Inference on Predicted Data: Examples from Verbal Autopsies and BMI

Adam Visokay

eScience Institute

November 12, 2024

Our Team!

Adam Visokay

Trinity Fan

Kentaro Hoffman

Stephen Salerno

Sasha Johfre

Jeff Leek

Li Liu

Tyler McCormick

Overview

Adam Visokay
Department of Sociology
University of Washington PhD Student
Max Planck Institute for Population Research Affiliated Student

Roadmap for today

1. Estimation versus Prediction
2. Inference on Predicted Data (IPD)
3. Verbal Autopsy Narratives
4. BMI as Prediction Algorithm
5. Q&A

Estimation versus Prediction

Inference on Predicted Data

Estimation: $y = \beta X_{\text{train}} \rightarrow \hat{\beta}$

Prediction: $\hat{\beta} X_{\text{test}} = \hat{y}$

Inference on Predicted Data (IPD)

Estimation: $y = \beta X_{\text{train}} \rightarrow \hat{\beta}$

Prediction: $\hat{\beta} X_{\text{test}} = \hat{y}$

IPD: $\hat{y}_{\text{AI}} = \beta X_{\text{train}} \rightarrow \hat{\beta}_{\text{AI}}$

Inference on Predicted Data (IPD)

What's the association between education (X) and income (y)? $\longrightarrow \hat{\beta}$

Predict income (y) given education (X) $= \hat{y}$

Inference on Predicted Data (IPD)

What's the association between education (X) and income (y)? $\longrightarrow \hat{\beta}$

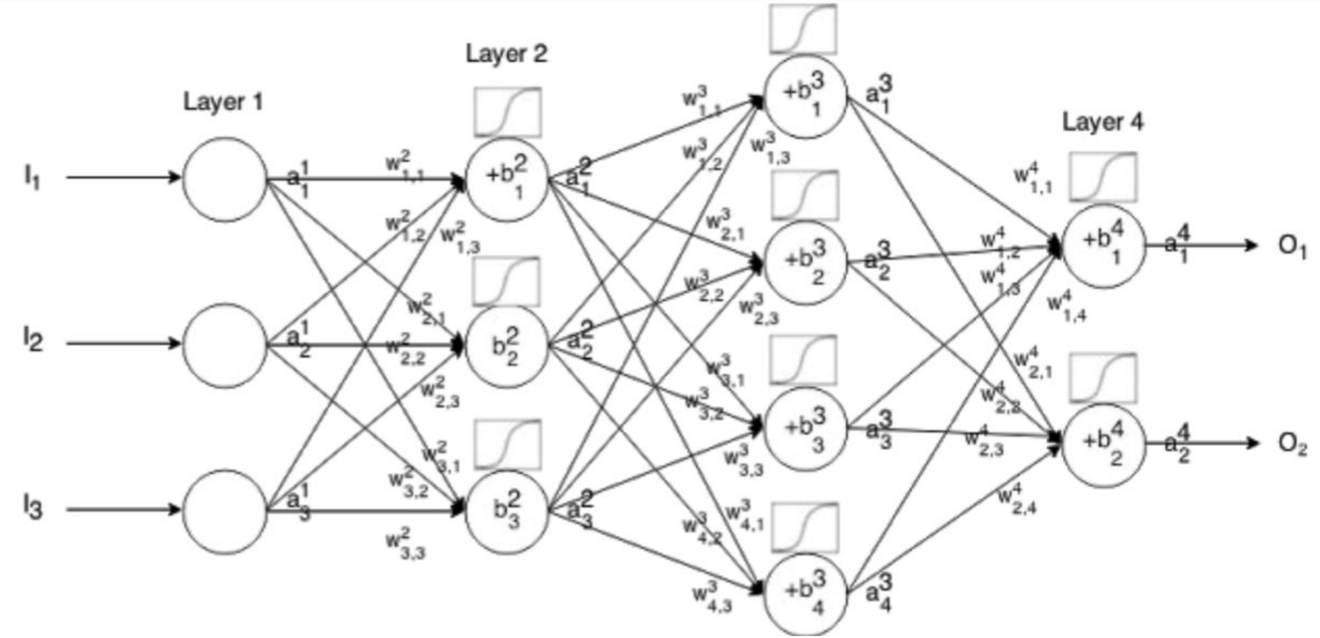
Predict income (y) given education (X) $= \hat{y}$

Estimate association between education (X) and <u>AI predicted</u> income (y) $\longrightarrow \hat{\beta}_{\mathrm{AI}}$

Predicted Data Is Often Practical

$ $$$$

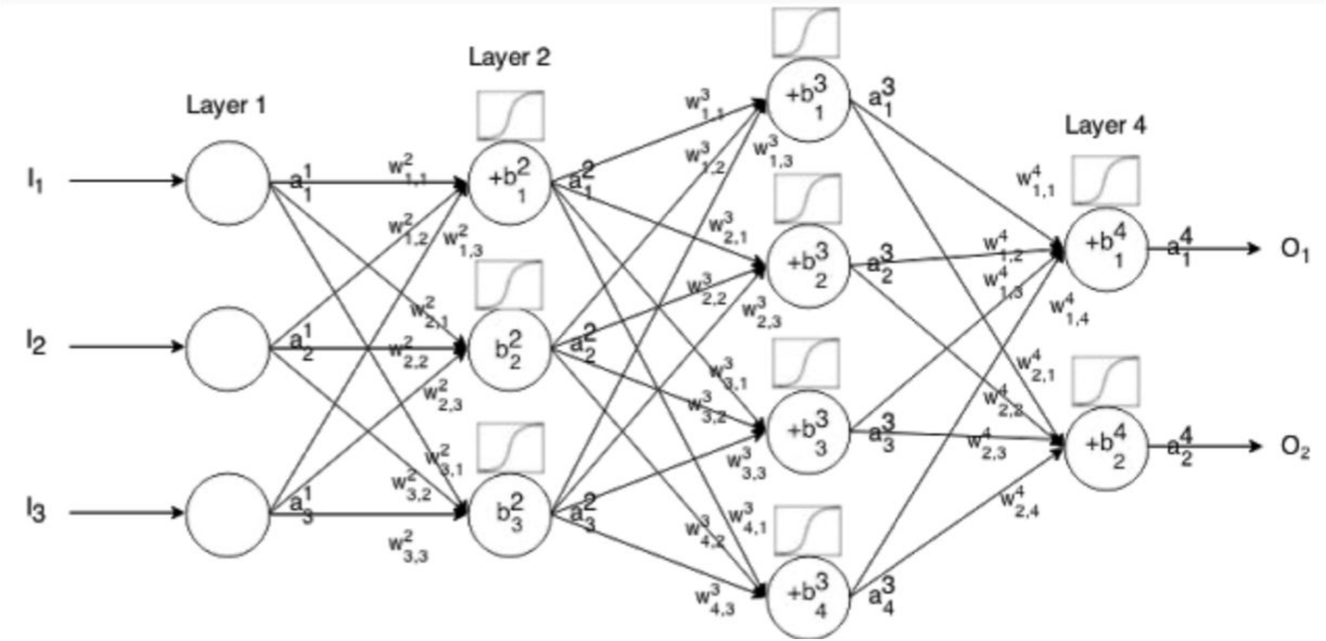
f(DataYouCanGet) = (DataYouWant)



