From Narratives to Numbers: Valid Inference Using Language Model Predictions from Verbal Autopsies

Adam Visokay
University of Washington

Session on Machine Learning Approaches in Mortality Research

The Team

Adam Visokay

Trinity Fan

Kentaro Hoffman

Stephen Salerno

Jeff Leek

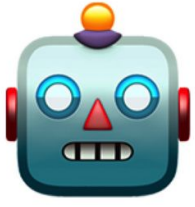
Li Liu

Tyler McCormick

Link to Paper

Motivation

You use an AI/ML algorithm to make predictions.



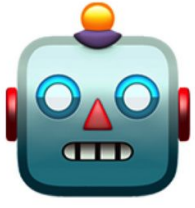
Confusion Matrix

	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)

?

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

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.

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]

🏥 $COD = \beta_1 X + \varepsilon_1$

🤖 $COD = \beta_2 X + \varepsilon_2$

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]

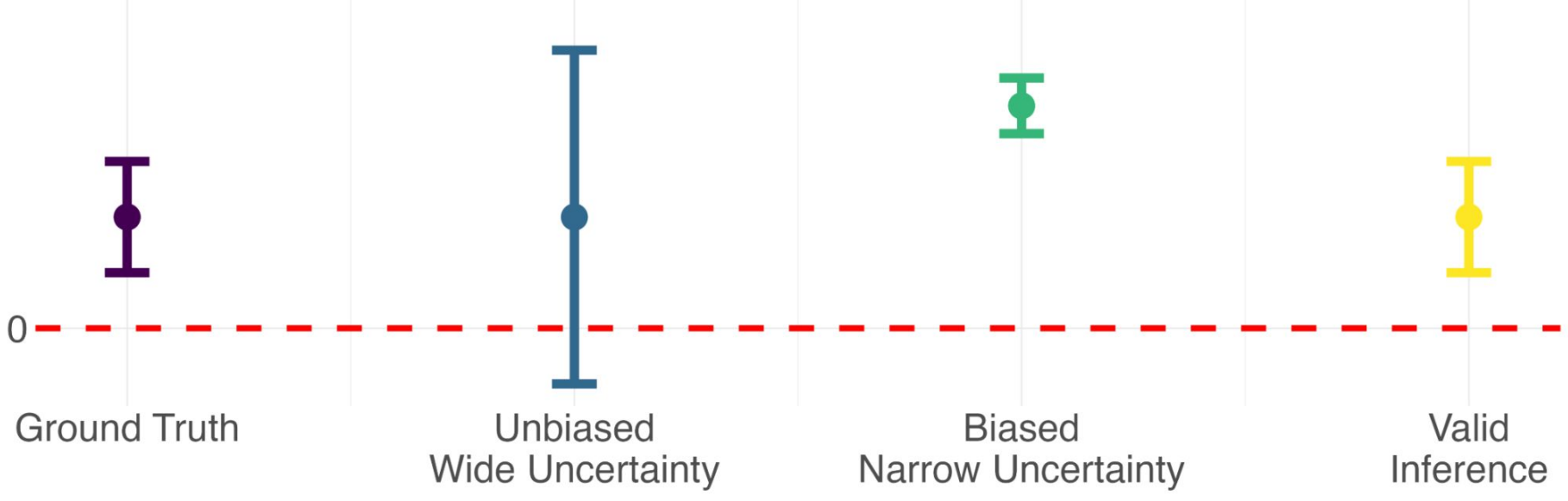
🏥 $COD = \beta_1 X + \varepsilon_1$

🤖 $COD = \beta_2 X + \varepsilon_2$

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

Inference with predicted data (IPD) can have:

