

# **Clinical Literature Entity-Relation Extraction using Pre-trained Language Models**

**HNSC 7200 Seminar  
Oct 19 2021**

**Yoonsik Park (University of Manitoba), Dr. Serena Jeblee (University of Toronto),  
Dr. Noah Crampton (University of Toronto, Mutuo Health)**

Background

Objectives

Dataset

Methods

Results

# Background

Objectives

Dataset

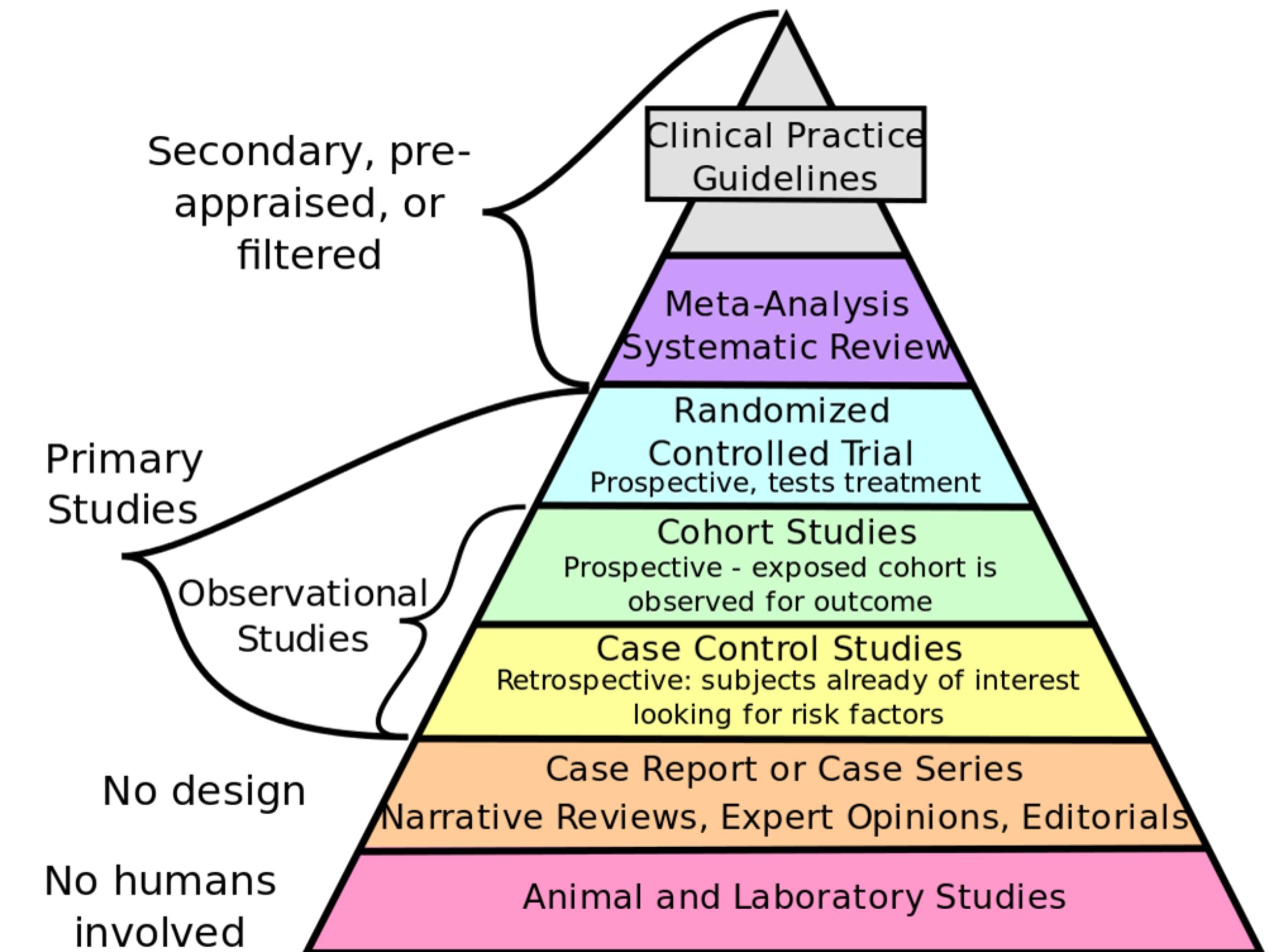
Methods

Results

# Literature Review and Meta-Analysis is Essential

## Background

- Systematic Literature Review is considered **Gold Standard**
  - PICO framework is commonly used
  - If results from primary studies are statistically combined, it is considered a **Meta-analysis**
    - Increased power, decreased bias, stronger conclusions
  - Basis for evidence-based medicine