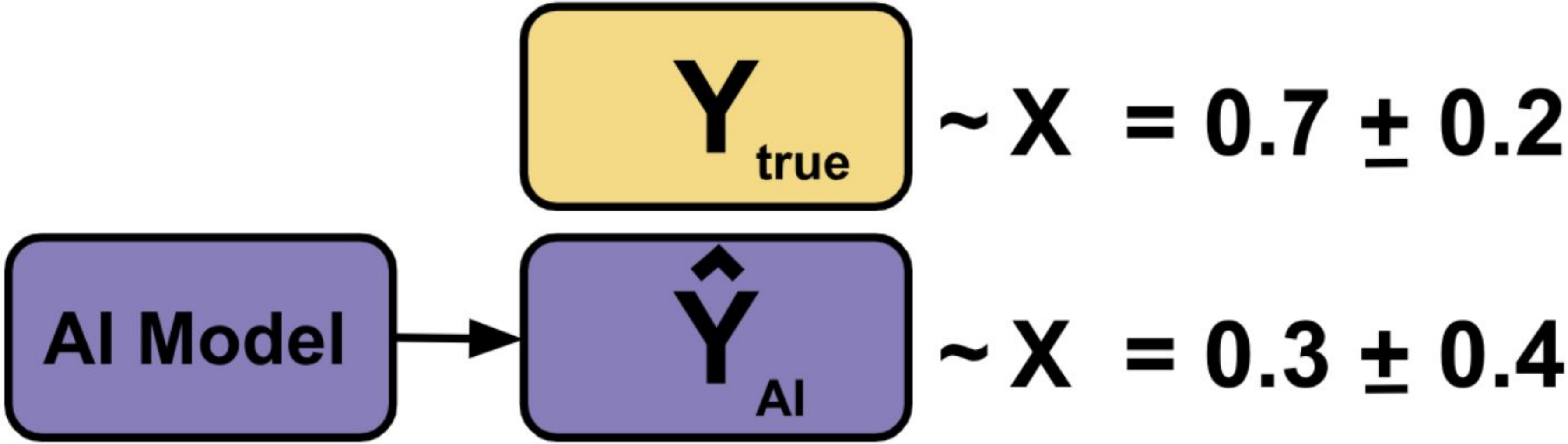
Predicted Data Is Often Practical

\$ *\$\$\$\$*

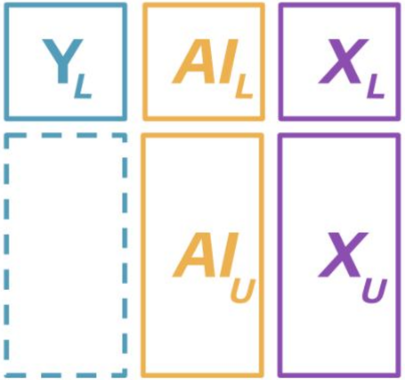
f(DataYouCanGet) = (DataYouWant)-ish



But IPD leads to Invalid Uncertainty *and* Potential Bias



IPD Correction Procedure

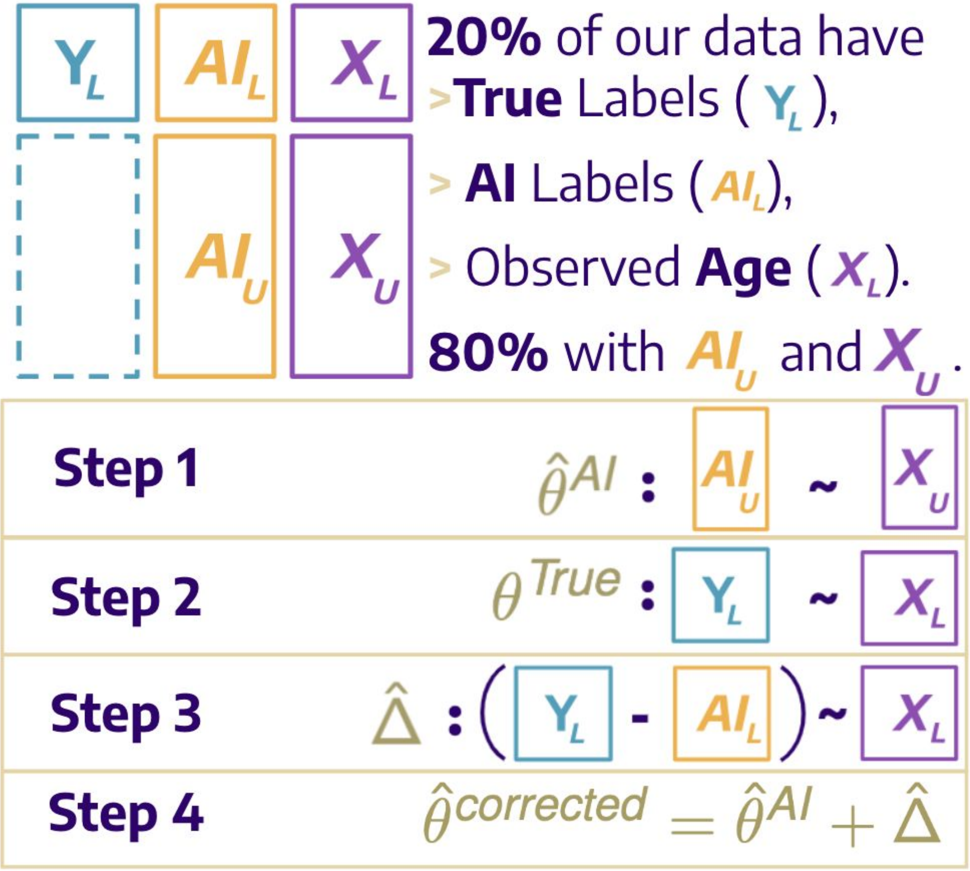


20% of our data have

- **True** Labels (Y_L),
- **AI** Labels (AI_L),
- Observed **Age** (X_L).

80% with AI_U and X_U .

IPD Correction Procedure



Verbal Autopsy (VA)

Verbal Autopsy (VA)

Fewer than one-third of deaths worldwide assigned medically certified cause (Horton, 2007)

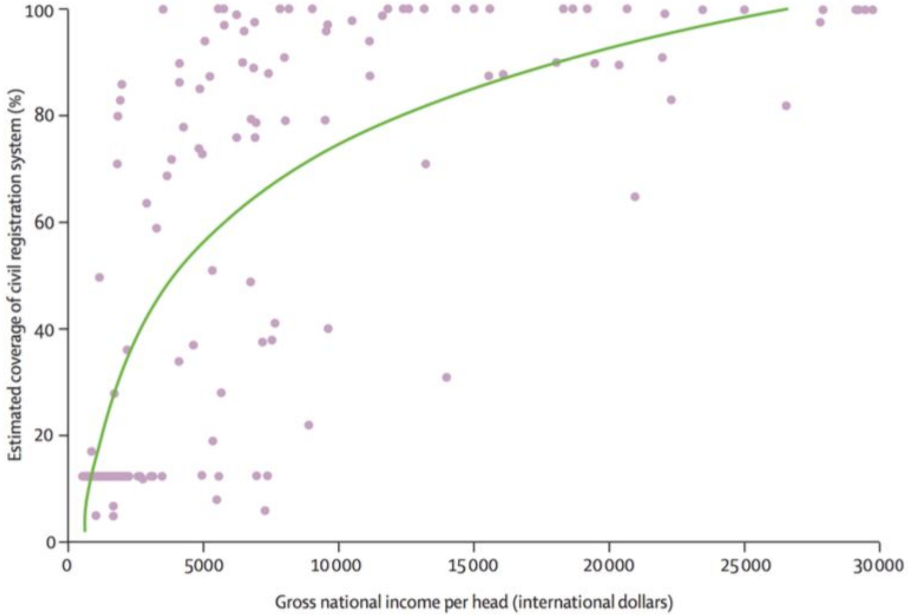
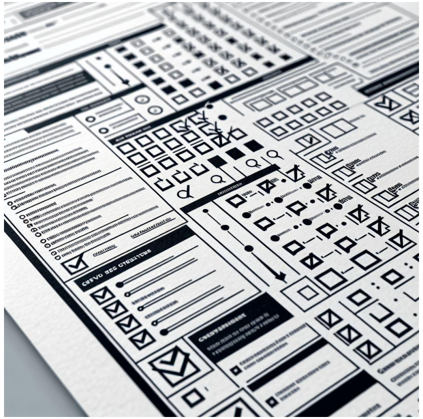


Figure: Source: Setel et al. 2007. Association between estimated coverage of civil registration and gross national income per head, 1998-2004.