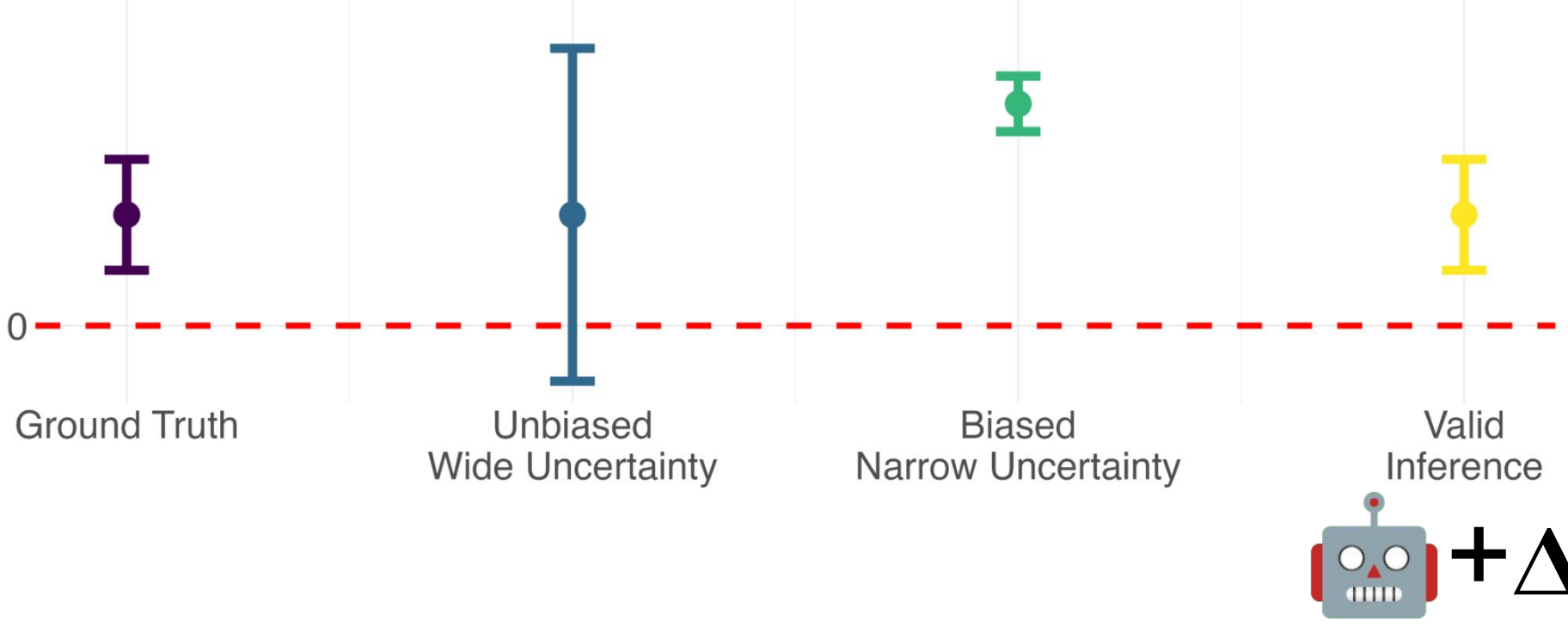
1. Biased estimates
2. Misleading uncertainty



Inference with predicted data (IPD) can have:

1. Biased estimates
2. Misleading uncertainty

Can be fixed with correction factor Δ



Contributions:

1. We use NLP models (from bag-of-words to LLMs) to predict Cause of Death from Verbal Autopsy Narratives

Contributions:

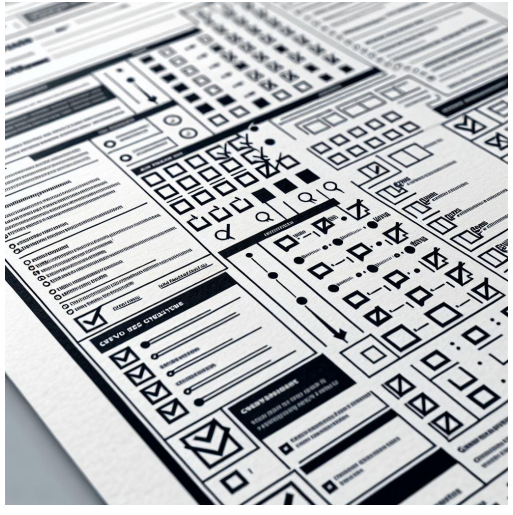
1. We use NLP models (from bag-of-words to LLMs) to predict Cause of Death from Verbal Autopsy Narratives

2. Estimate statistically valid associations between predicted cause-specific mortality and key demographic variables (e.g. Age)

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

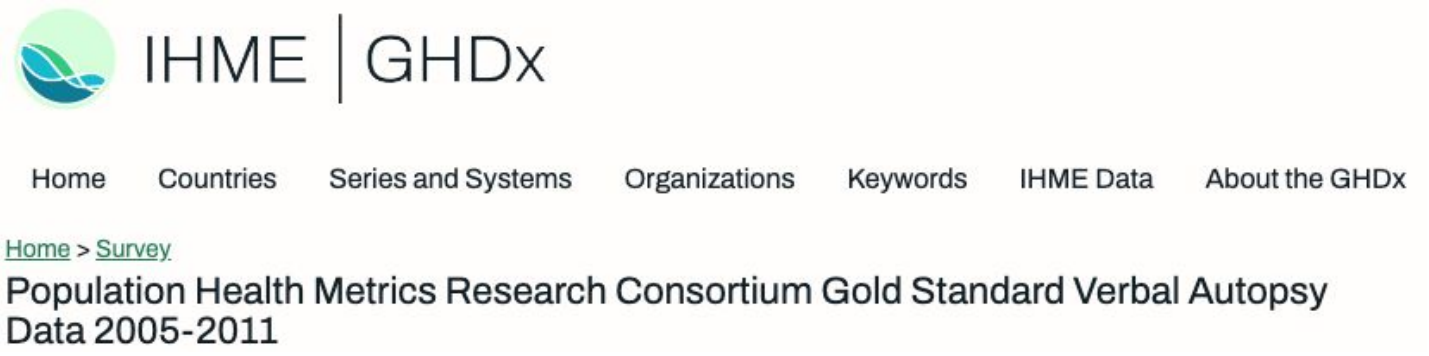
free text narrative

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).

