Inference with Predicted Data: Examples from Verbal Autopsies and the BMI

Adam Visokay
University of Washington

CENTER for STATISTICS and the SOCIAL SCIENCES

ELDER RESEARCH

DATA SCIENCE · AI · MACHINE LEARNING

TECH TALK — APRIL 19 2025

A bit about me

University of Virginia
B.A. in
History & Economics

Syracuse University
M.A. in
Economics

University of Washington
Ph.D. in
Sociology (current)

Research interests

1. Computational Social Science Methods
    - Text as data
    - The Rashomon Effect (model selection)
    - **Inference on Predicted Data (IPD)**

2. Health
    - Mortality Estimation with Verbal Autopsy
    - Morbidity - Obesity
    - Palliative Care

3. Social construction of categories
4. Science, Technology and Society (STS)

Outline for today

1. Inference with predicted data (IPD)
    Motivation
    Methodology

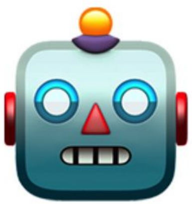
2. Examples:
    a. Verbal Autopsies for cause of death estimation
        NLP prediction models

    b. BMI for obesity research
        Conceptualization vs measurement

3. Looking ahead

Motivation

You use an AI/ML algorithm to make predictions.



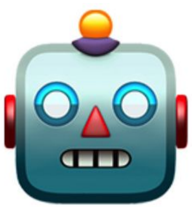
Confusion Matrix

	Actually Positive (1)	Actually Negative (0)
Predicted Positive (1)	True Positives (TPs)	False Positives (FPs)
Predicted Negative (0)	False Negatives (FNs)	True Negatives (TNs)

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)^2}{n}}$$

Motivation

You use an AI/ML algorithm to make predictions. Now what?



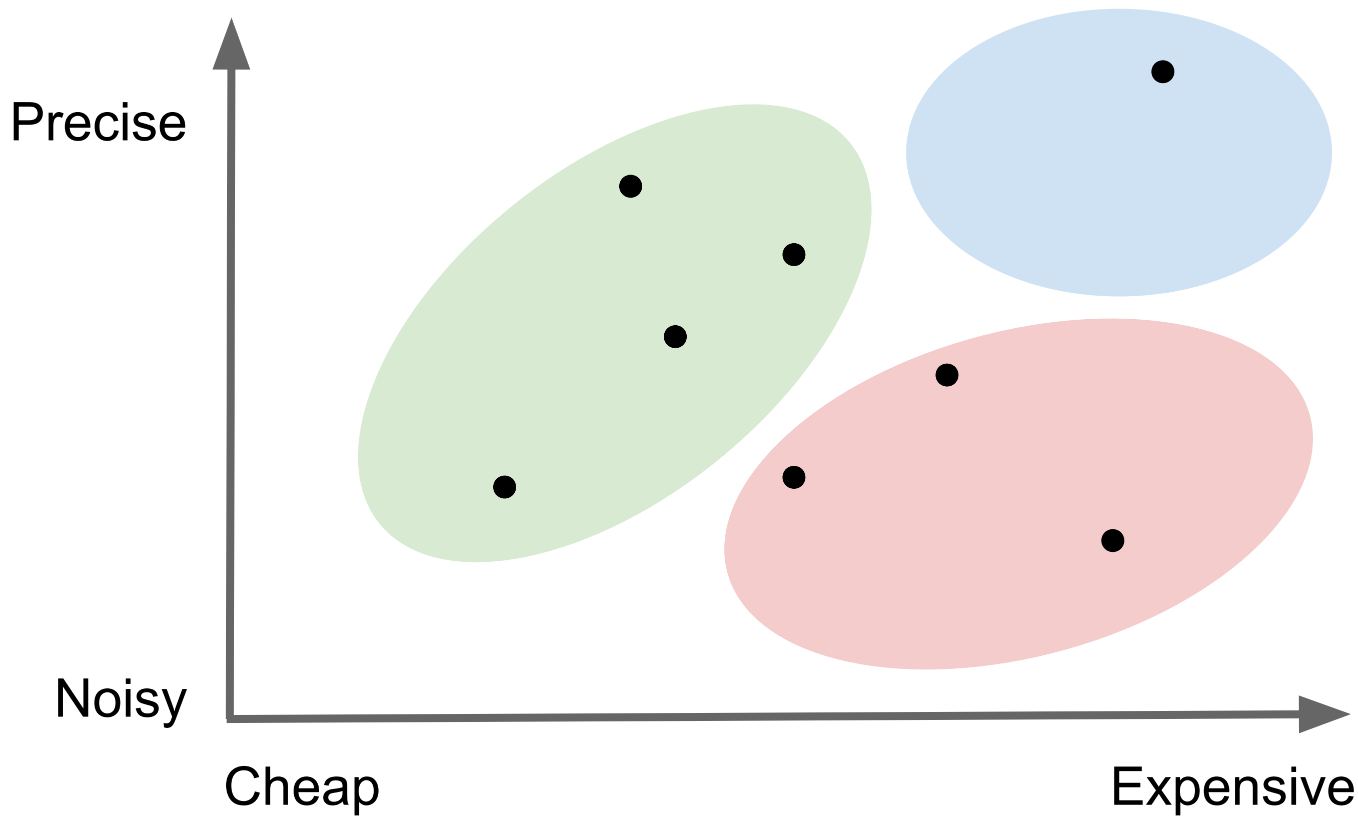
Confusion Matrix

	Actually Positive (1)	Actually Negative (0)
Predicted Positive (1)	True Positives (TPs)	False Positives (FPs)
Predicted Negative (0)	False Negatives (FNs)	True Negatives (TNs)

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)^2}{n}}$$

?

Measurements Vary in Cost and Precision



Example: Global Mortality Estimation 🌎

1. Observe COD directly (🏥)

    Expensive but precise.

2. Predict COD based on symptoms (🤖)

    Cheap but noisy.

Goal: learn association between COD and demographics, X.

Example: Global Mortality Estimation 🌎

1. Observe COD directly (🏥)
   Expensive but precise.
2. Predict COD based on symptoms (🤖)
   Cheap but noisy.

Goal: learn association between COD and demographics, X.

Specify regression with demographics X: [Age, Sex, Race, etc]

🏥 $COD = \beta_1 X + \varepsilon_1$

🤖 $COD = \beta_2 X + \varepsilon_2$

Example: Global Mortality Estimation

1. Observe COD directly ( )

   Expensive but precise.

2. Predict COD based on symptoms ( )

   Cheap but imprecise.

Goal: learn association between COD and demographics, X.

Specify regression with demographics X: [Age, Sex, Race, etc]

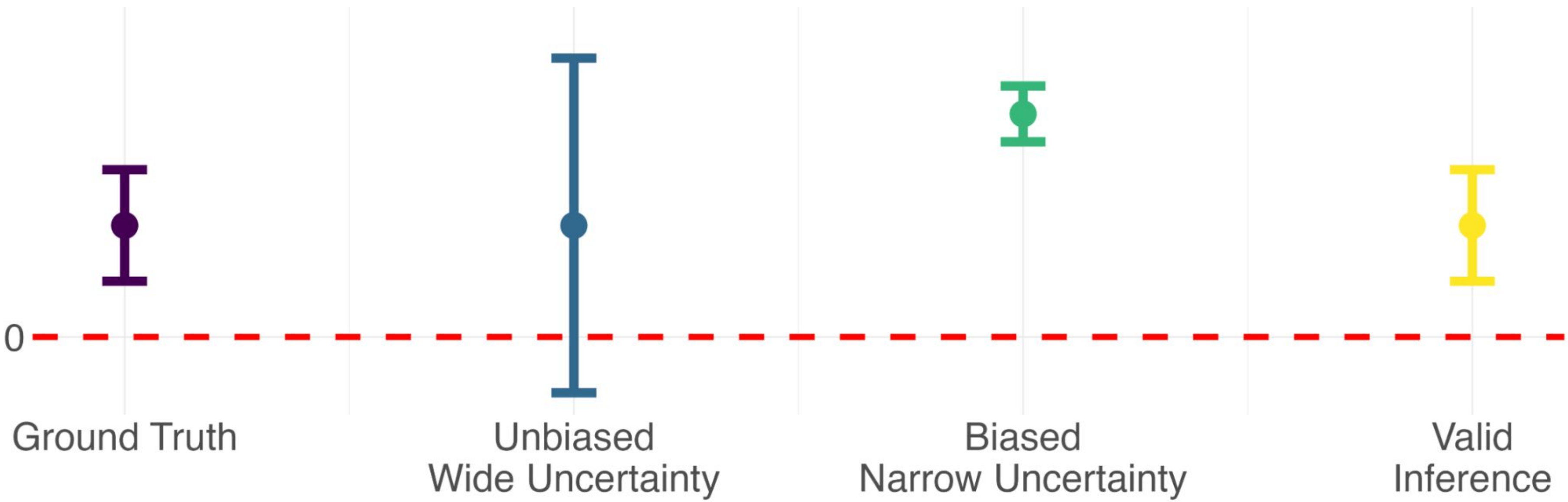
COD = β_1 X + ε_1

COD = β_2 X + ε_2

β_1 and ε_1 are <u>different</u> from β_2 and ε_2

Inference with predicted data (IPD) can have:

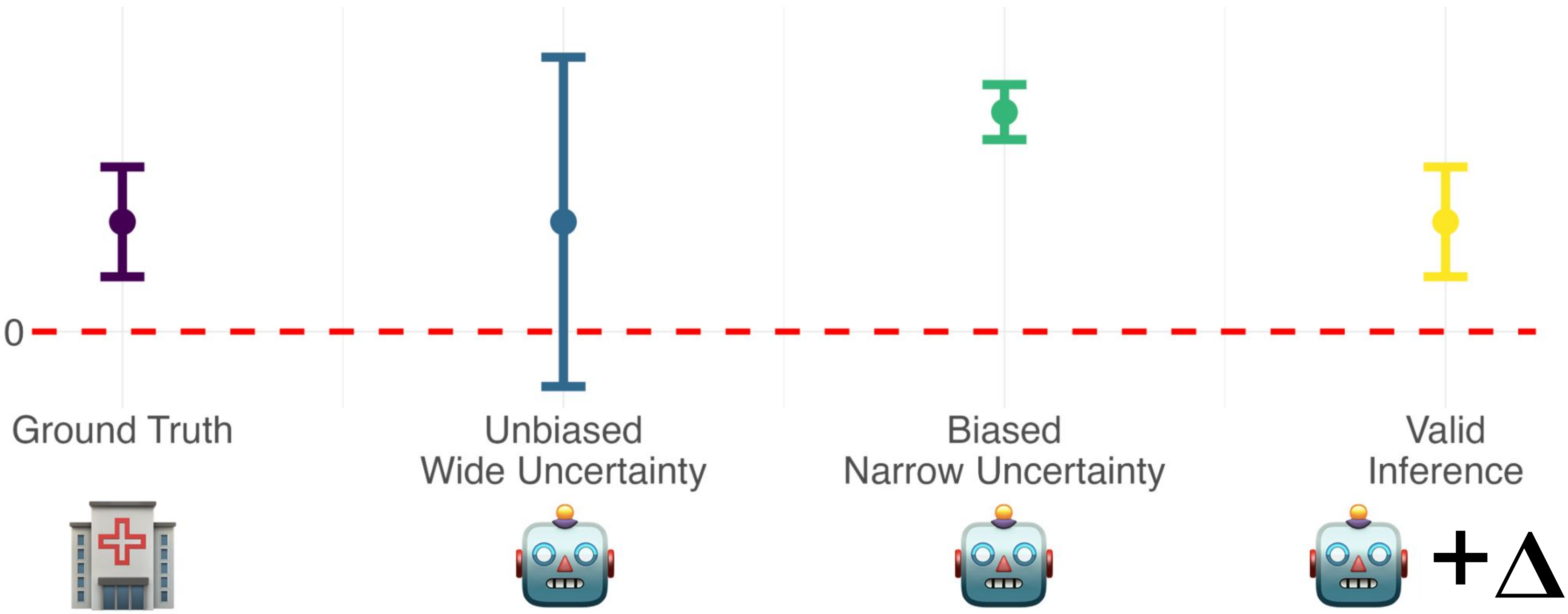
1. Biased estimates
2. Misleading uncertainty