Verbal Autopsy (VA)

Interviews with caregivers of the deceased.

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

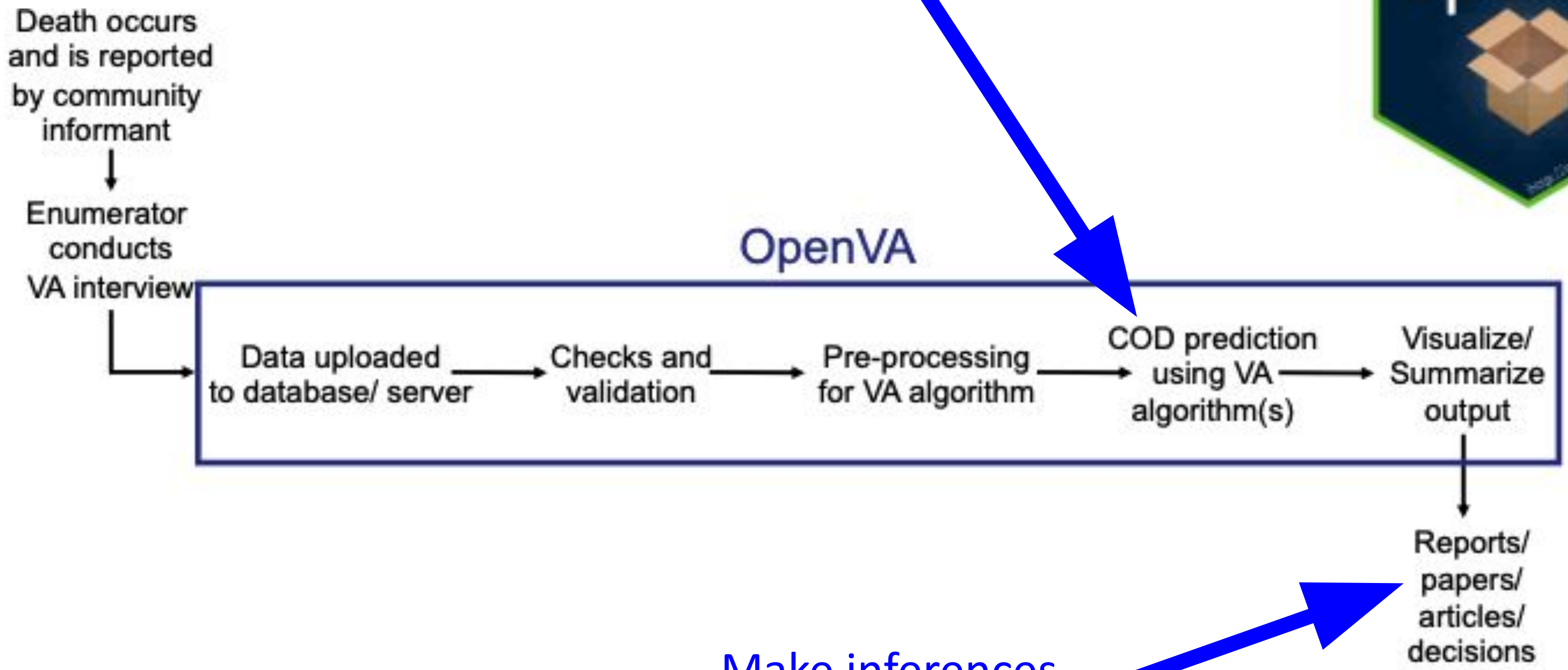
['cough', cough',' chest',' pain',' tiredness',' blood',' vision',' stomach',' pain',' ' vomit',' worms','diarrhea',' sneezed',' worm',' hot',' chest',' throat',' ' lip',' rash',' chapping',' blood',' lip',' pain',' cough',' headache',' hiccup','' defecating',' defecate',' stool',' yellowish',' ring','worms']

Mapundu et al. 2024

Interviews are burdensome on respondents (~2hr, repetitive, impersonal).

Structured Questionnaire

Make predictions



Make inferences
with those predictions

Text Narratives and Language Modeling

Research Questions:

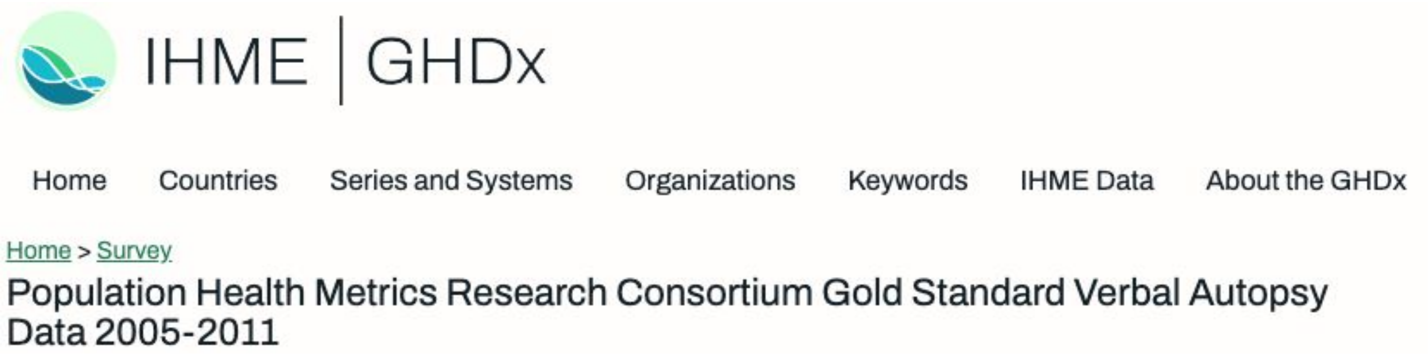
1. What if we use <u>only the text</u> narratives of the VA?