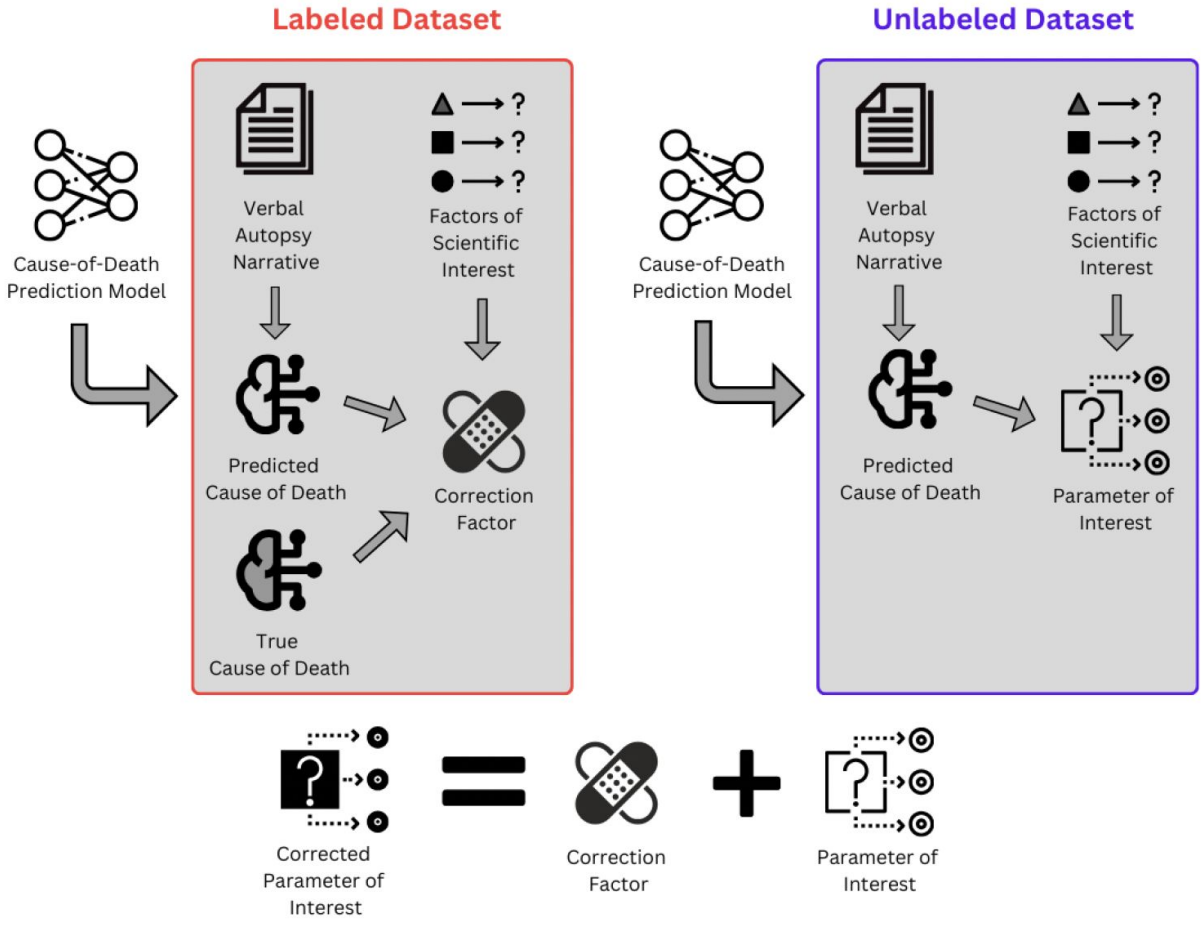
Data



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