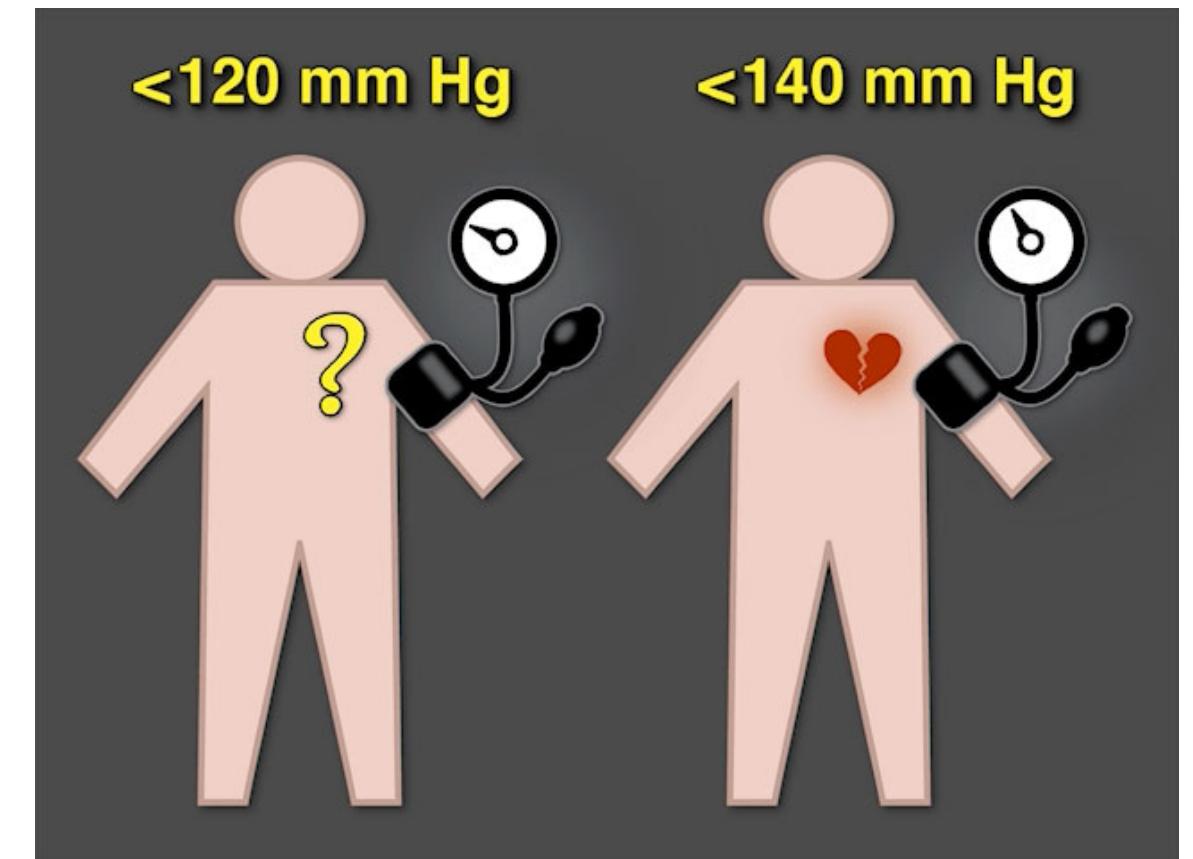


# PICO Framework Helps us Understand Clinical Trials

## Background

### Example Trial: The SPRINT, Intensive vs. standard blood pressure control (2016)

- Population: Patients aged 75 years or older
- Intervention: Systolic Blood Pressure <120 mmHg
- Comparator: Systolic Blood Pressure <140 mmHg
- Outcome (Primary): Major Adverse Cardiac Events (MACE)
  - Result: Hazard Ratio (HR) **0.66**
- Outcome (Secondary): All-cause Mortality
  - Result: Hazard Ratio (HR) **0.67**