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

2. Does IPD correction change our conclusions?

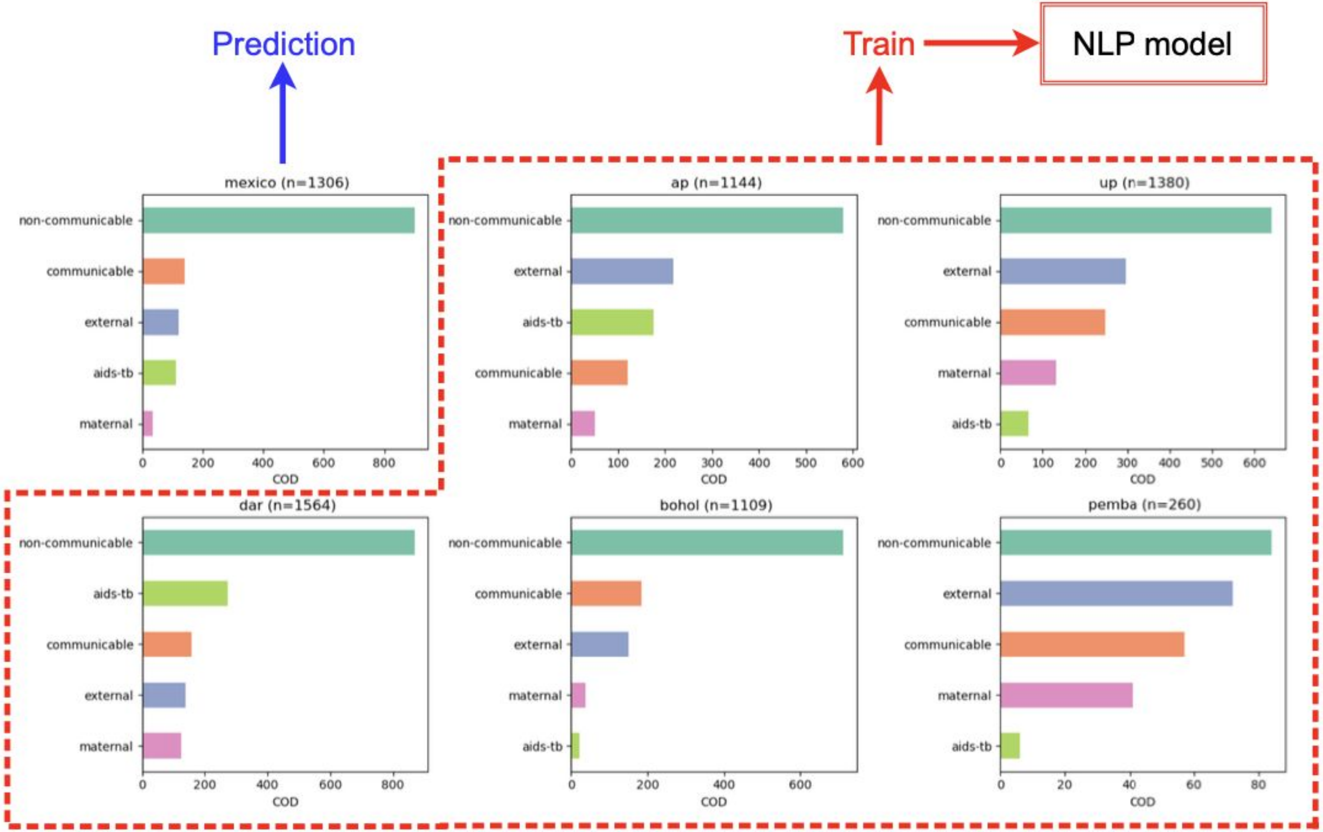
Data



- 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!

Leave-One-Out Prediction



LLM Prompt

<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

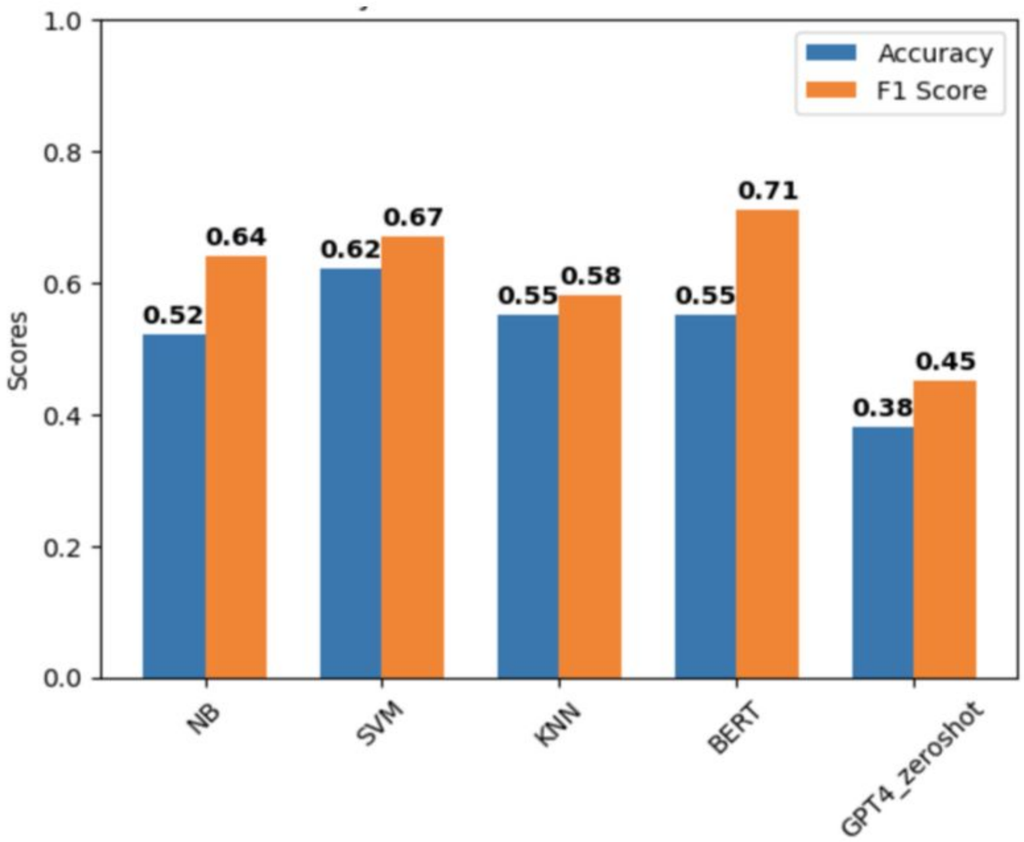
<options>
aids-tb,
communicable,
external,
maternal,
non-communicable,
unclassified
</options>