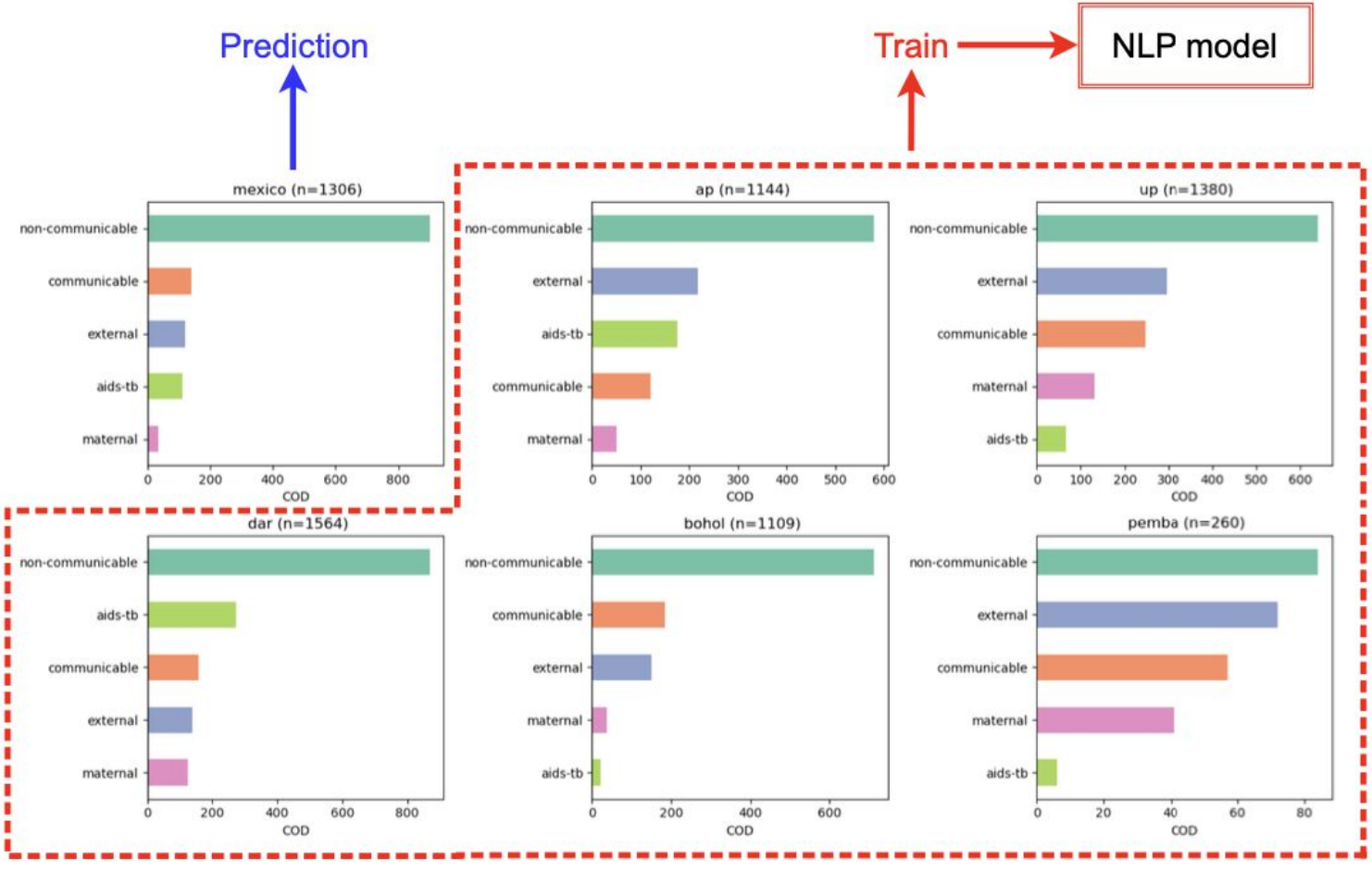
Validation set allows us to evaluate our experiment!

Inference with Predicted Data (multiPPI++)