Inference with predicted data (IPD) can have:

1. Biased estimates
2. Misleading uncertainty

Can be fixed with correction factor Δ

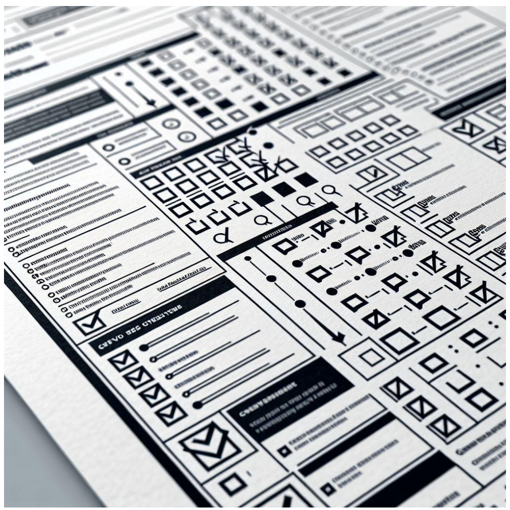


Verbal Autopsy (VA)

Verbal Autopsy (VA)

Interviews with caregivers of the deceased, used to assign COD.

structured questionnaire



free text narrative

UNPROCESSED VA TEXT NARRATIVE
Deceased started to ill while at working place, He came home while experiencing cough with chest pain, difficult in breathing, tiredness and blood vision. The after visited Belfast clinic to get treatment but no improvement. Afterwards deceased complained of stomach pain. Then after experienced diarrhea. He was given traditional medicine but did not change. Afterwards he vomiting worms and diarrhea continued. He continued using traditional medicine and the condition remains the same. Three days before death deceased sneezed a thing like a worm. He died at home and he also experienced hot body. It was examined that his chest and throat developed wounds. Treatment given but no change. His lower lip also had rash that at time chapping and a lot of blood will comes out. After treatment that lip became healed He was taken to traditional healer, but condition unchanged. He was taken Tintswalo hospital, where he was admitted Oxygen supplier was given but he finally passed away on the third day at hospital. A week before death he complained about body pain. At the beginning deceased also had cough and complained of headache during the night only throughout the illness. A month before death he experienced hiccup which continued until death but recurrent, he skips days not defecating When defecate the stool were hard then after yellowish and black few days before death. Deceased also developed ring worms on both checks but healed before death
PROCESSED VA TEXT NARRATIVE
['cough', cough',' chest',' pain',' tiredness',' blood',' vision',' stomach',' pain',' ' vomit',' worms','diarrhea',' sneezed',' worm',' hot',' chest',' throat',' ' lip',' rash',' chapping',' blood',' lip',' pain',' cough',' headache',' hiccup'," defecating',' defecate',' stool',' yellowish',' ring','worms']

Mapundu et al. 2024

Burdensome on respondents (~2hr, repetitive, impersonal).

Verbal Autopsy (VA)

Interviews with caregivers of the deceased, used to assign COD.

structured questionnaire

free text narrative

UNPROCESSED VA TEXT NARRATIVE

Deceased started to ill while at working place, He came home while experiencing cough with chest pain, difficult in breathing, tiredness and blood vision. The after visited Belfast clinic to get treatment but no improvement. Afterwards deceased complained of stomach pain. Then after experienced diarrhea. He was given traditional medicine but did not change. Afterwards he vomiting worms and diarrhea continued. He continued using traditional medicine and the condition remains the same. Three days before death deceased sneezed a thing like a worm. He died at home and he also experienced hot body. It was examined that his chest and throat developed wounds. Treatment given but no change. His lower lip also had rash that at time chapping and a lot of blood will comes out. After treatment that lip became healed He was taken to traditional healer, but condition unchanged. He was taken Tintswalo hospital, where he was admitted Oxygen supplier was given but he finally passed away on the third day at hospital. A week before death he complained about body pain. At the beginning deceased also had cough and complained of headache during the night only throughout the illness. A month before death he experienced hiccup which continued until death but recurrent, he skips days not defecating When defecate the stool were hard then after yellowish and black few days before death. Deceased also developed ring worms on both checks but healed before death

PROCESSED VA TEXT NARRATIVE

['cough', cough',' chest',' pain',' tiredness',' blood',' vision',' stomach',' pain',' ' vomit',' worms','diarrhea',' sneezed',' worm',' hot',' chest',' throat',' ' lip',' rash',' chapping',' blood',' lip',' pain',' cough',' headache',' hiccup','' defecating',' defecate',' stool',' yellowish',' ring','worms']

Mapundu et al. 2024

Burdensome on respondents (~2hr, repetitive, impersonal).

Data

IHME | GHDx

Home Countries Series and Systems Organizations Keywords IHME Data About the GHDx

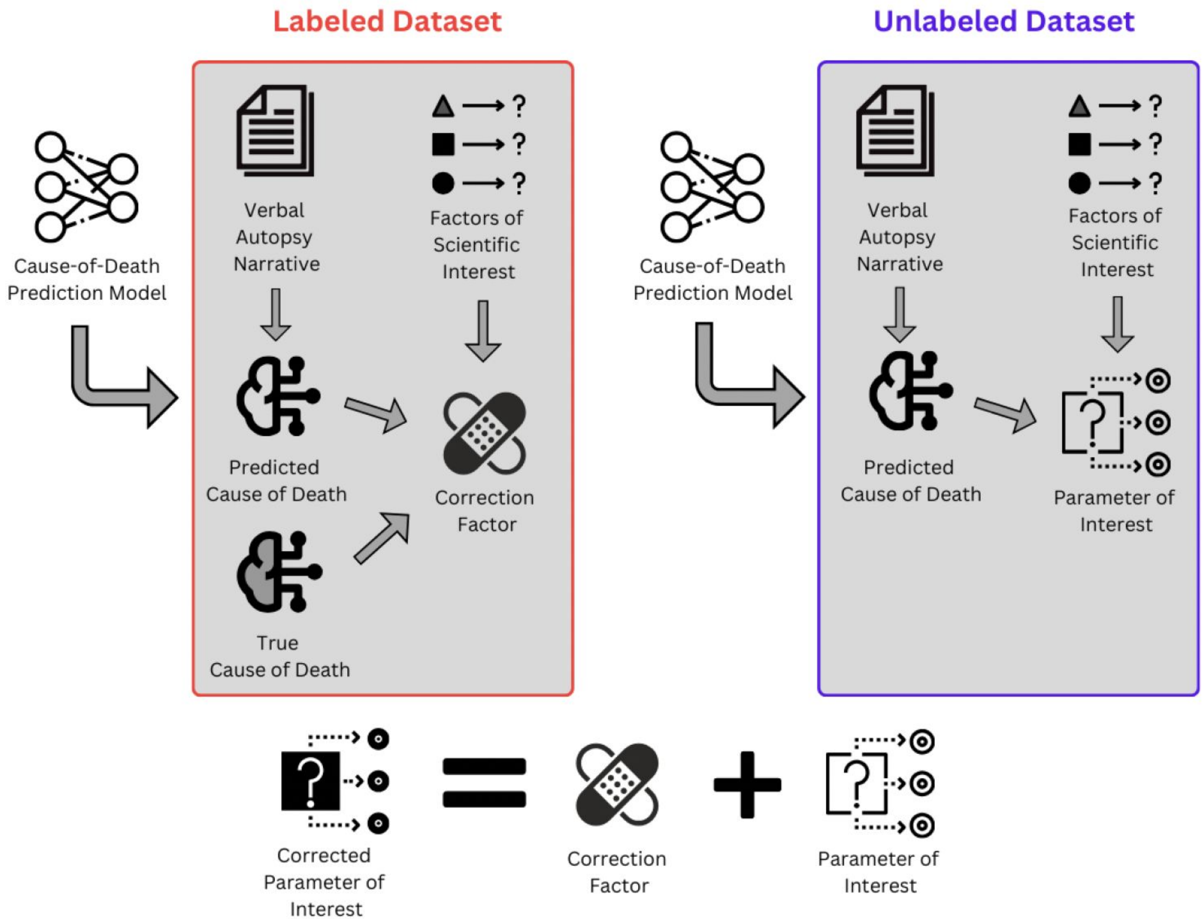
Home > Survey

Population Health Metrics Research Consortium Gold Standard Verbal Autopsy Data 2005-2011

- adult deaths (n=6763)
- both traditional **and** verbal autopsies
- 6 sites, 4 countries
- 5 COD - [*Communicable, Non-communicable, Maternal, AIDS-TB, External*]

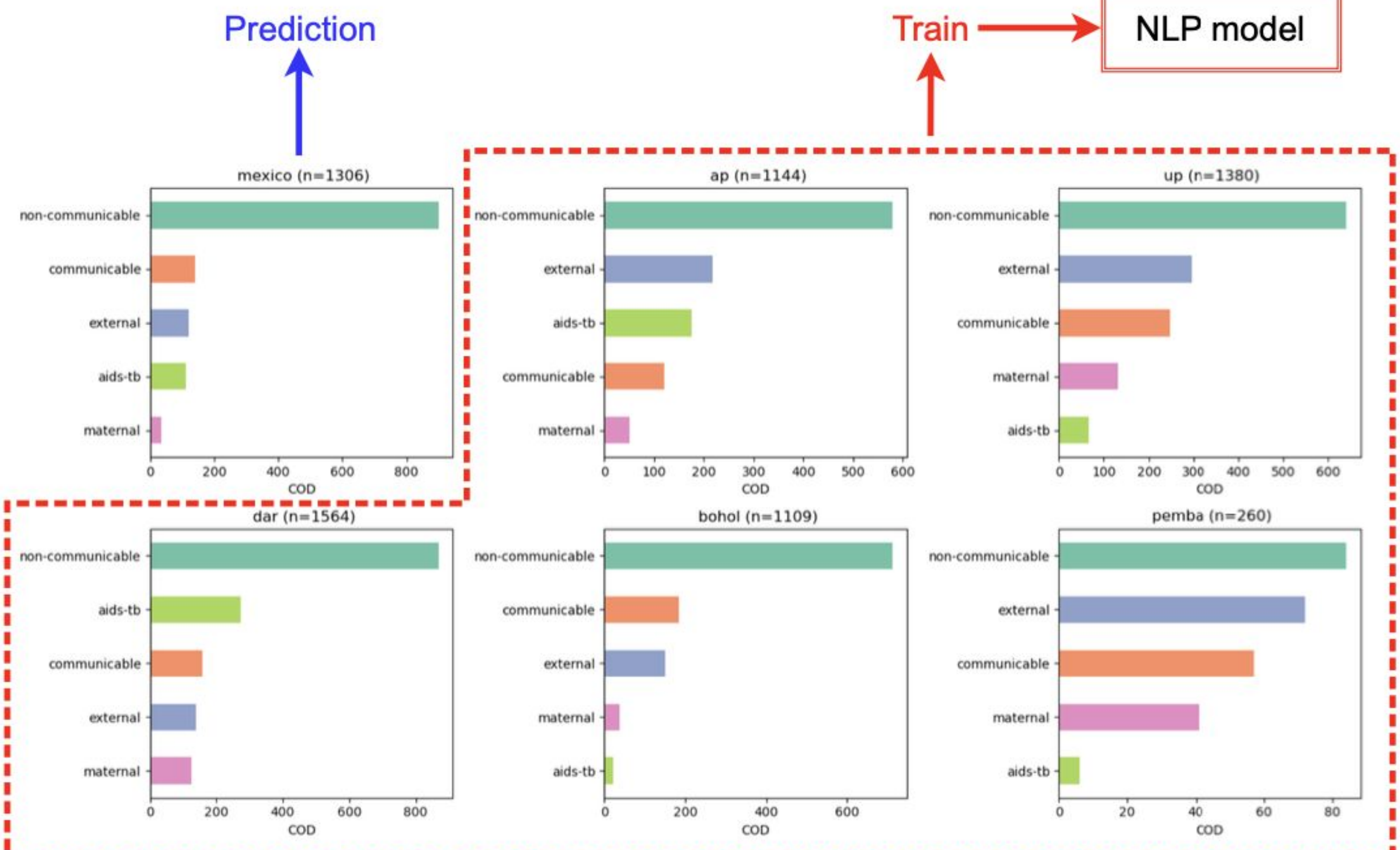
Validation set allows us to evaluate our experiment!