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

Prediction Accuracy

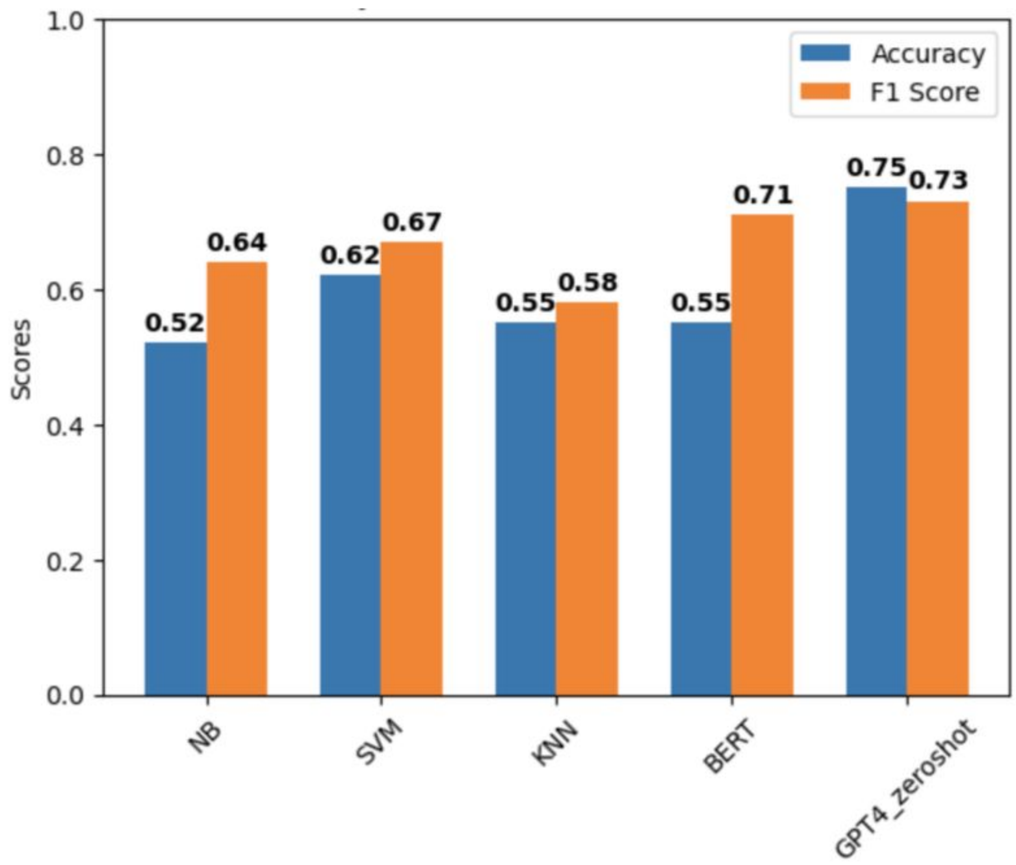


Closer Look at GPT-4 Predictions

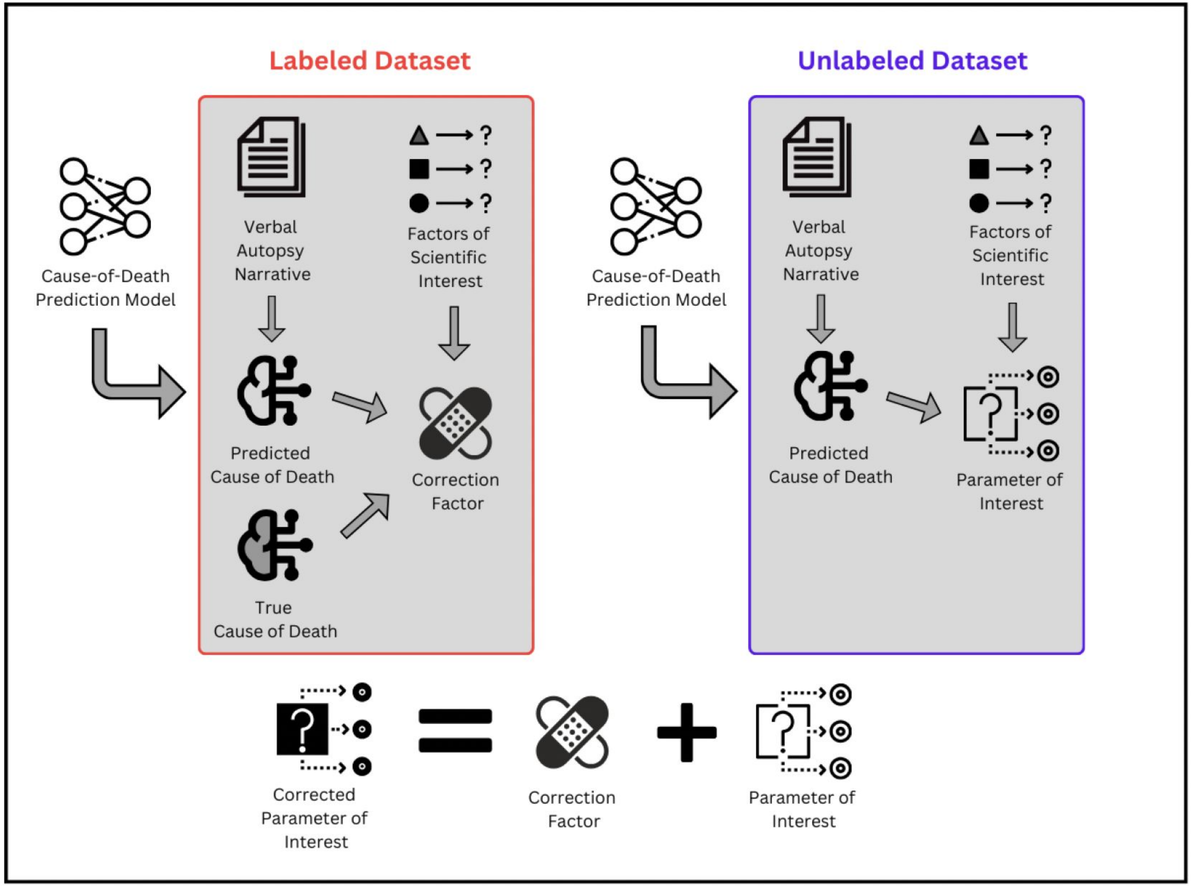
narrative	gs_cod	prediction
respondent thanked for being visited	aids-tb	The narrative does not provide enough information to determine a cause of death.
client had no additional point	non-communicable	The narrative does not provide enough information to determine the appropriate label.
the client thanked 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.

GPT-4 Actually Makes Good Predictions!



IPD on Predicted COD



Regularized Loss Function

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

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

Regularized Loss Function

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

Lambda is a tuning parameter in [0,1]

Lambda = 0 when the predicted data are all **<u>noise</u>**

Lambda = 1 when the predicted data are all **<u>signal</u>**

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

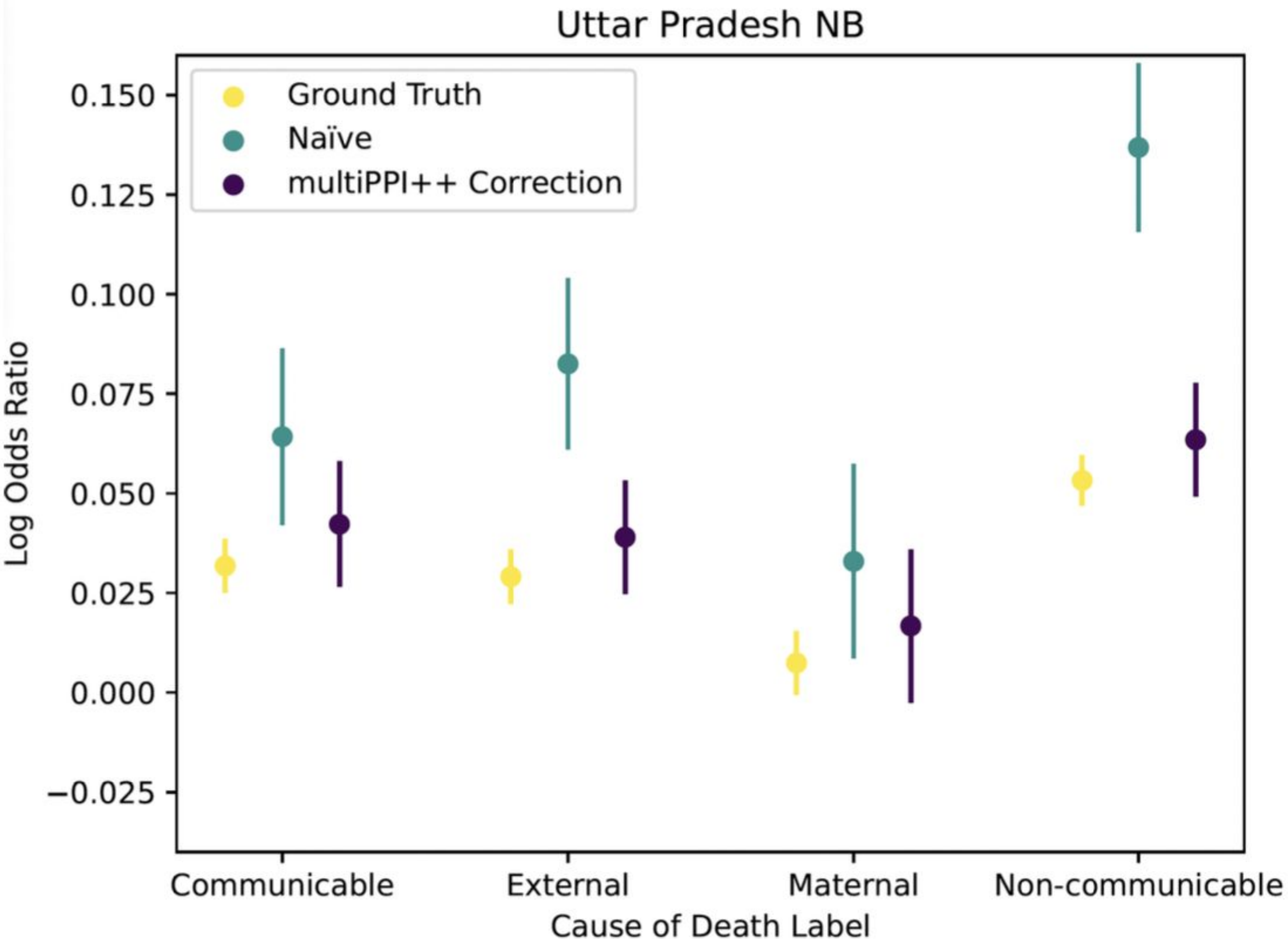
How does Age (X) vary with Cause of Death (y)?

