,3], KOK-LEONG ONG[4], JUSTIN CROW[2], and MEG E. MORRIS[1,5]

[1]La Trobe Sport and Exercise Medicine Research Centre, College of Science, Health and Engineering, La Trobe University, Melbourne, AUSTRALIA; [2]Essendon Football Club, Melbourne, AUSTRALIA; [3]Rehabilitation Department, Aspetar Orthopedic and Sports Medicine Hospital, Doha, QATAR; [4]Research Centre for Data Analytics and Cognition, La Trobe University, Melbourne, AUSTRALIA; and [5]Healthscope, Northpark Private Hospital, Melbourne, AUSTRALIA

ABSTRACT

CAREY, D. L., K. M. CROSSLEY, R. WHITELEY, A. MOSLER, K.-L. ONG, J. CROW, and M. E. MORRIS. Modeling Training Loads and Injuries: The Dangers of Discretization. *Med. Sci. Sports Exerc.*, Vol. 50, No. 11, pp. 2267–2276, 2018. **Purpose**: To evaluate common modeling strategies in training load and injury risk research when modeling continuous variables and interpreting continuous risk estimates; and present improved modeling strategies. **Method**: Workload data were pooled from Australian football (n = 2550) and soccer (n = 23,742) populations to create a representative sample of acute:chronic workload ratio observations for team sports. Injuries were simulated in the data using three predefined risk profiles (U-shaped, flat and S-shaped). One-hundred data sets were simulated with sample sizes of 1000 and 5000 observations. Discrete modeling methods were compared with continuous methods (spline regression and fractional polynomials) for their ability to fit the defined risk profiles. Models were evaluated using measures of discrimination (area under receiver operator characteristic [ROC] curve) and calibration (Brier score, logarithmic scoring). **Results**: Discrete models were inferior to continuous methods for fitting the true injury risk profiles in the data. Discrete methods had higher false discovery rates (16%–21%) than continuous methods (3%–7%). Evaluating models using the area under the ROC curve incorrectly identified discrete models as superior in over 30% of simulations. Brier and logarithmic scoring was more suited to assessing model performance with less than 6% discrete model selection rate. **Conclusions**: Many studies on the relationship between training loads and injury that have used regression modeling have significant limitations due to improper discretization of continuous variables and risk estimates. Continuous methods are more suited to modeling the relationship between training load and injury. Comparing injury risk models using ROC curves can lead to inferior model selection. Measures of calibration are more informative judging the utility of injury risk models. **Key Words:** ACUTE:CHRONIC WORKLOAD RATIO, INJURY RISK, ROC CURVES, CALIBRATION

LA TROBE UNIVERSITY

All kinds of clever

Thank you

- http://www.fharrell.com/2017/01/classification-vs-prediction.html
- Harrell, Frank E., et al. "Regression modelling strategies for improved prognostic prediction." *Statistics in medicine* 3.2 (1984): 143-152.
- Hulin, Billy T., et al. "The acute: chronic workload ratio predicts injury: high chronic workload may decrease injury risk in elite rugby league players." *Br J Sports Med* (2015): bjsports-2015.
- Murray, N. B., et al. "Individual and combined effects of acute and chronic running loads on injury risk in elite Australian footballers." *Scandinavian journal of medicine & science in sports* (2016).
- Carey, David L., et al. "Training loads and injury risk in Australian football—differing acute: chronic workload ratios influence match injury risk." *Br J Sports Med* (2016): bjsports-2016.
- Stares, Jordan, et al. "Identifying high risk loading conditions for in-season injury in elite Australian football players." *Journal of Science and Medicine in Sport* (2017).
- Malone, Shane, et al. "The acute: chonic workload ratio in relation to injury risk in professional soccer." *Journal of science and medicine in sport* 20.6 (2017): 561-565.