Inference with Predicted Data (multiPPI++)



Experimental Design - leave one out validation

Bag of words (Naive Bayes, KNN, SVM), BERT, GPT-4

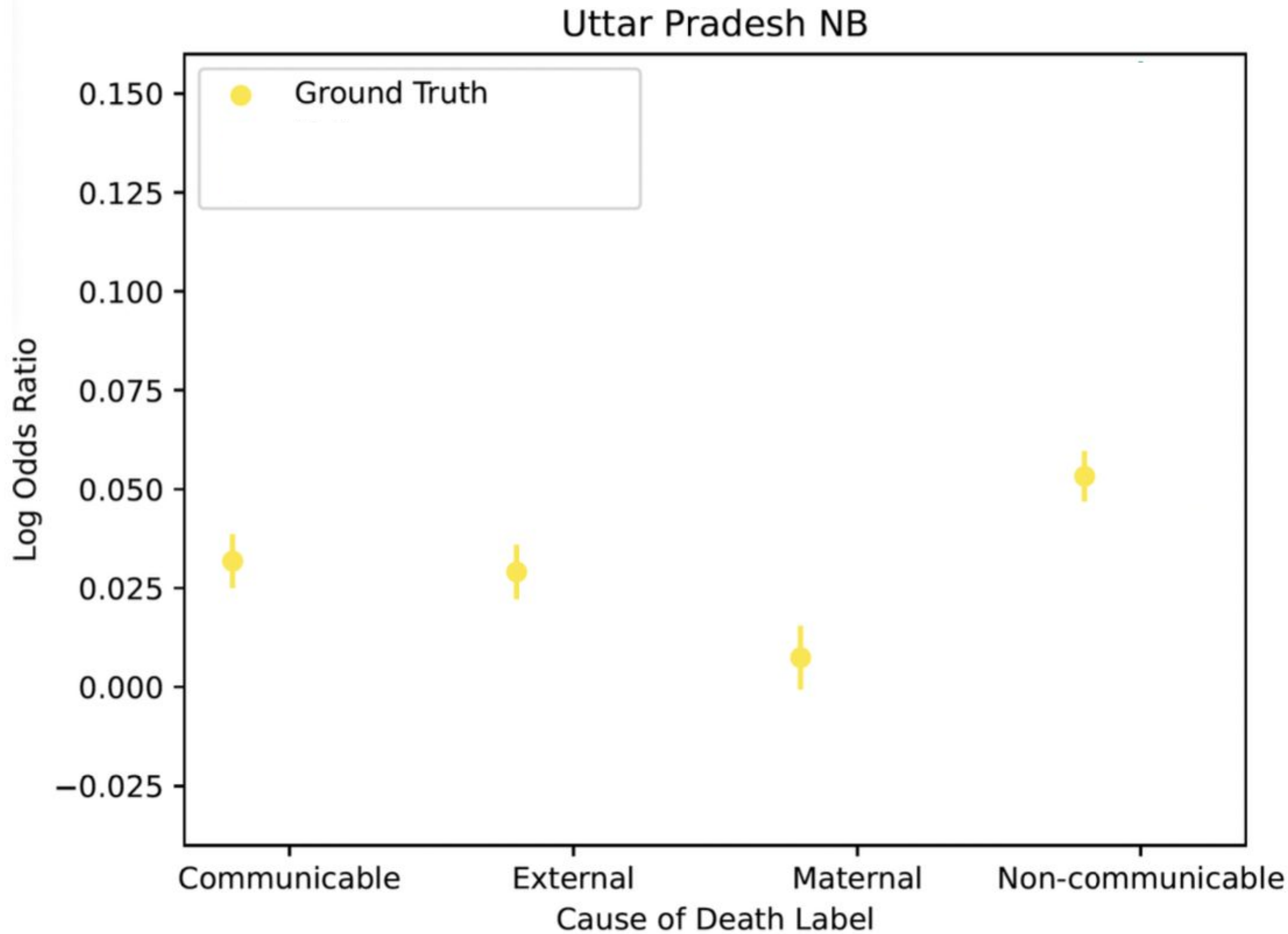


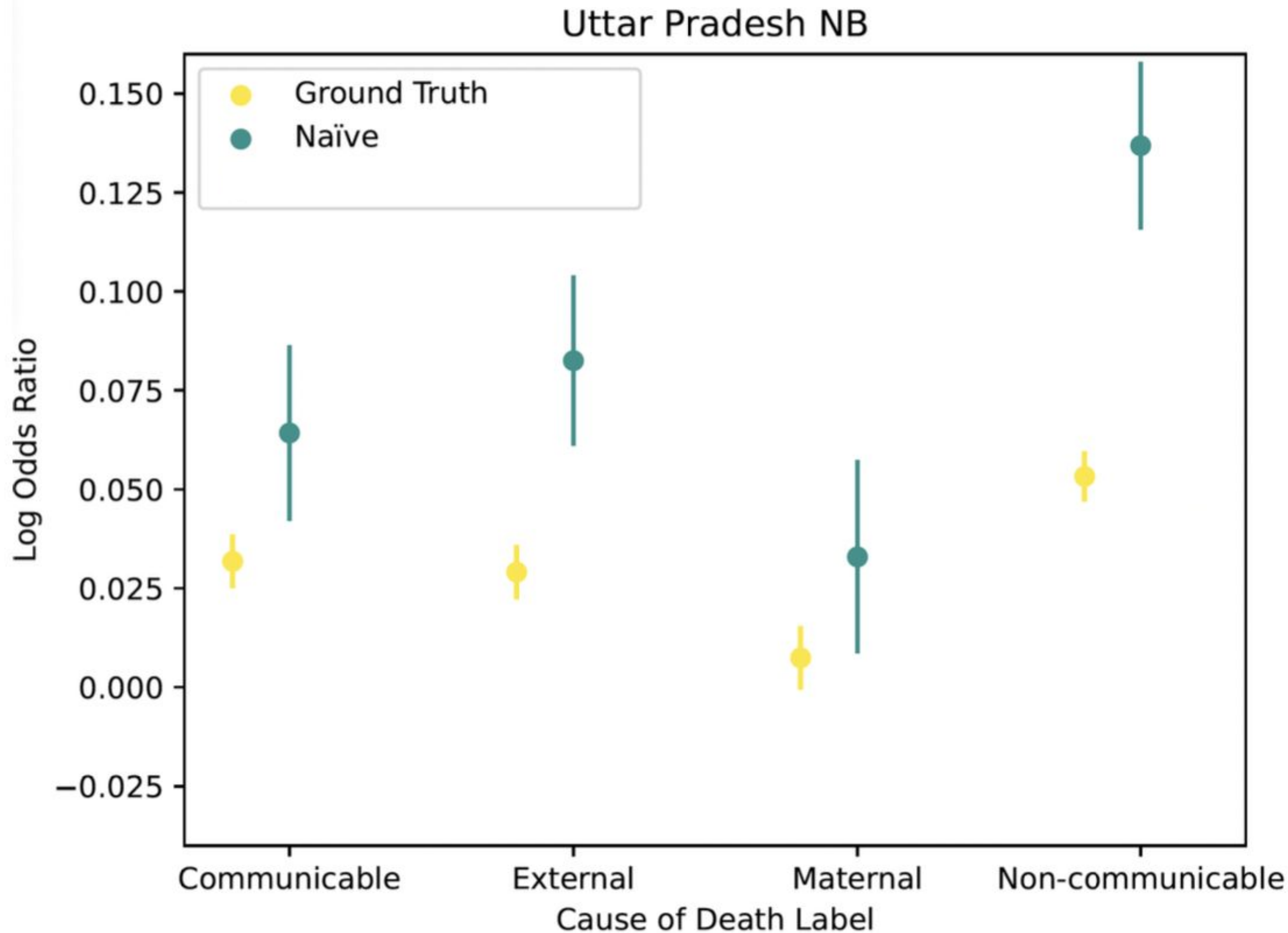
Multinomial Logistic Regression

Cause specific mortality associated with Age.

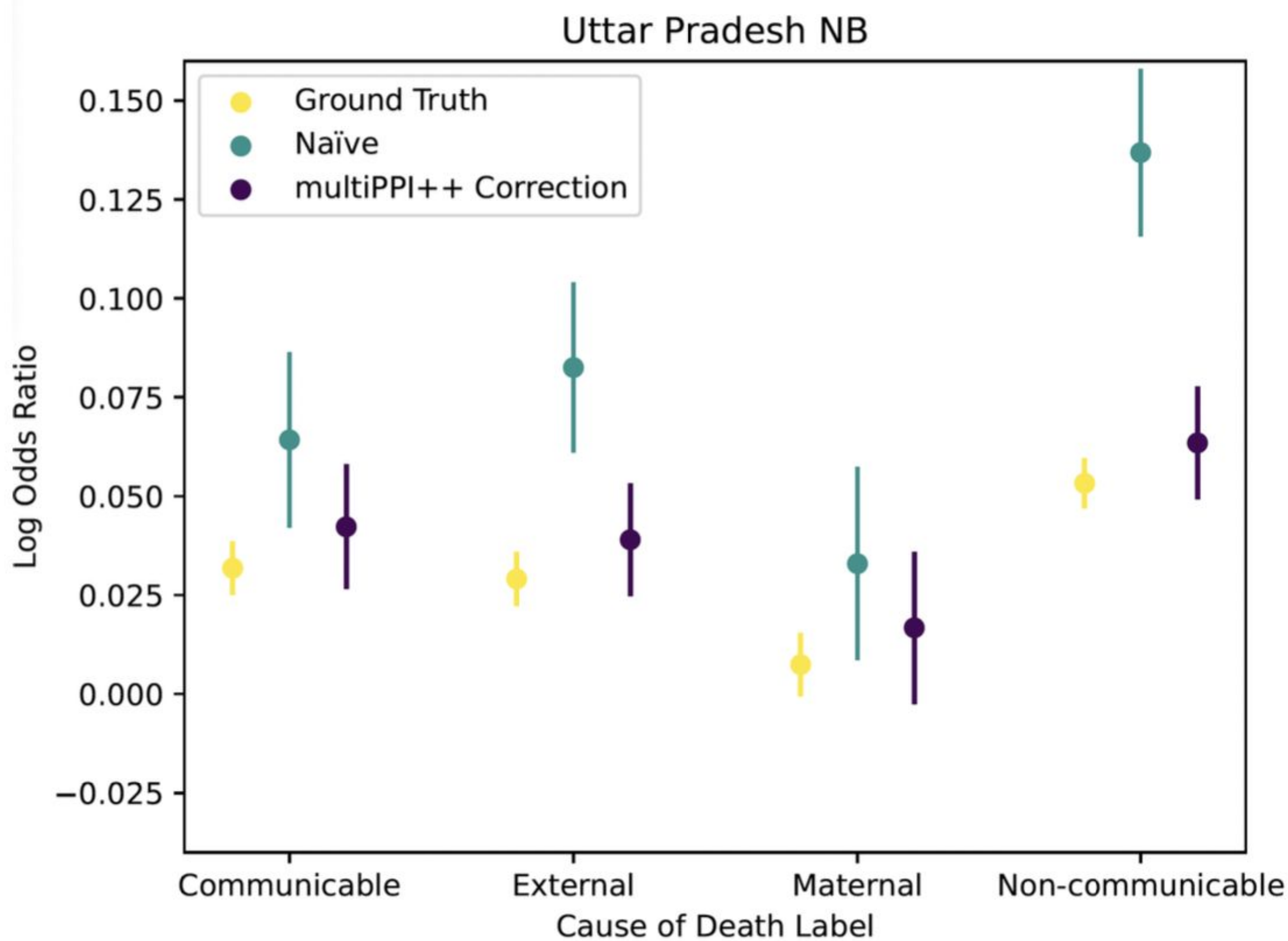
$$\log\left(\frac{p_{COD_i}}{p_{COD_{reference}}}\right) = \theta_0 + X_{age} * \theta_i$$

where θ_i is the change in log-odds of dying to cause *i* relative to the reference COD (aids-tb).