multinomial logistic regression:

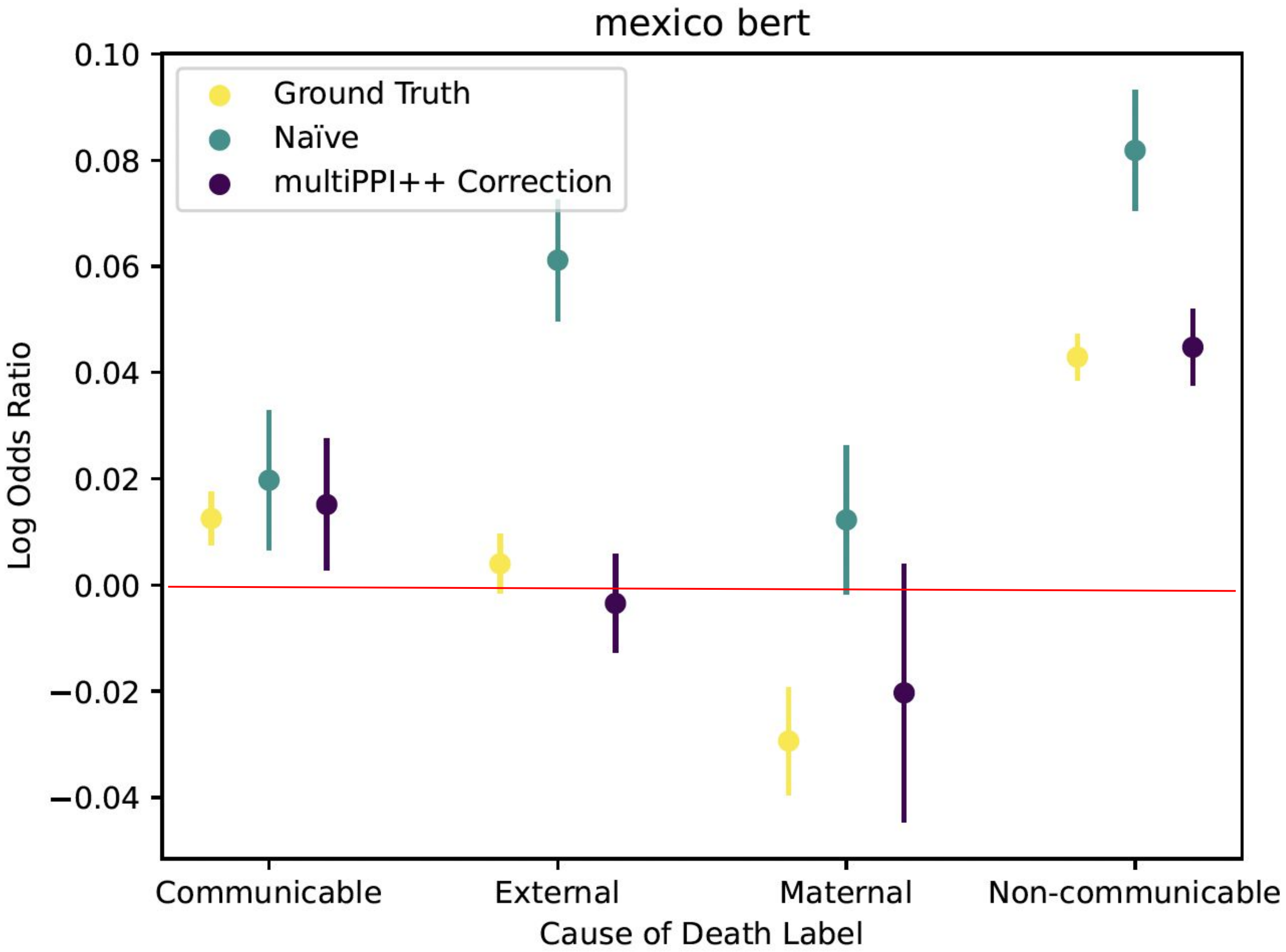
$$\log\left(\frac{p_{COD_i}}{p_{COD_{reference}}}\right) = \theta_0 + X_{age} * \theta_i$$

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

- $\theta_1, \theta_2, \theta_3, \theta_4$ 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.

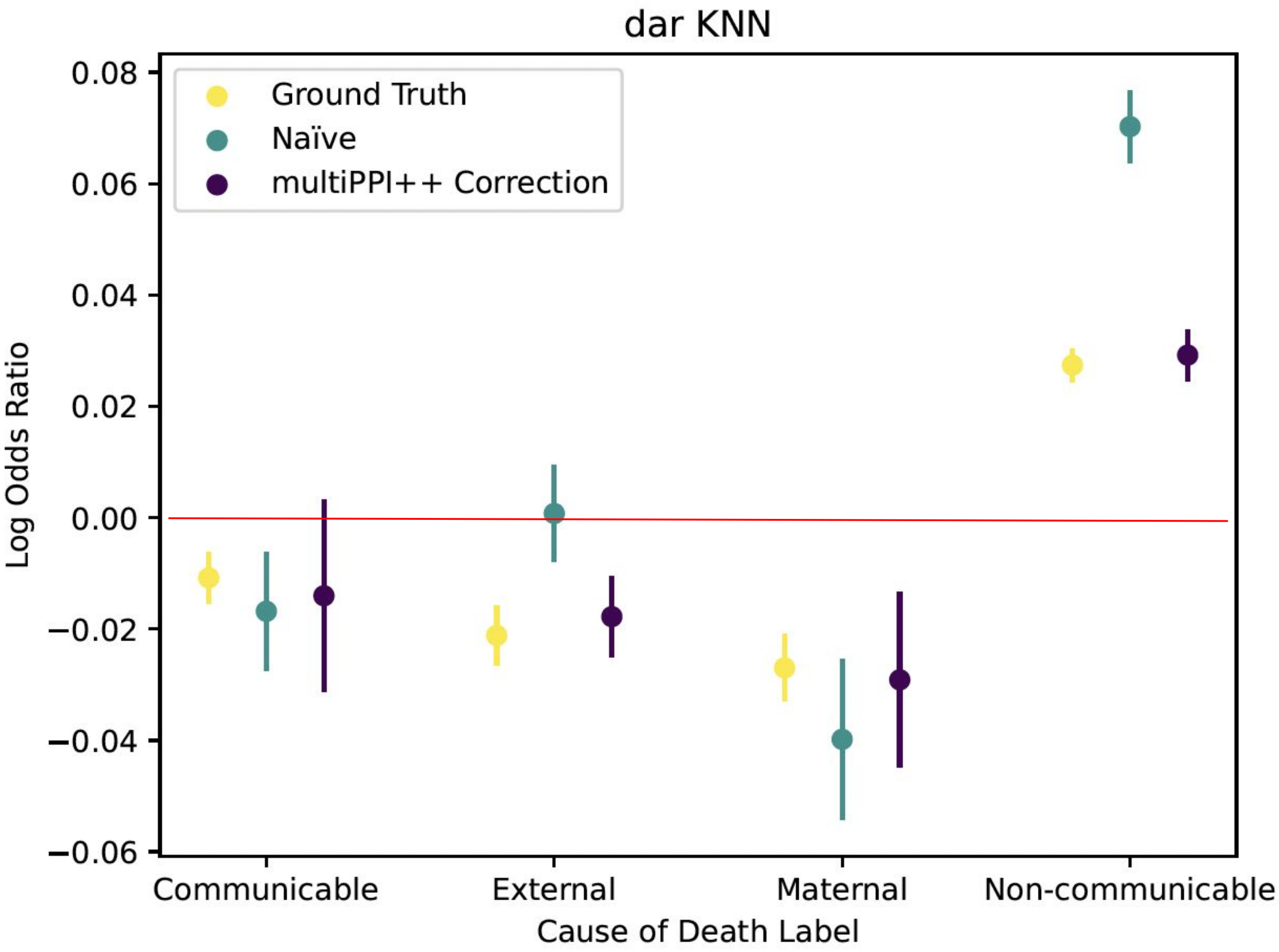


Conclusions change dramatically!!!