Experimental Design - leave one out validation

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

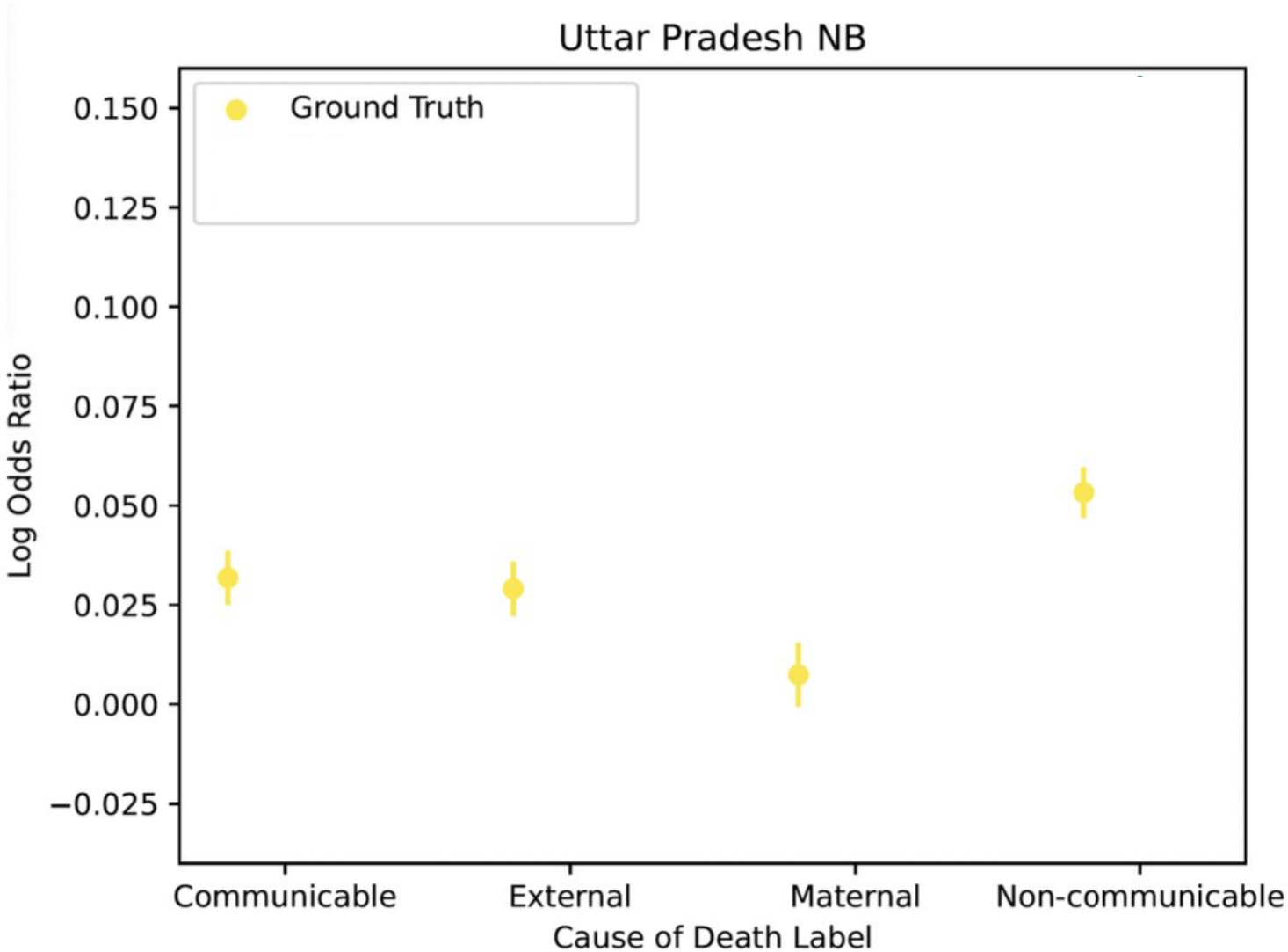