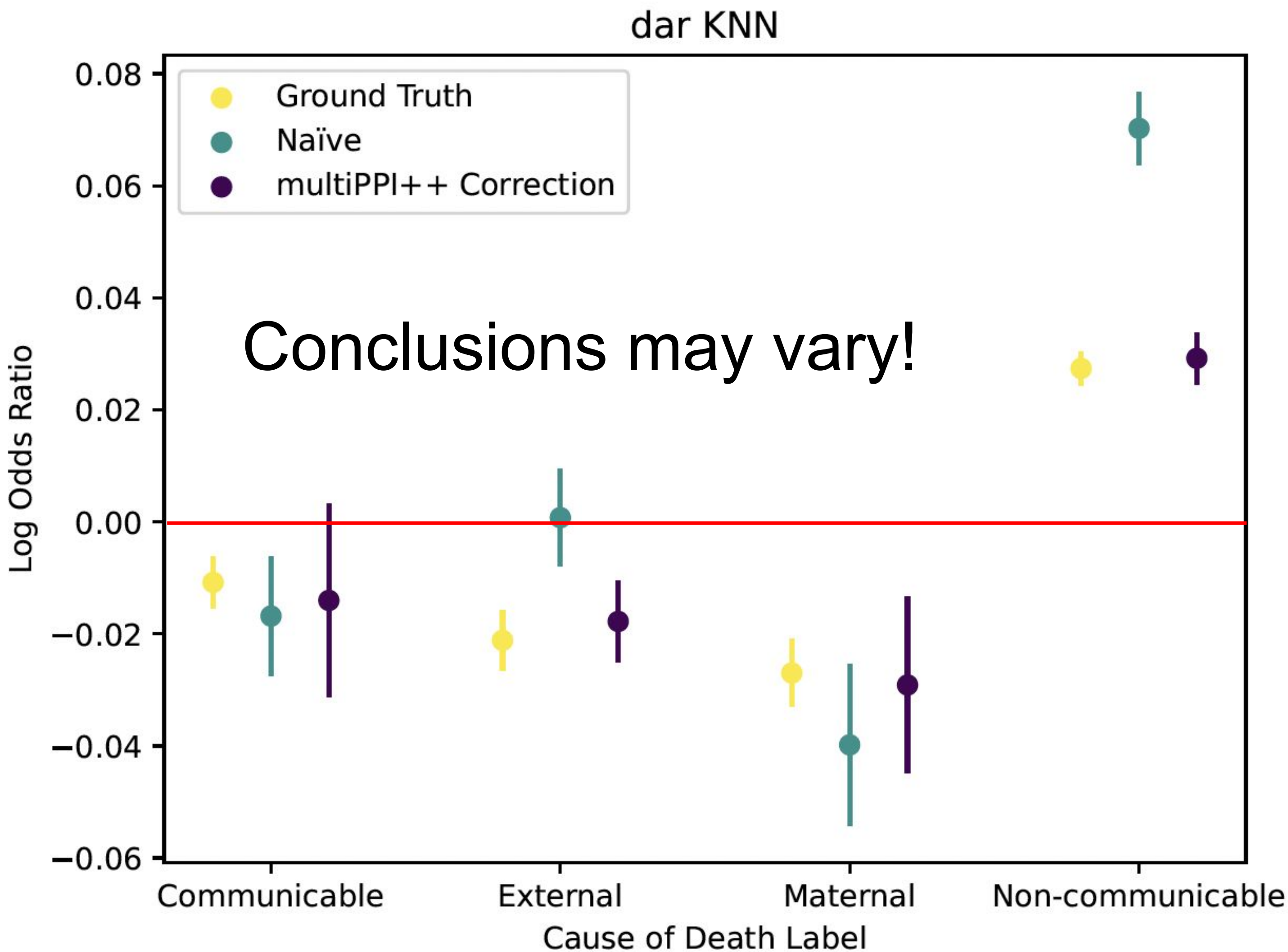

Uttar Pradesh NB
Ground Truth
0.150
0.125
0.100
0.075
0.050
0.025
0.000
−0.025
Log Odds Ratio
Communicable
External
Maternal
Non-communicable
Cause of Death Label



Uttar Pradesh NB
Ground Truth
Naïve
Log Odds Ratio
0.150
0.125
0.100
0.075
0.050
0.025
0.000
−0.025
Communicable
External
Maternal
Non-communicable
Cause of Death Label



Uttar Pradesh NB
Ground Truth
Naïve
multiPPI++ Correction
Log Odds Ratio
0.150
0.125
0.100
0.075
0.050
0.025
0.000
−0.025
Communicable
External
Maternal
Non-communicable
Cause of Death Label


dar KNN
Ground Truth
Naïve
multiPPI++ Correction
Conclusions may vary!
Log Odds Ratio
0.08
0.06
0.04
0.02
0.00
−0.02
−0.04
−0.06
Communicable
External
Maternal
Non-communicable
Cause of Death Label

The Body Mass Index (BMI)

1. “in **population** studies BMI is a **reasonable** surrogate measure of body and visceral fat, but it lacks sensitivity and specificity when applied to individuals.”

   - Nature, International Journal of Obesity (2009)

2. “BMI remains the most commonly used metric for **population-level** assessments and provides the most extensive data.”

   - the Lancet, Volume 405 March 08, 2025

Contributions:

1. We test the assumption that BMI is “good enough” for population-level inference, and find that **it is not**.

Contributions:

1. We question the assumption that BMI is “good enough” for population-level inference, and find that <u>**it is not**</u>.

2. We offer a practical solution (with caveats):

   a <u>**statistical calibration**</u> from inexpensive BMI-based measures of obesity towards better but less accessible measures.

Obesity

excessive fat accumulation that presents a risk to health

- World Health Organization

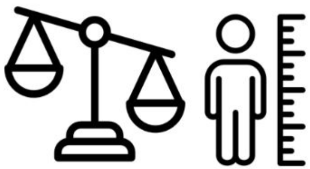
excessive ***fat accumulation*** *that presents a risk to health*

aka **Adiposity**

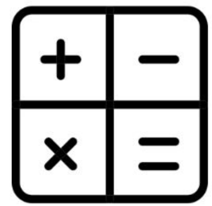
- World Health Organization

BMI is a cheap adiposity prediction algorithm

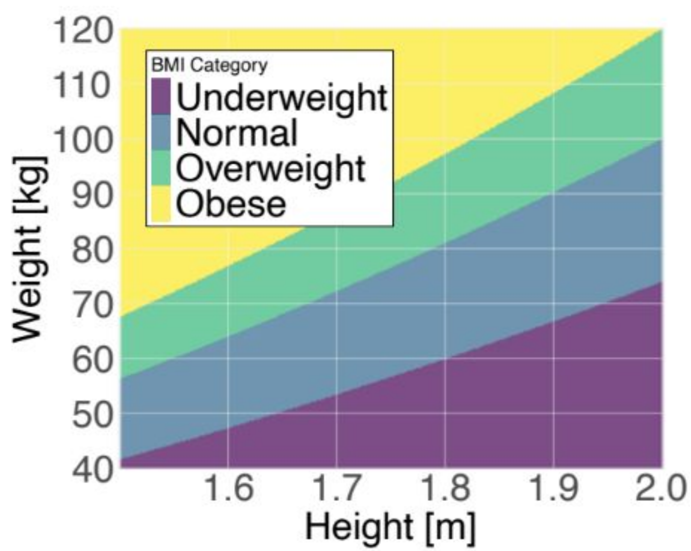
Measure Height (kg) and Weight (m)



Calculate BMI Score
kg / m^2



Create BMI Categories



The Body Mass Index: the Good, the Bad, and the Horrid

Barry Bogin and Ines Varela-Silva

OPEN ACCESS

GAVIN PUBLISHERS

Journal of Obesity and Nutritional Disorders

Research Article

Abrahim M and Hand B. J Obes Nutr Disord 06: 145.

DOI: 10.29011/2577-2244.100045

Is it Time to Consider Body Mass Index to be Bad Medical Information (BMI)?

Mohammed Abrahim*, Brittany Hand

HARVARD T.H. CHAN SCHOOL OF PUBLIC HEALTH

Home / News / BMI a poor metric for measuring people's health, say experts

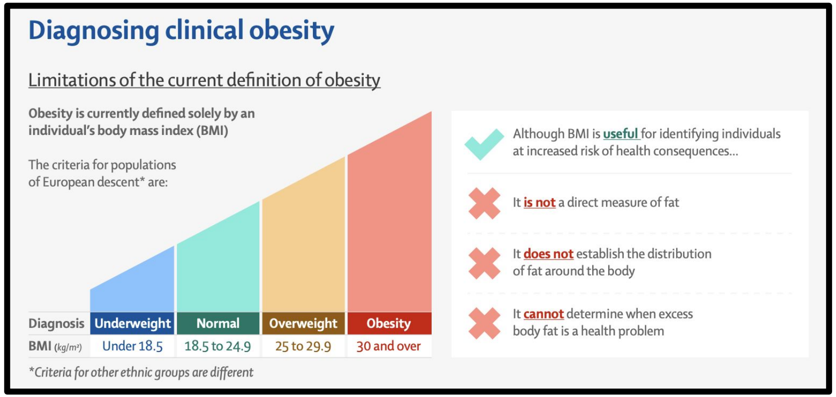
BMI a poor metric for measuring people's health, say experts

NUTRITION RESEARCH

Body Mass Index

Obesity, BMI, and Health A Critical Review

Nuttall, Frank Q. MD, PhD



"You Just Need to Lose Weight" and 19 Other Myths About Fat People

Aubrey Gordon

Co-host of *Maintenance Phase*

Why You Shouldn't Rely on BMI Alone

The Body Mass Index: the Good, the Bad, and the Horrid

HARVARD T.H. CHAN SCHOOL OF PUBLIC HEALTH

Home / News / BMI a poor metric for measuring people's health, say experts

BMI is discussed everywhere.

Ubiquity legitimates its use in research.

but BMI ≠ Adiposity!

Why You Shouldn't Rely on BMI Alone

AUBREY GORDON

CO-HOST OF *MAINTENANCE PHASE*

What is the “gold standard” measure of adiposity?

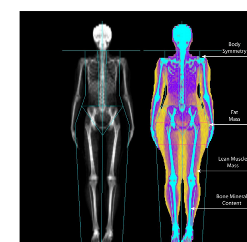
Dual-Energy X-Ray Absorptiometry

DXA scans are the “gold standard” measure of adiposity.