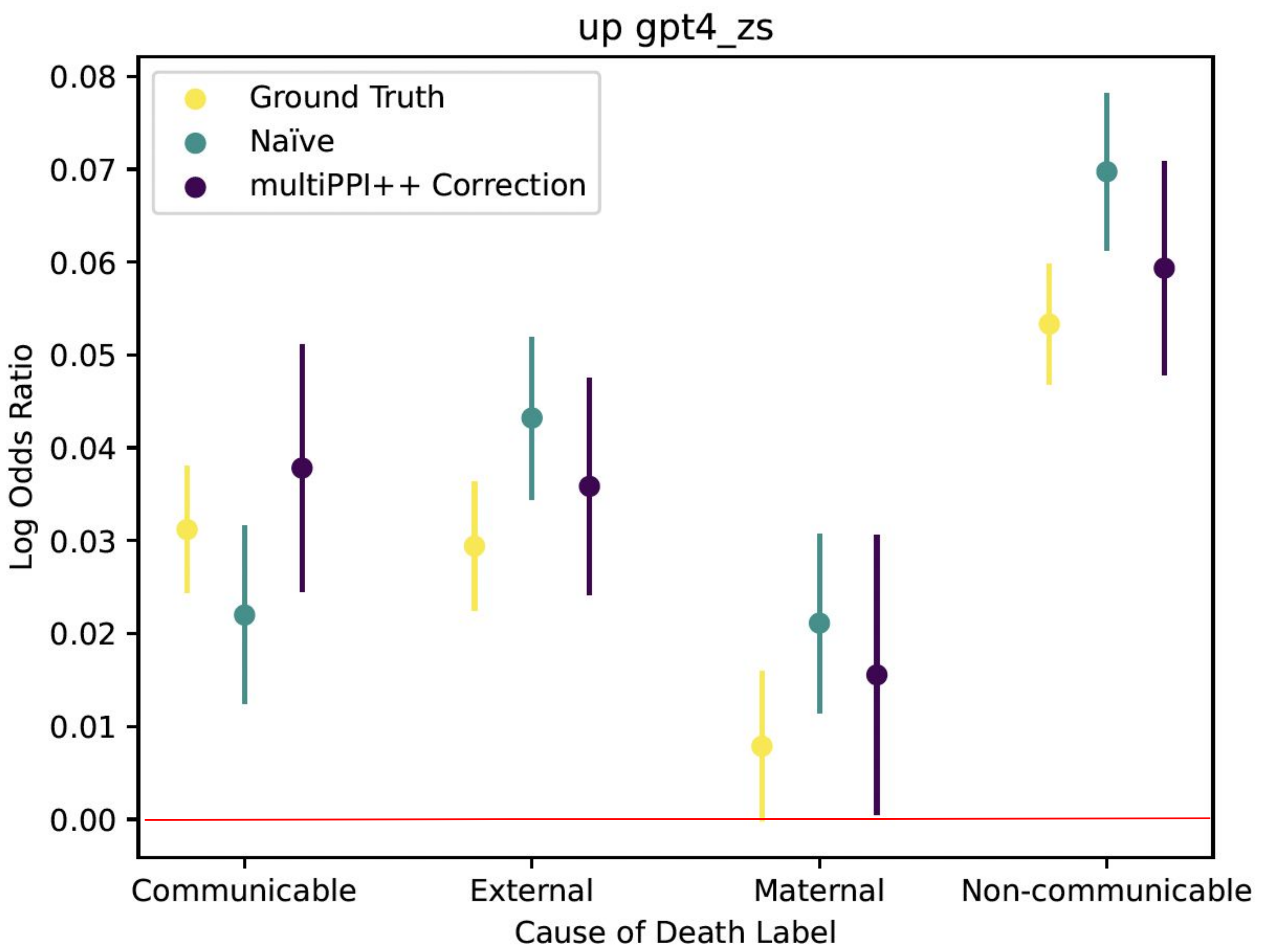


Conclusions change dramatically!!!

This time with BERT.



One more example with classical NLP



Not as dramatic, but naive estimates still exhibit substantial bias.

Switching Gears

Obesity

Despite obesity's designation as a disease, it lacks a biologically specific definition (Kraemer, Berkowitz, and Hammer 1990).

Obesity

Despite obesity's designation as a disease, it lacks a biologically specific definition (Kraemer, Berkowitz, and Hammer 1990).

Can be measured a number of ways:

1. Total percentage body fat (DXA scan)
2. Body Mass Index (BMI)
3. Waist circumference ratio

Obesity

Despite obesity's designation as a disease, it lacks a biologically specific definition (Kraemer, Berkowitz, and Hammer 1990).

Can be operationalized a number of ways:

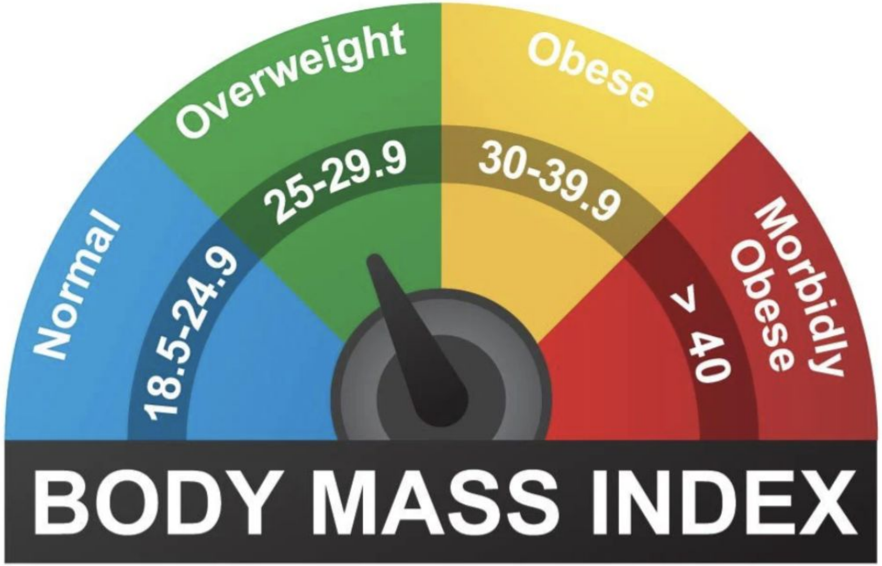
1. Total percentage body fat (DXA scan)
2. Body Mass Index (BMI)
3. Waist circumference ratio

Claim: The BMI is a Prediction Algorithm

BMI as Prediction Algorithm

BMI = weight (kg) / height (m)^2

“Healthy Weight”