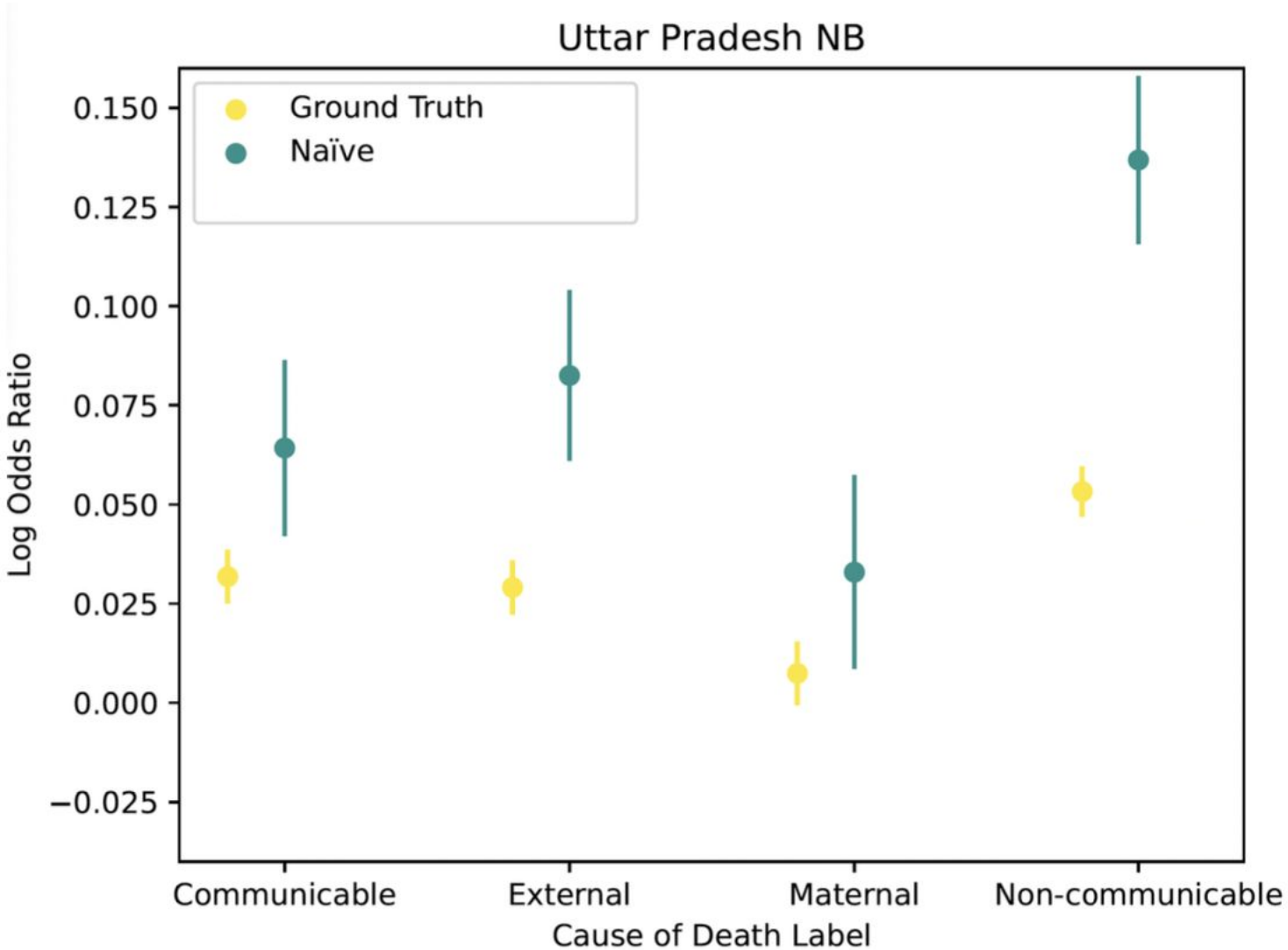
Multinomial Logistic Regression

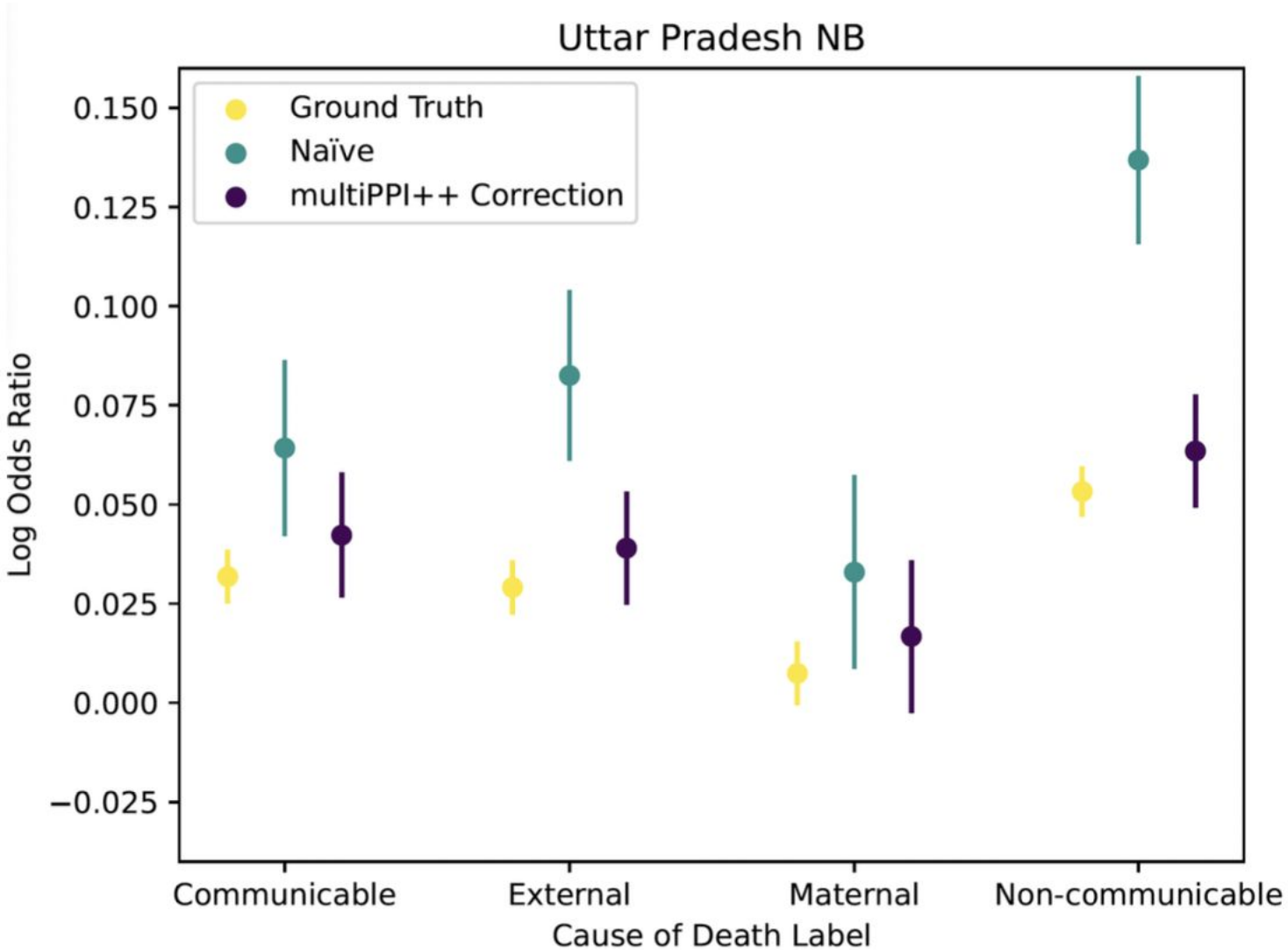
Cause specific mortality associated with Age.