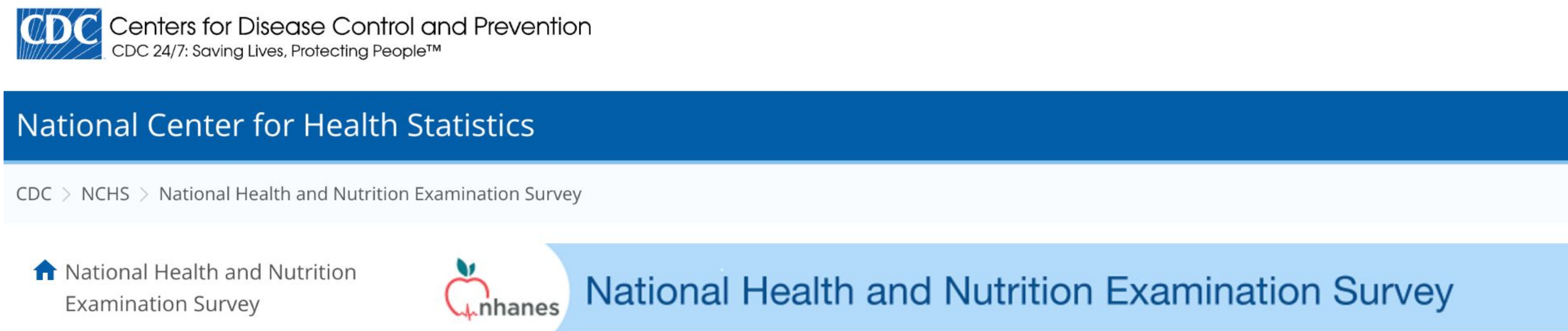
- Encyclopedia of Human Nutrition (Fourth Edition), 2013

As opposed to BMI and WC which measure **body proportions**, DXA measures **body composition** directly.

Whole-body percentage fat



Data



2011-2017	2021-2023
✅ BMI	✅ BMI
✅ Waist circumference (WC)	✅ Waist circumference (WC)
✅ Whole-body fat % (DXA)	❌ Whole-body fat % (DXA)

Data

Centers for Disease Control and Prevention
CDC 24/7: Saving Lives, Protecting People™

National Center for Health Statistics

CDC > NCHS > National Health and Nutrition Examination Survey

National Health and Nutrition Examination Survey

nhanes National Health and Nutrition Examination Survey

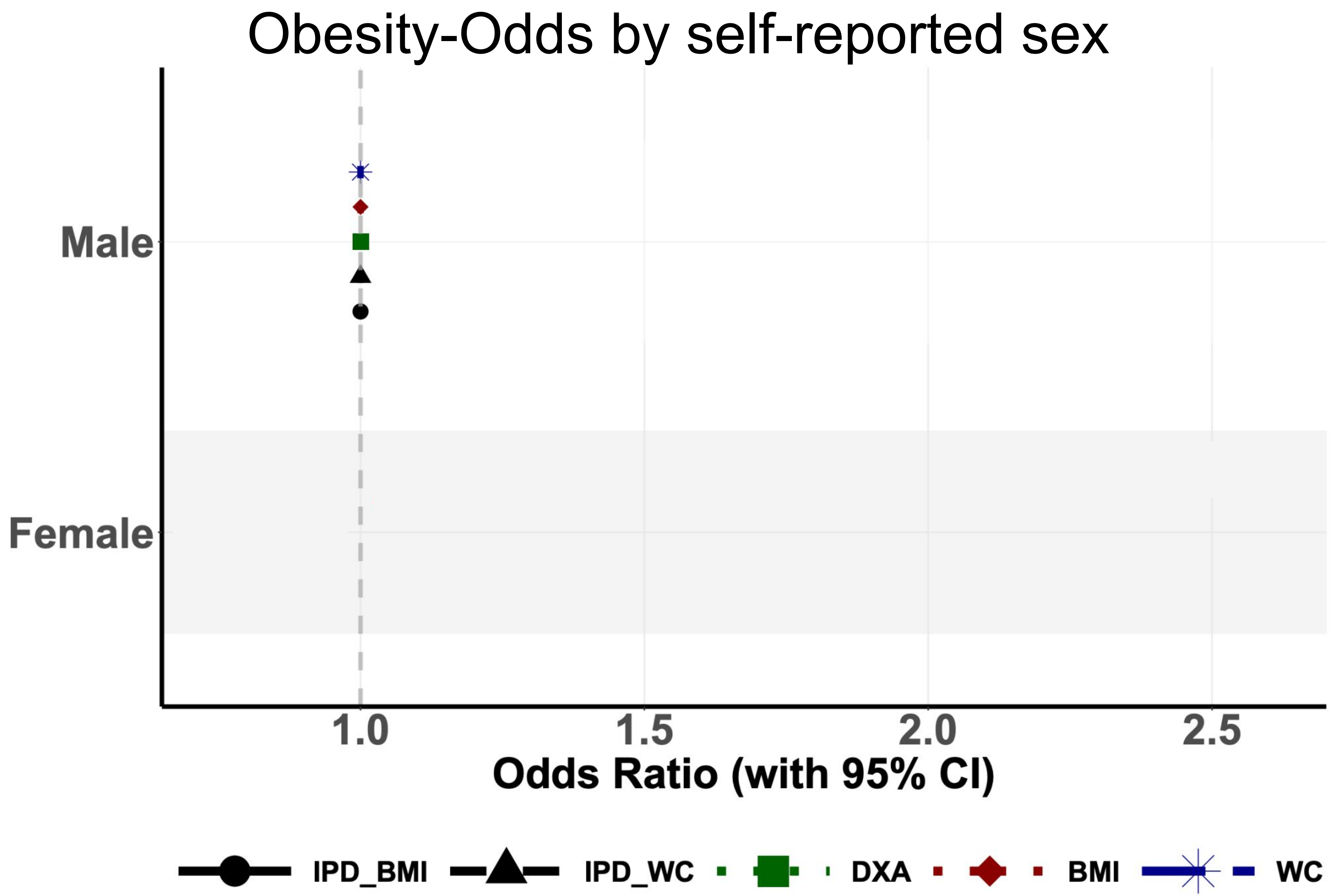
Obesity Threshold

BMI	> **30 kg/m^2** for females and males
Waist circumference (WC)	> **88cm** (female) or **102cm** (male)
Whole-body fat % (DXA)	> **42%** (female) or **30%** (male)

Results!

Compared to DXA-based obesity odds, what do WC and BMI estimates look like?

Odds Ratio (with 95% CI)

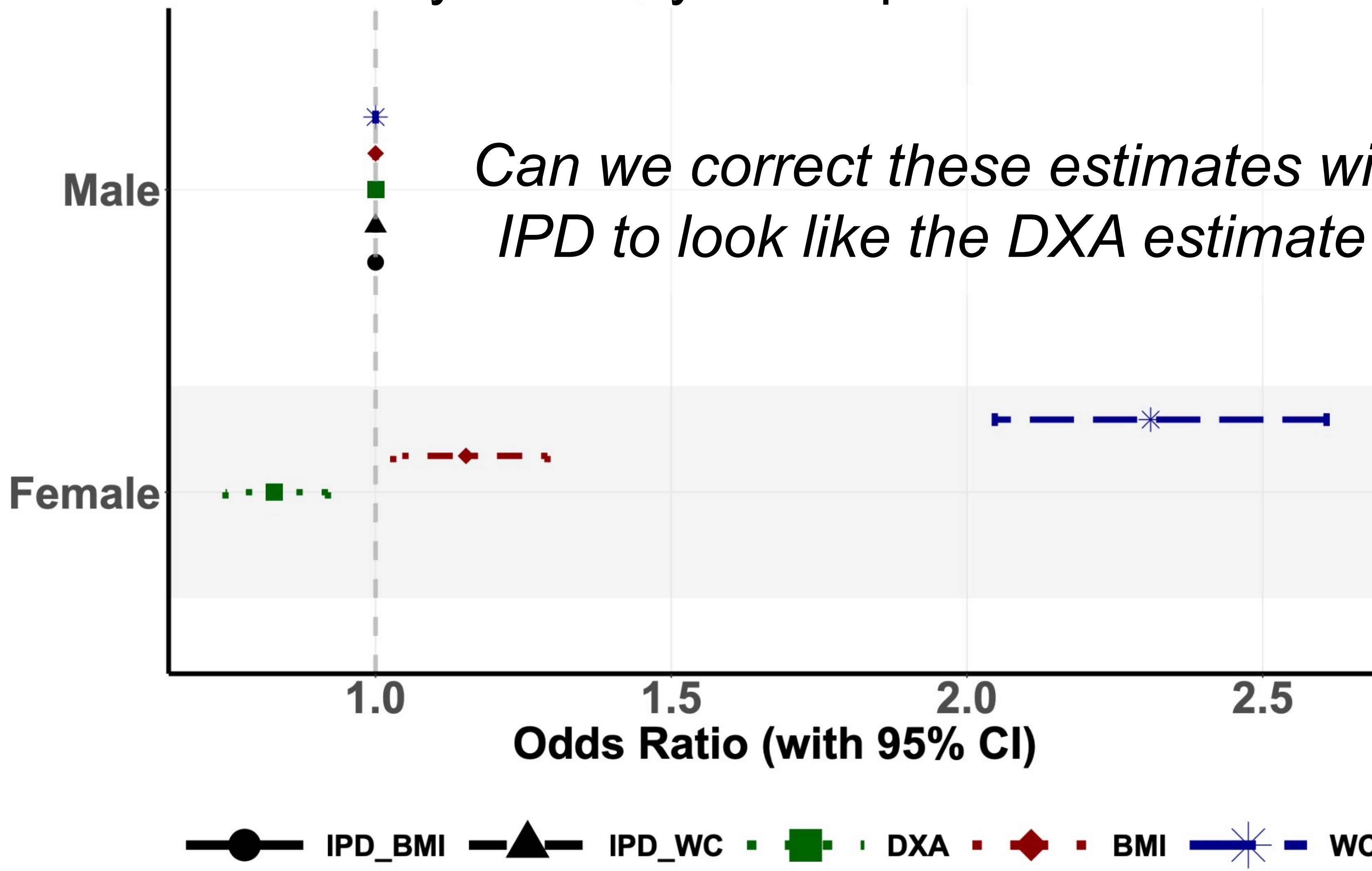

Obesity-Odds by self-reported sex
Male
Female
1.0
1.5
2.0
2.5
Odds Ratio (with 95% CI)
IPD_BMI
IPD_WC
DXA
BMI
WC


Obesity-Odds by self-reported sex
Male
Female
1.0
1.5
2.0
2.5
Odds Ratio (with 95% CI)
IPD_BMI
IPD_WC
DXA
BMI
WC


Obesity-Odds by self-reported sex
Male
Female
1.0
1.5
2.0
2.5
Odds Ratio (with 95% CI)
IPD_BMI
IPD_WC
DXA
BMI
WC