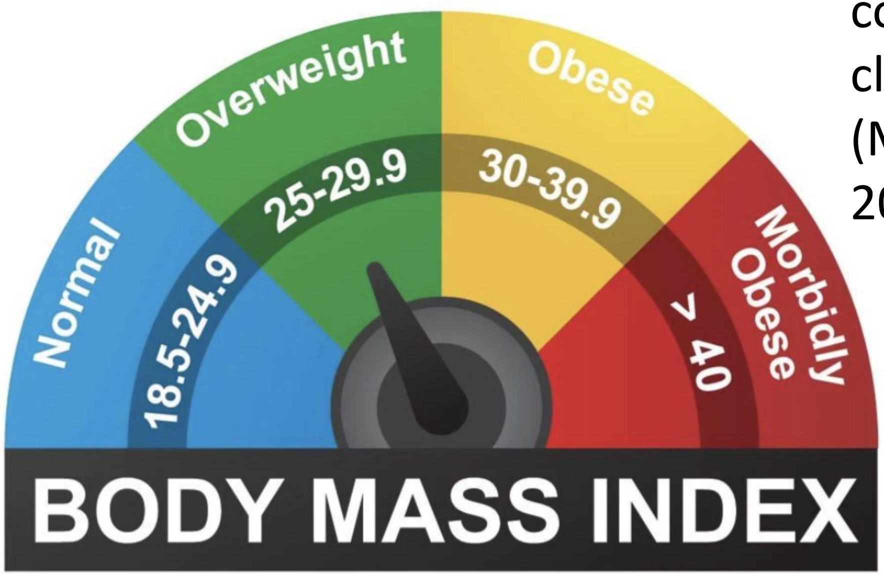


https://nutriactiva.com/blogs/bmi

BMI as Prediction Algorithm (noisy)

BMI = weight (kg) / height (m)^2

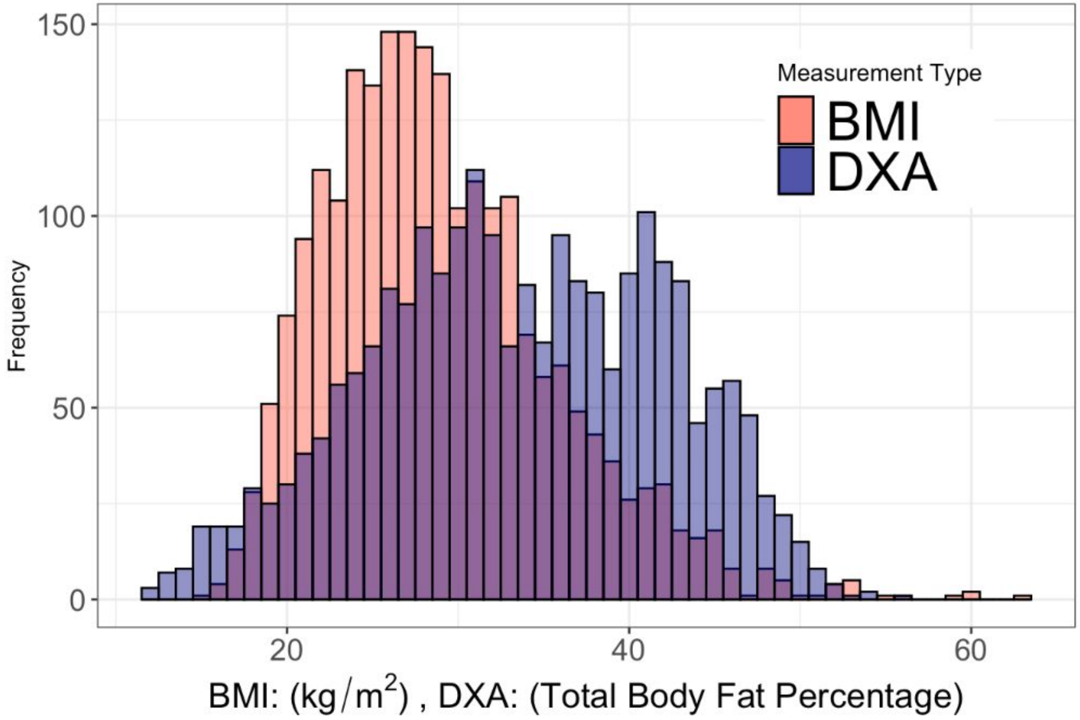
“Healthy Weight”



https://nutriactiva.com/blogs/bmi

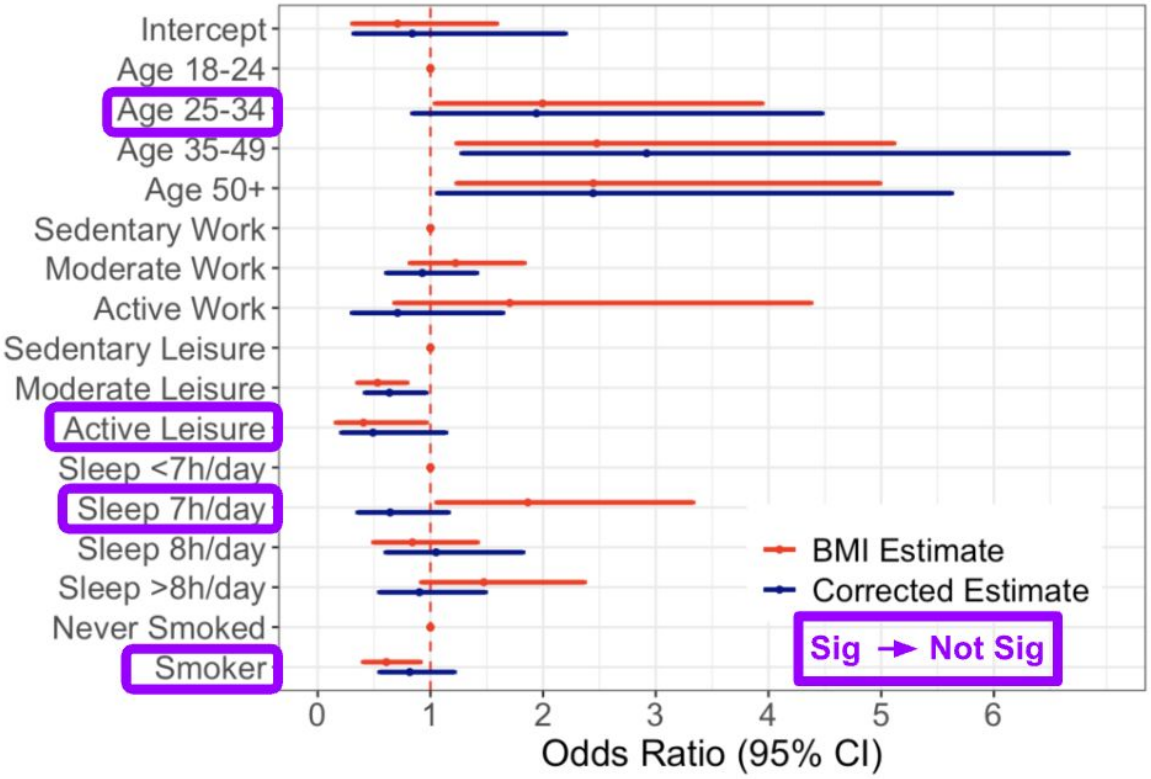
47% of patients had a fat percentage that did not correspond to their BMI classification (Monasor-Ortolá et al. 2021)

NHANES 2017



47% of patients had a fat percentage that did not correspond to their BMI classification (Monasor-Ortolá et al. 2021)

IPD Correction Procedure BMI -> DXA



Limitations

Verbal Autopsy

- Multilingual translations are not lossless.
- 5 Cause of Death categories are too broad.
- Even “ground truth” traditional autopsies can be biased.

BMI

- Healthy weight as a concept is contested.
- No obvious “ground truth” measure of obesity.

Conclusions

1. IPD calibrates statistical inference when using predicted outcomes.

2. Text narratives can be used in place of the structured VA questionnaire.

3. Performing IPD on inference using BMI can lead to different conclusions.

Thank you!!

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

IPD software is available!
Paper
Github
CRAN

Full Paper Here