# Systematic Reviews are Well-defined and Structured

## Background

- Steps include: search strategy, selection criteria, quantitative/qualitative analysis, summary
- Some steps require a human to read article text and extract key terms, a few examples:
  - **Applying selection criteria:** date of publication, geographical location, cohort characteristics (age, sex, etc.), methodology
  - **Study quality appraisal:** methodology, sources of bias
  - **Data collection:** risk estimates (ratios), outcome, explanatory, control variables

European Journal of Epidemiology  
<https://doi.org/10.1007/s10654-019-00576-5>

### GUIDELINES



A 24-step guide on how to design, conduct, and successfully publish a systematic review and meta-analysis in medical research

Taulant Muka<sup>1</sup> · Marija Glisic<sup>1,2</sup> · Jelena Milic<sup>3,4</sup> · Sanne Verhoog<sup>1</sup> · Julia Bohlius<sup>1</sup> · Wichor Brammer<sup>5</sup> · Rajiv Chowdhury<sup>6</sup> · Oscar H. Franco<sup>1</sup>

Received: 21 June 2019 / Accepted: 29 October 2019  
© Springer Nature B.V. 2019

### Abstract

To inform evidence-based practice in health care, guidelines and policies require accurate identification, collation, and

# Machine Learning is Rapidly Improving

## Background

- ML can be loosely defined as "training" computer algorithms to automatically perform a task that we desire
- A growing subfield of ML is Natural Language Processing (NLP), i.e. the field of processing and analyzing large amounts of natural language data (text)