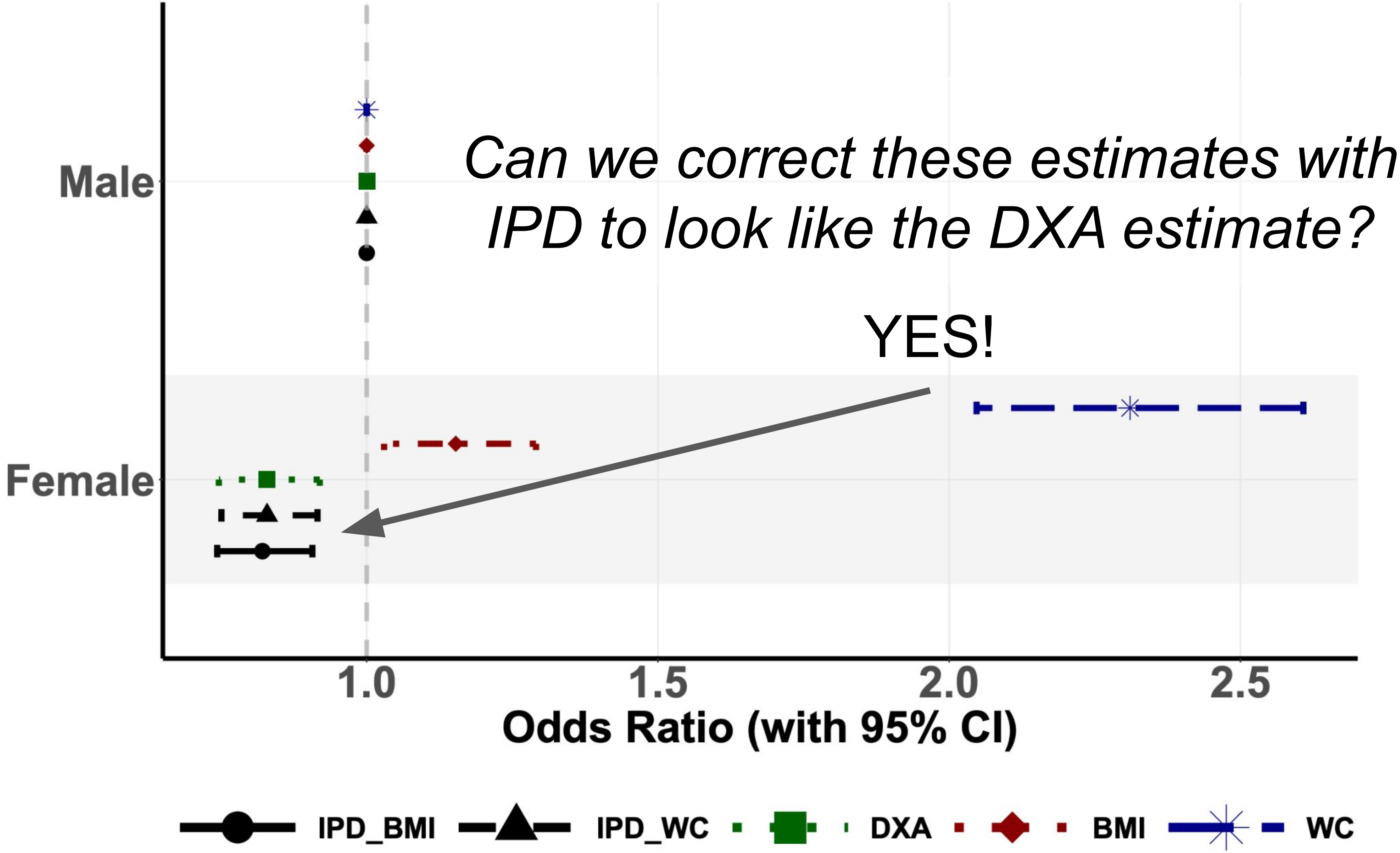

Obesity-Odds by self-reported sex
Male
Female
1.0
1.5
2.0
2.5
Odds Ratio (with 95% CI)
IPD_BMI
IPD_WC
DXA
BMI
WC

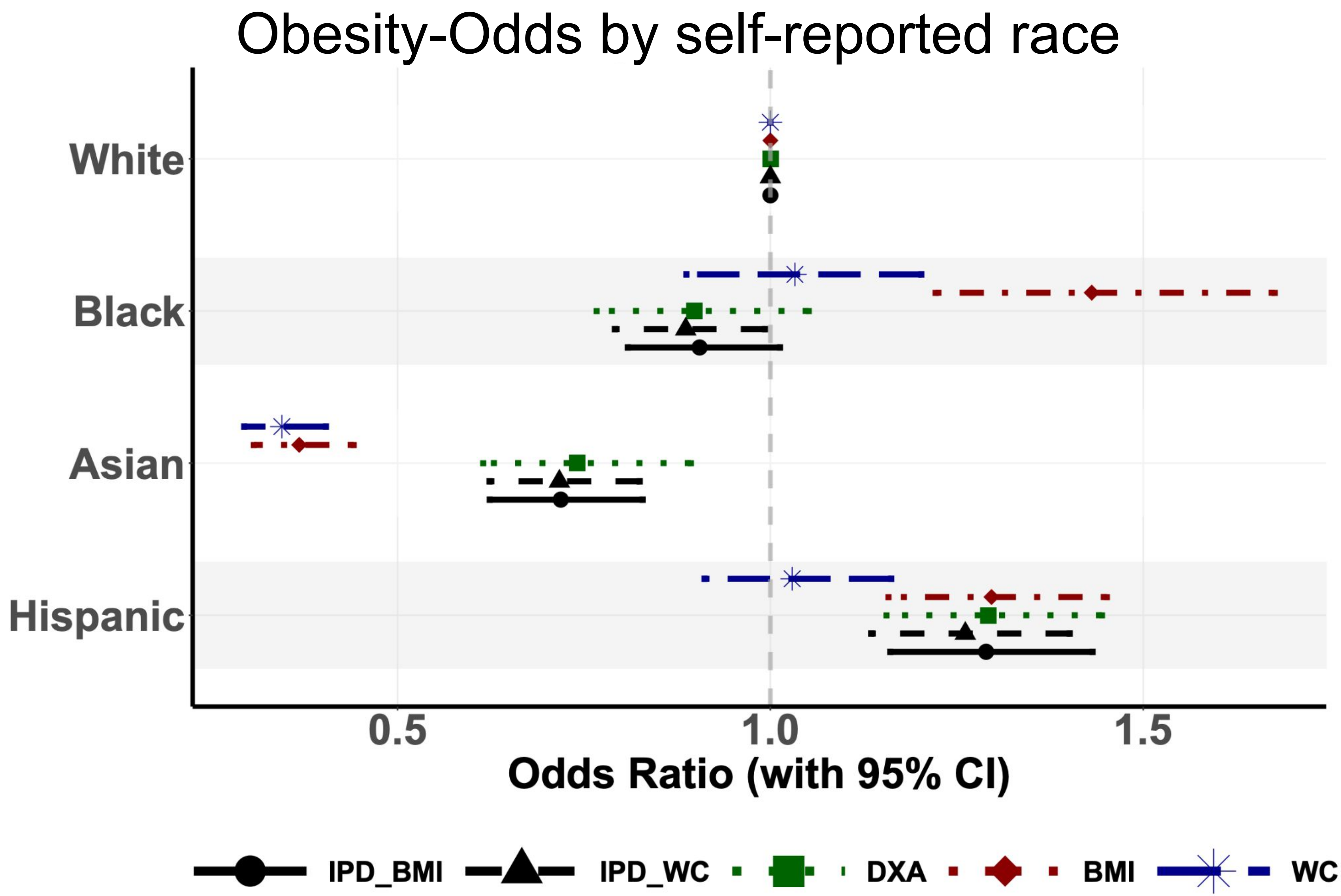
Obesity-Odds by self-reported sex



Can we correct these estimates with IPD to look like the DXA estimate?

Obesity-Odds by self-reported sex




Obesity-Odds by self-reported race
White
Black
Asian
Hispanic
0.5
1.0
1.5
Odds Ratio (with 95% CI)
IPD_BMI
IPD_WC
DXA
BMI
WC

Takeaways

Are you using predictions in downstream inference?

Consider an IPD calibration!

Here’s an explainer with a numerical example!

Thank you!!

Contact:
Adam Visokay
avisokay@uw.edu
https://avisokay.github.io/

IPD software is available!

Paper

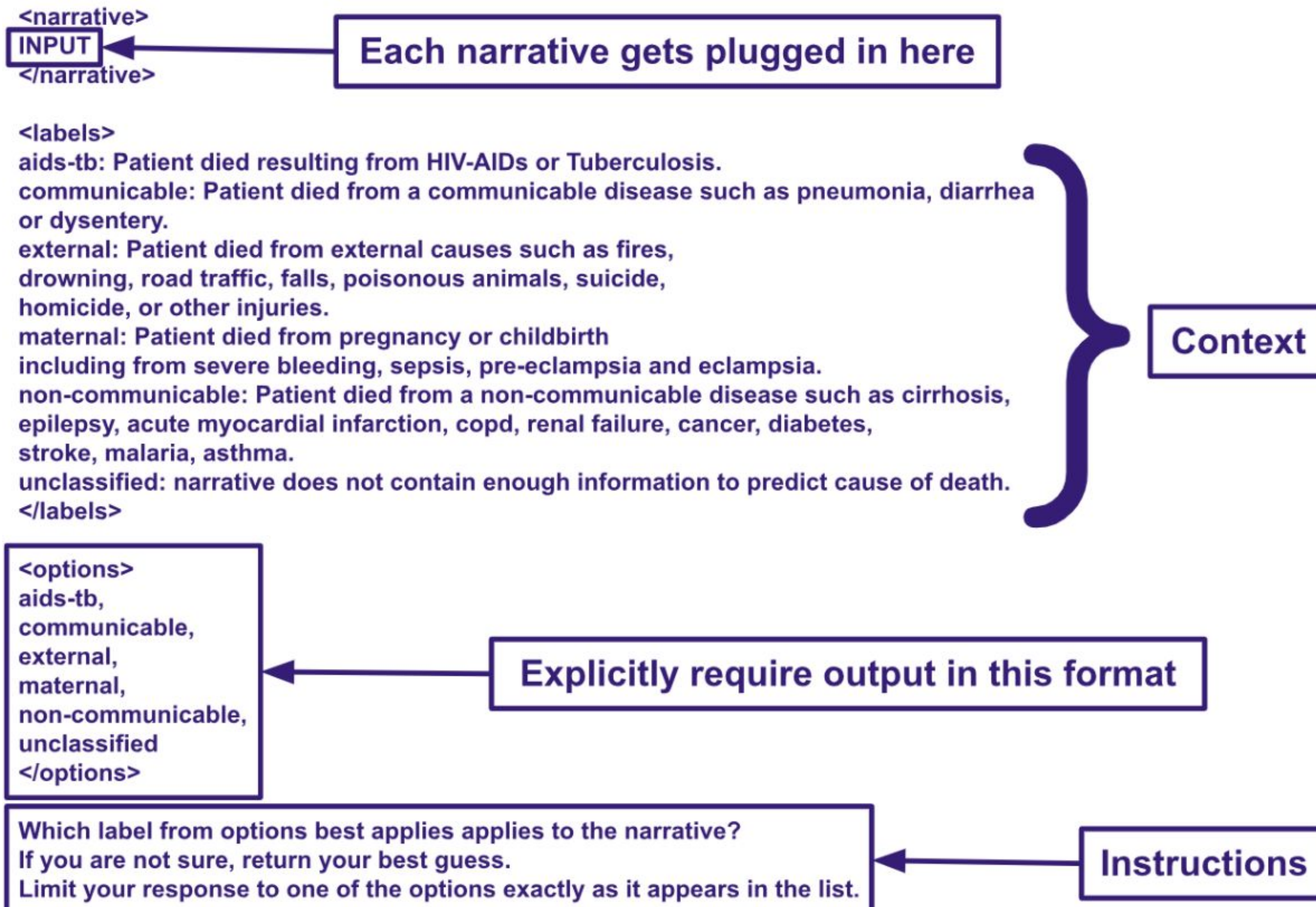
Github

CRAN

COLM

medRxiv
THE PREPRINT SERVER FOR HEALTH SCIENCES

Appendix


<narrative>
INPUT
</narrative>
Each narrative gets plugged in here
<labels>
aids-tb: Patient died resulting from HIV-AIDs or Tuberculosis.
communicable: Patient died from a communicable disease such as pneumonia, diarrhea or dysentery.
external: Patient died from external causes such as fires, drowning, road traffic, falls, poisonous animals, suicide, homicide, or other injuries.
maternal: Patient died from pregnancy or childbirth including from severe bleeding, sepsis, pre-eclampsia and eclampsia.
non-communicable: Patient died from a non-communicable disease such as cirrhosis, epilepsy, acute myocardial infarction, copd, renal failure, cancer, diabetes, stroke, malaria, asthma.
unclassified: narrative does not contain enough information to predict cause of death.
</labels>
Context
<options>
aids-tb,
communicable,
external,
maternal,
non-communicable,
unclassified
</options>
Explicitly require output in this format
Which label from options best applies applies to the narrative?
If you are not sure, return your best guess.
Limit your response to one of the options exactly as it appears in the list.
Instructions