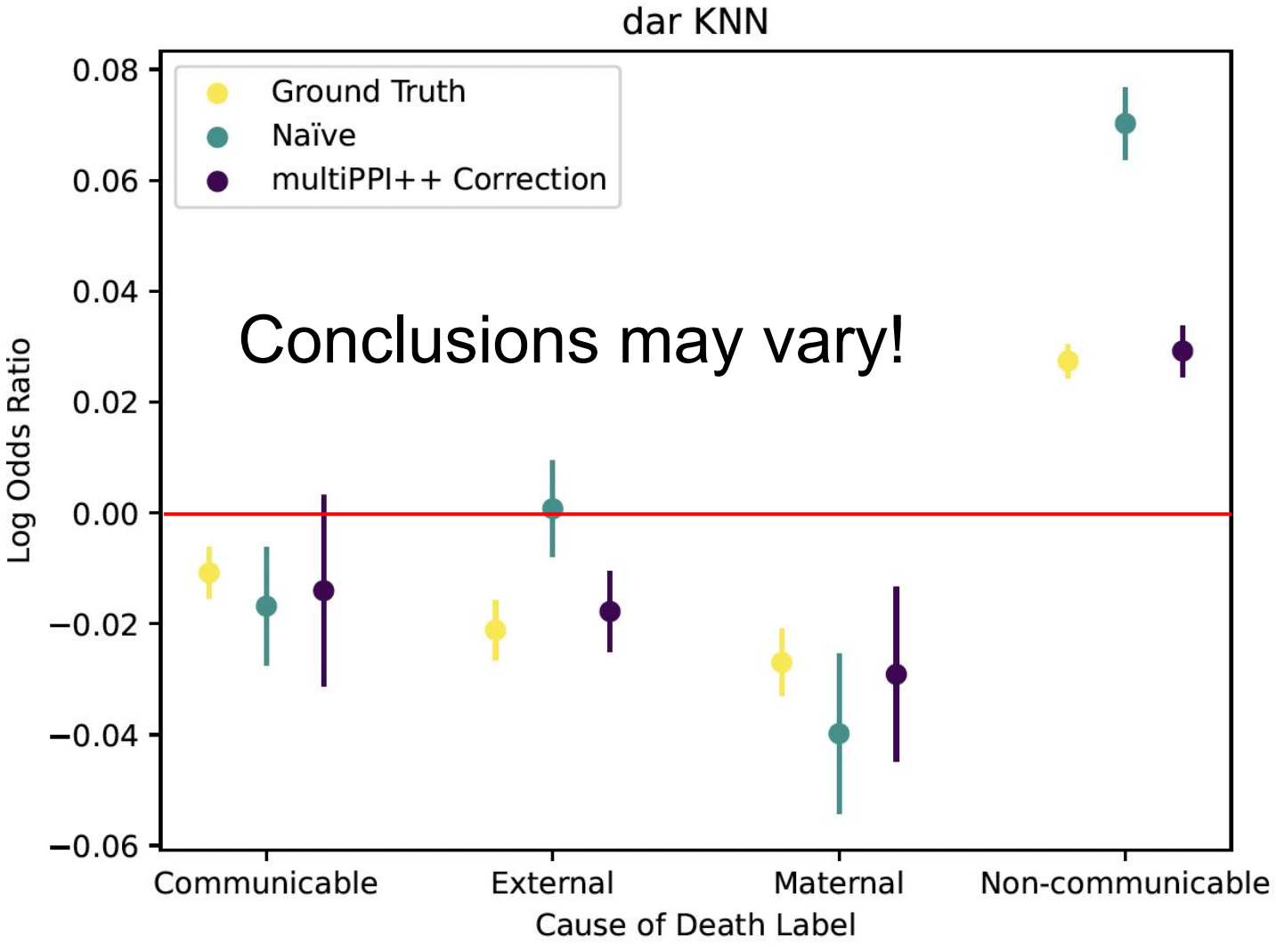
$$\log\left(\frac{p_{COD_i}}{p_{COD_{reference}}}\right) = \theta_0 + X_{age} * \theta_i$$

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
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
multiPPI++ Correction
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


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

Takeaways

1. Are you using predictions in downstream inference?

Consider a statistical calibration!