### Image Classification

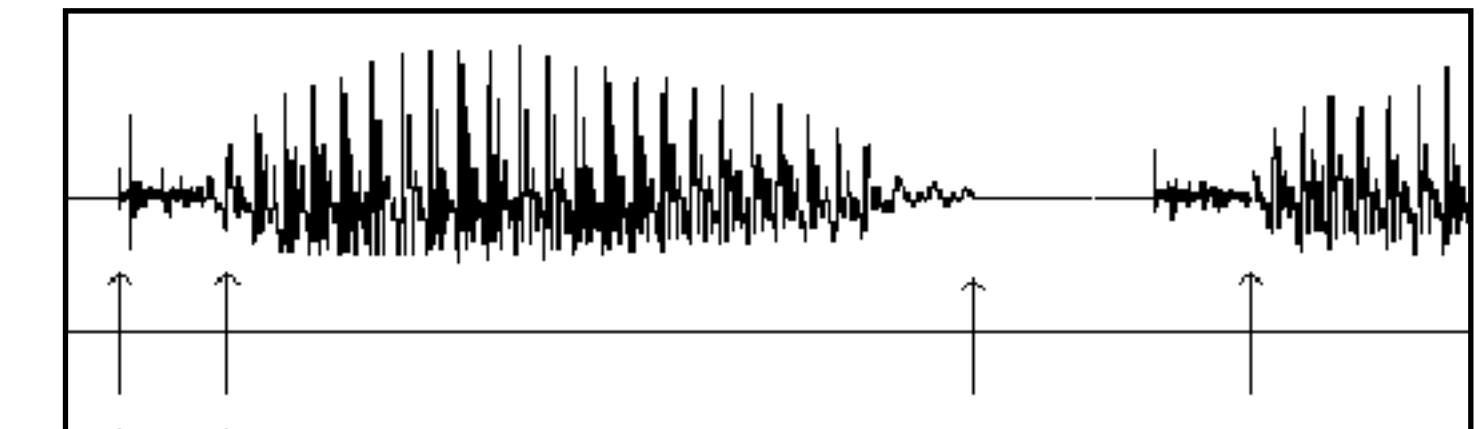


Cat



Dog

### Speech Recognition



PyTorch



TensorFlow



# Biomedical NLP is Showing Success

## Background

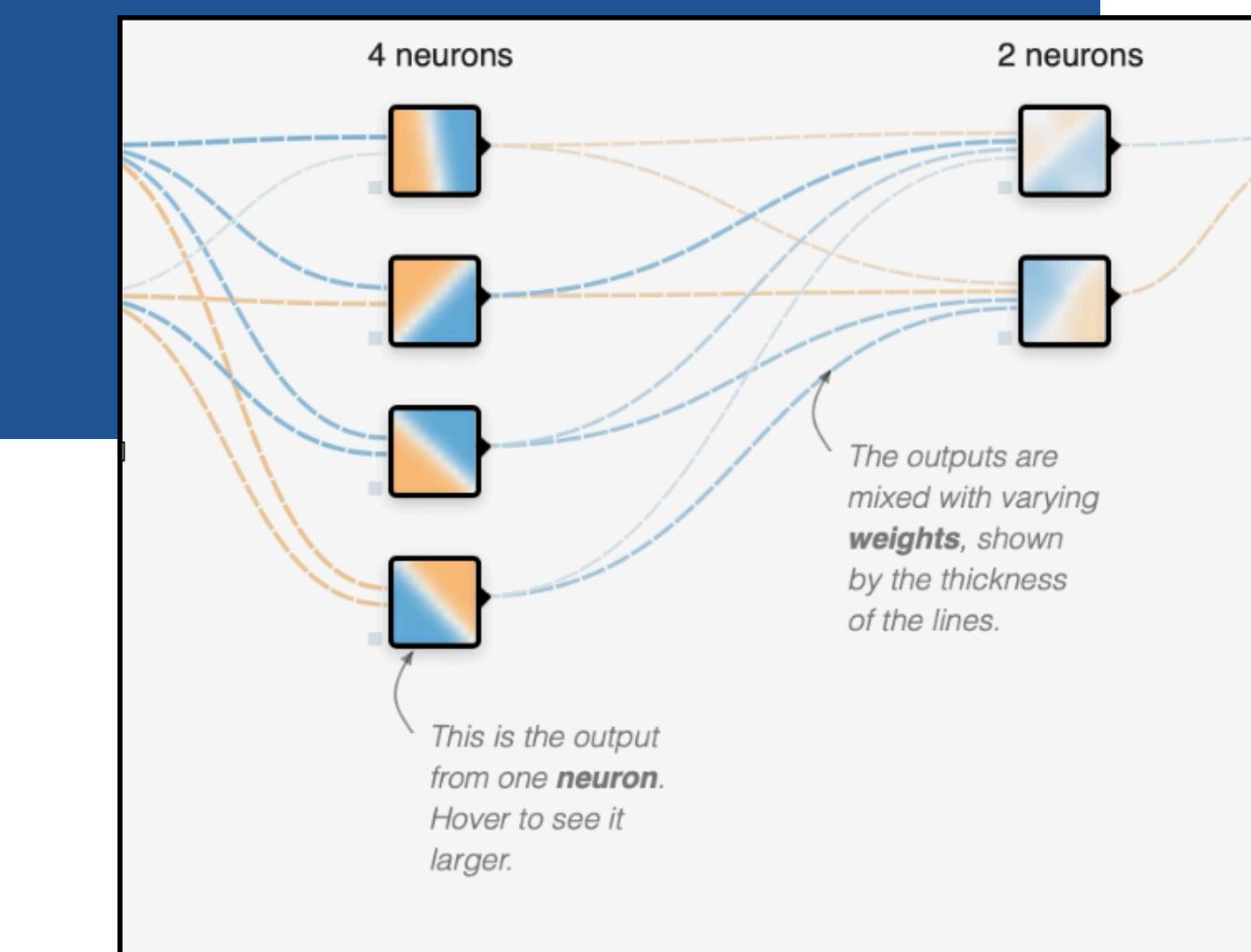
Biomedical NLP is the process of extracting and synthesizing important information from biomedical data

- Notably, Schmidt *et al.* (2020) showed that pre-trained language models were able to support systematic reviews by
  - classifying each sentence as: Population, Intervention, Comparator, or Outcome (PICO)
  - predicting a single span in a sentence that best answers a PICO question
- Limitation: a sentence may contain information from all PICO categories & multiple sections from a sentence may be needed to answer a PICO question

# Rationale

## Background

- Clinical systematic reviews are the gold-standard for evidence based medicine
- However, they are time consuming, requiring researchers, planning, and money
- 1.4 million articles released on PUBMED in 2019! We need a way to scale this process
- The structured format of a systematic review implies that we can automate parts of this process using biomedical NLP



Background

## Objectives

Dataset

Methods

Results

# **Primary Research Objective**

## Objectives

Extract risk estimates (Hazard, Risk, Odds Ratio) with their linked study variables (Explanatory, Baseline, Outcome) using an NLP model that can extract multiple entity-relations from a sentence.