Regularized Loss Function

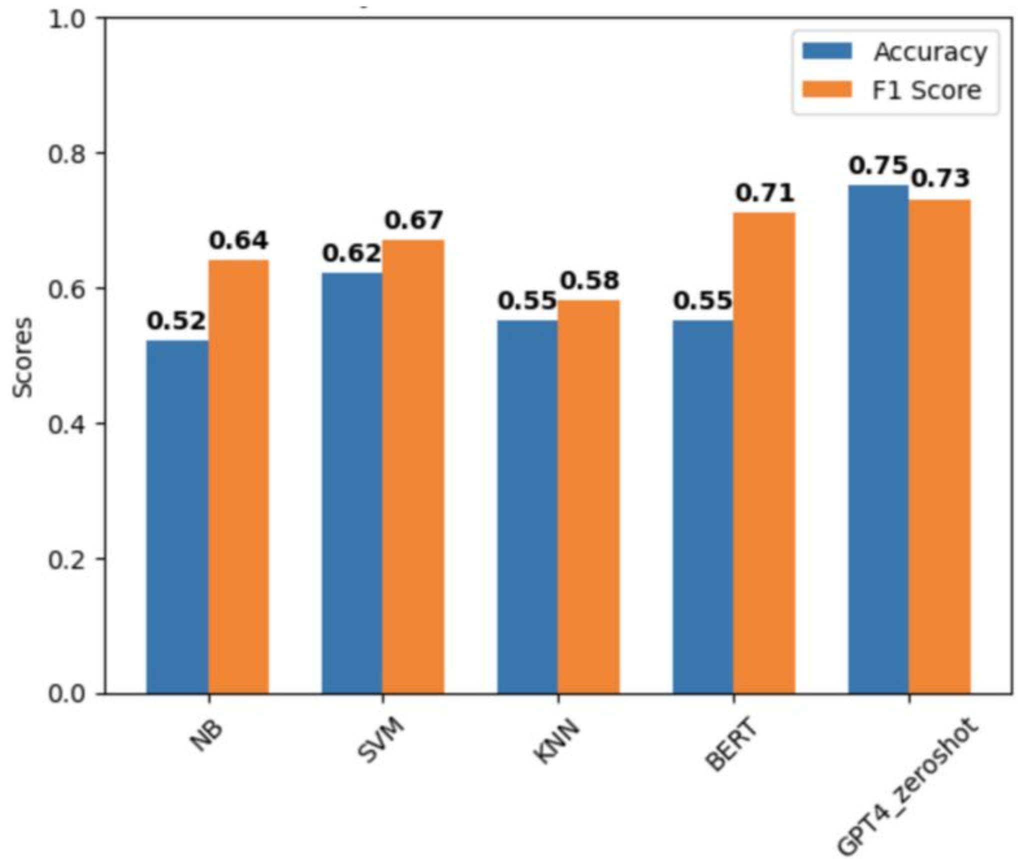
$$\mathbb{E}[\ell_\theta(X_L, Y_L)] \quad +$$

$$\lambda\left(\mathbb{E}[\ell_\theta(X_U, \hat{Y}_U^{AI})] - \mathbb{E}[l_\theta(X_L, \hat{Y}_L^{AI})]\right)$$

Lambda is a tuning parameter in [0,1]

Lambda = 0 when the predicted data are all **noise**

Lambda = 1 when the predicted data are all **signal**


1.0
0.8
0.6
0.4
0.2
0.0
Scores
Accuracy
F1 Score
0.52
0.64
0.62
0.67
0.55
0.58
0.55
0.71
0.75
0.73
NB
SVM
KNN
BERT
GPT4_zeroshot

Closer Look at GPT-4 Predictions

narrative	gs_cod	prediction
respondent thanked for being visited	aids-tb	The narrative does not provide enough information to determine a cause of death.
client had no additional point	non-communicable	The narrative does not provide enough information to determine the appropriate label.
the client thanked for service which provided in the hospital_x000d__x000d_\nthe client transfer death certificate to their original home [place]	non-communicable	The narrative does not provide enough information to determine the cause of death.
the client thanked for the service	communicable	The narrative does not provide information related to any of the labels.
no comment	communicable	The narrative does not provide enough information to determine the cause of death.

- GPT-4 fails to classify 1503 of the 6763 cases. These 1503 text narratives contain no useful information.

How does Age (X) vary with Cause of Death (y)?

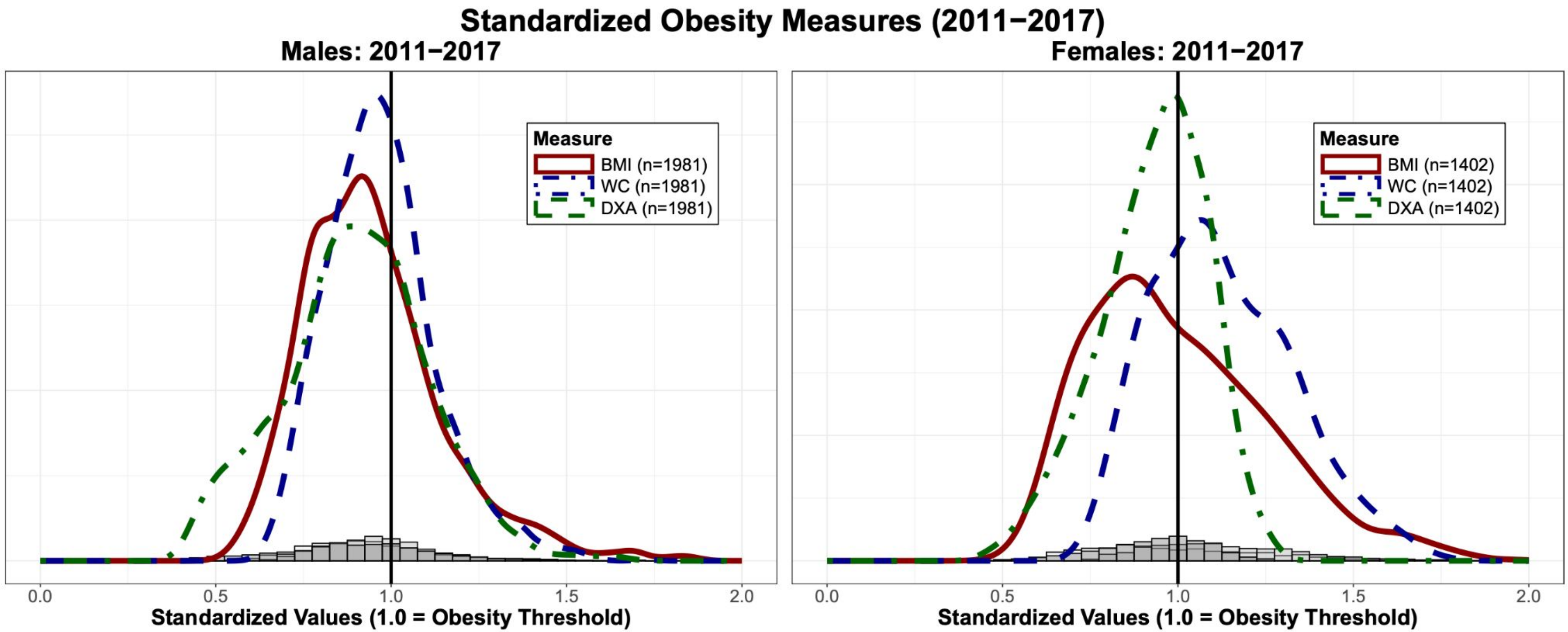
multinomial logistic regression:

$$\log\left(\frac{p_{COD_i}}{p_{COD_{reference}}}\right) = \theta_0 + X_{age} * \theta_i$$

for $\theta \in \{1, ..., 4\}$

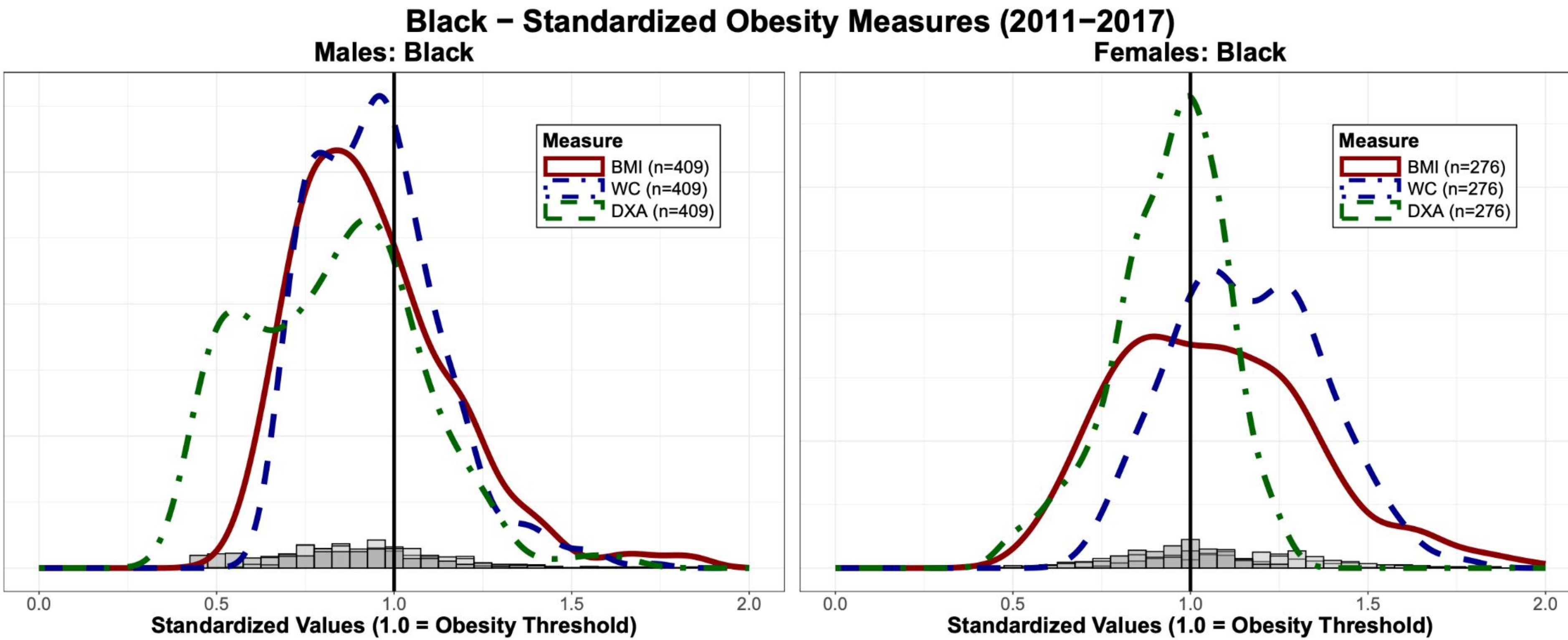
- θ_1, θ_2, θ_3, θ_4 are the multinomial regression coefficients when we regress $COD \sim Age$.
- With AIDS-TB as the left out reference category we have:
  - θ_1: For every one-unit increase in Age(yr), the log-odds of P(Y=**communicable**) (compared to AIDS-TB) are expected to increase by θ_1.
  - θ_2: P(Y=**external**) are expected to increase by θ_2.
  - θ_3: P(Y=**maternal**) are expected to increase by θ_3.
  - θ_4: P(Y=**non-communicable**) are expected to increase by θ_4.