2. Verbal Autopsy text narratives are extremely valuable data sources, but vary a lot in quality.

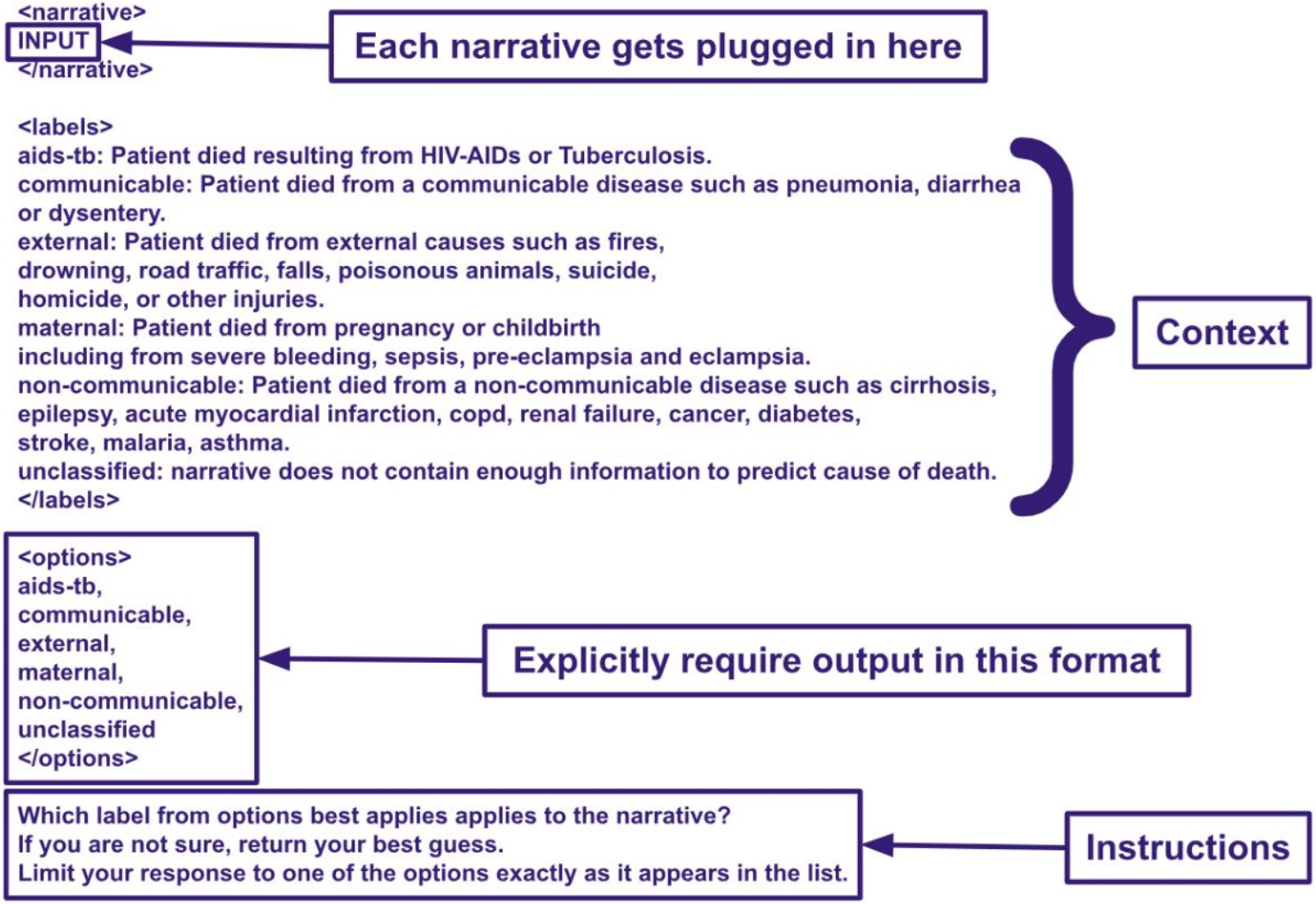
Thank you!!

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

IPD software is available!
Paper
Github
CRAN

arxiv

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

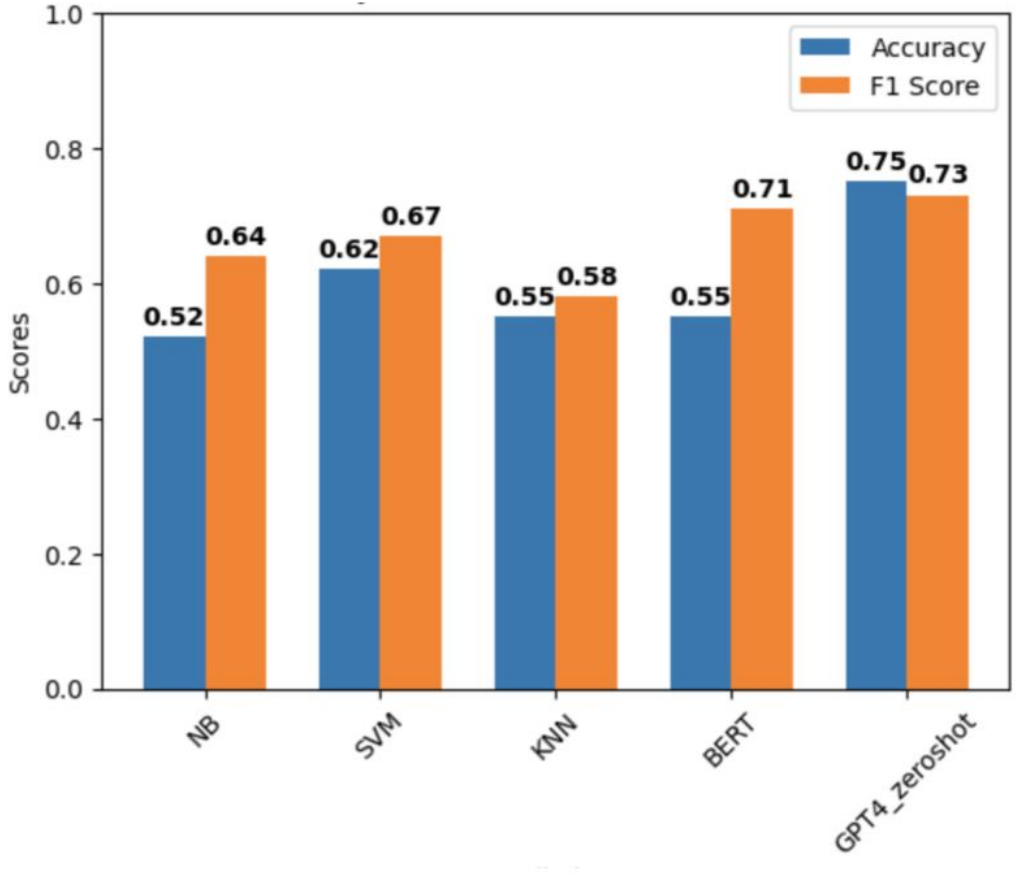
Regularized Loss Function

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

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

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