Appendix



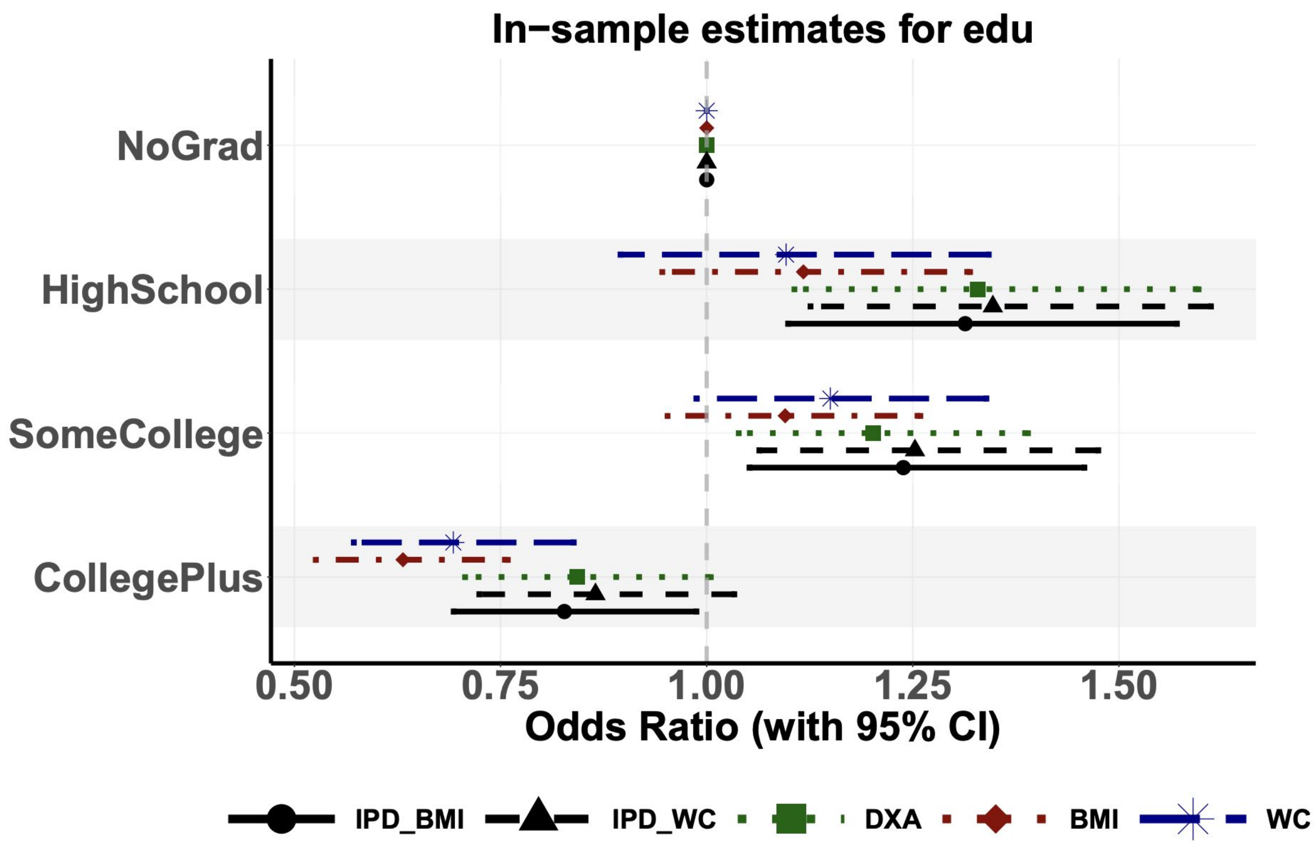
Standardized Obesity Measures (2011−2017)
Males: 2011−2017
Females: 2011−2017
Measure
BMI (n=1981)
WC (n=1981)
DXA (n=1981)
BMI (n=1402)
WC (n=1402)
DXA (n=1402)
0.0
0.5
1.0
1.5
2.0
Standardized Values (1.0 = Obesity Threshold)

Appendix



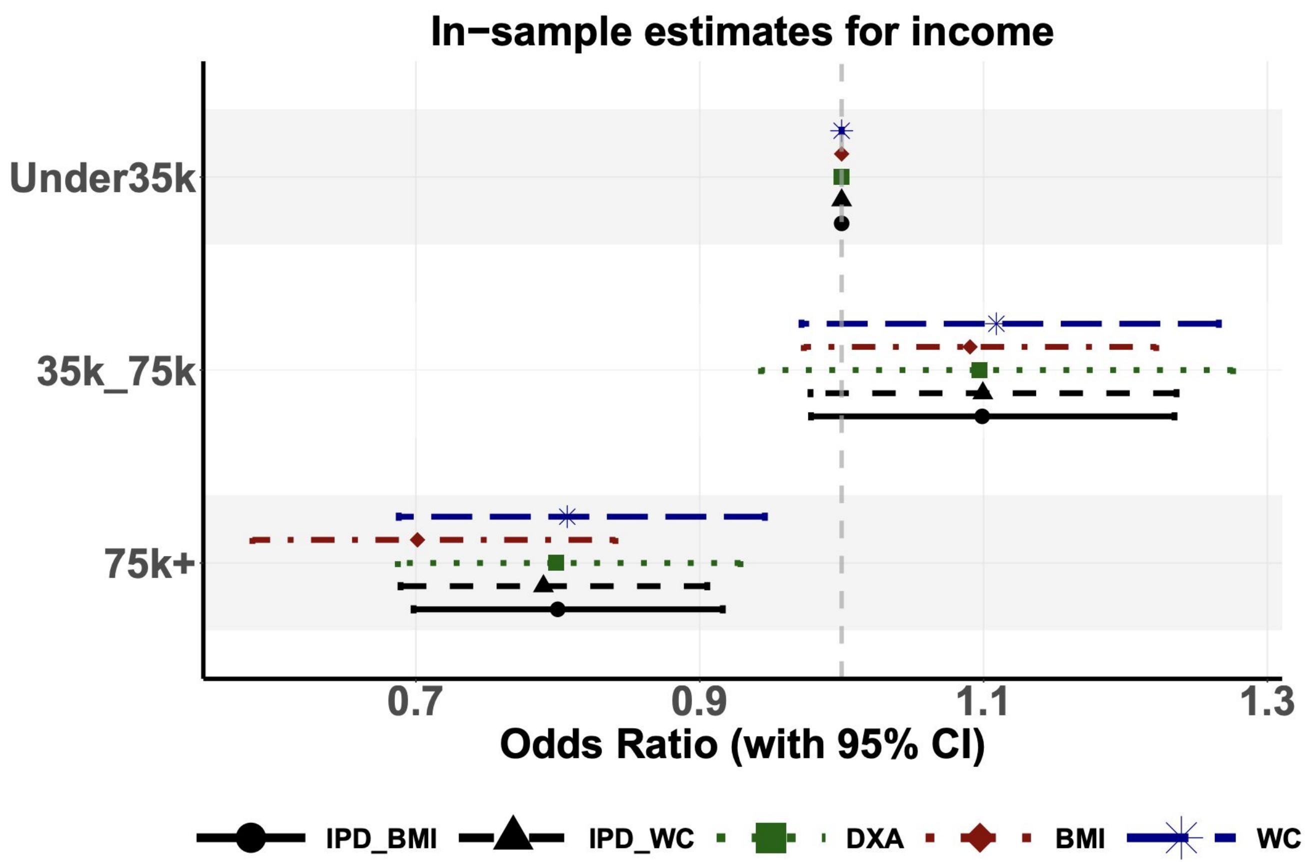
Black − Standardized Obesity Measures (2011−2017)
Males: Black
Females: Black
Measure
BMI (n=409)
WC (n=409)
DXA (n=409)
BMI (n=276)
WC (n=276)
DXA (n=276)
0.0
0.5
1.0
1.5
2.0
Standardized Values (1.0 = Obesity Threshold)

Appendix



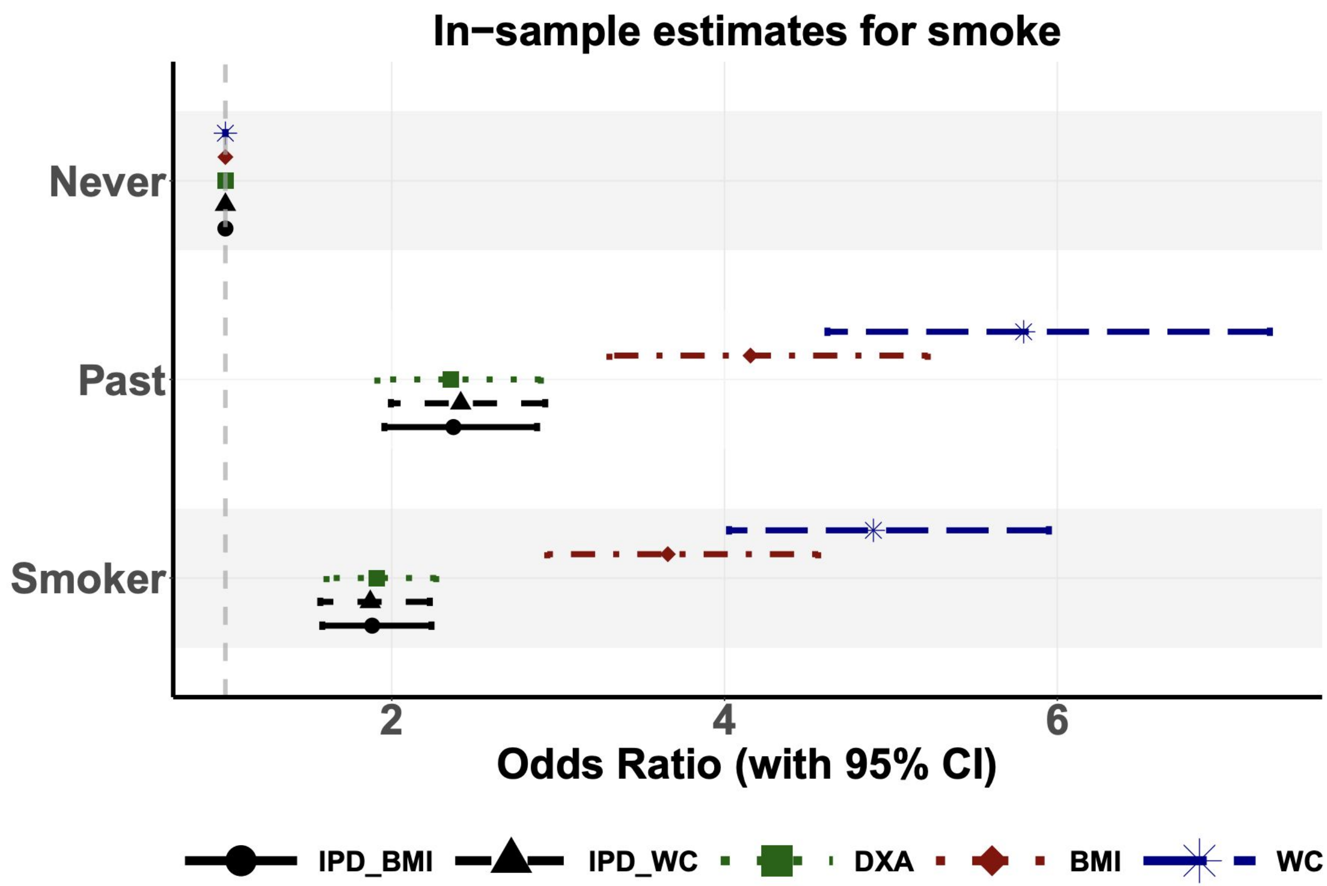
In−sample estimates for edu
NoGrad
HighSchool
SomeCollege
CollegePlus
0.50
0.75
1.00
1.25
1.50
Odds Ratio (with 95% CI)
IPD_BMI
IPD_WC
DXA
BMI
WC

Appendix



In−sample estimates for income
Under35k
35k_75k
75k+
0.7
0.9
1.1
1.3
Odds Ratio (with 95% CI)
IPD_BMI
IPD_WC
DXA
BMI
WC

Appendix



In−sample estimates for smoke
Never
Past
Smoker
2
4
6
Odds Ratio (with 95% CI)
IPD_BMI
IPD_WC
DXA
BMI
WC