In terms of PICO: Explanatory → Intervention, Baseline → Comparator

# **Secondary Research Objective**

## Objectives

Explore different variations of model to improve performance:

- Combine NER model and RE model into a single Joint Entity-Relation Extraction model
- Try different language model variations (size + corpus)
- Add context, i.e. the title + first sentence from an abstract

Background

Objectives

**Dataset**

Methods

Results

# PubMed Abstracts

## Dataset

- 79,468 PubMed abstracts published from Nov 11 2019 to Dec 31 2019 were downloaded and filtered for risk estimates → pool of 6,421 articles
  - Training set: 82 articles → 95 sentences with risk estimates
  - Test set: 31 random articles → 32 sentences with risk estimates
- Annotated entities: "RR", "OR", "HR", "Outcome", "Baseline", "Explanatory"
- Relations between the risk estimates and study variables were tagged

# Training - Example

## Dataset

Rates of MACE (RR 0.91 95% CI 0.58-1.41; p = 0.66; I<sup>2</sup> 75%), MI (RR 1.75 95% CI 0.87-3.55; p = 0.12; I<sup>2</sup> 0%) and ischemic stroke (RR 0.83 95% CI 0.53-1.31; p = 0.42; I<sup>2</sup> 0%) did not differ between the OAC monotherapy and the OAC combination therapy.

# Training - NER Task

## Dataset

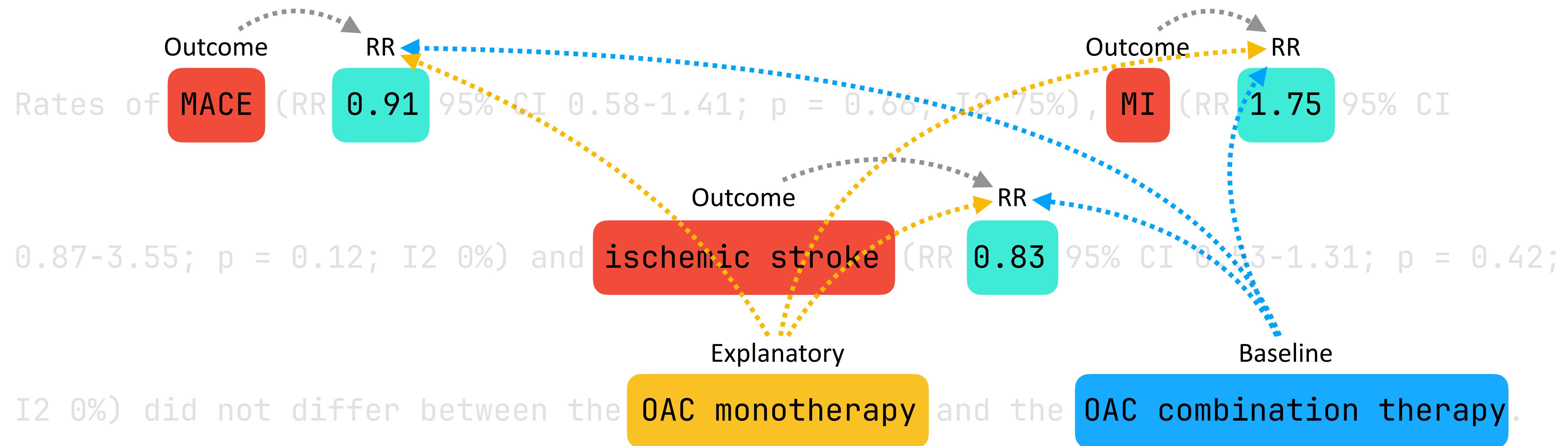
Outcome RR  
Rates of MACE (RR 0.91 95% CI 0.58-1.41; p = 0.66; I<sup>2</sup> 75%), MI (RR 1.75 95% CI

Outcome RR  
0.87-3.55; p = 0.12; I<sup>2</sup> 0%) and ischemic stroke (RR 0.83 95% CI 0.53-1.31; p = 0.42;

Explanatory Baseline  
I<sup>2</sup> 0%) did not differ between the OAC monotherapy and the OAC combination therapy.

# Training - Relation Task

## Dataset



Background

Objectives

Dataset

**Methods**

Results

# Pre-trained Language Models are a Starting Point

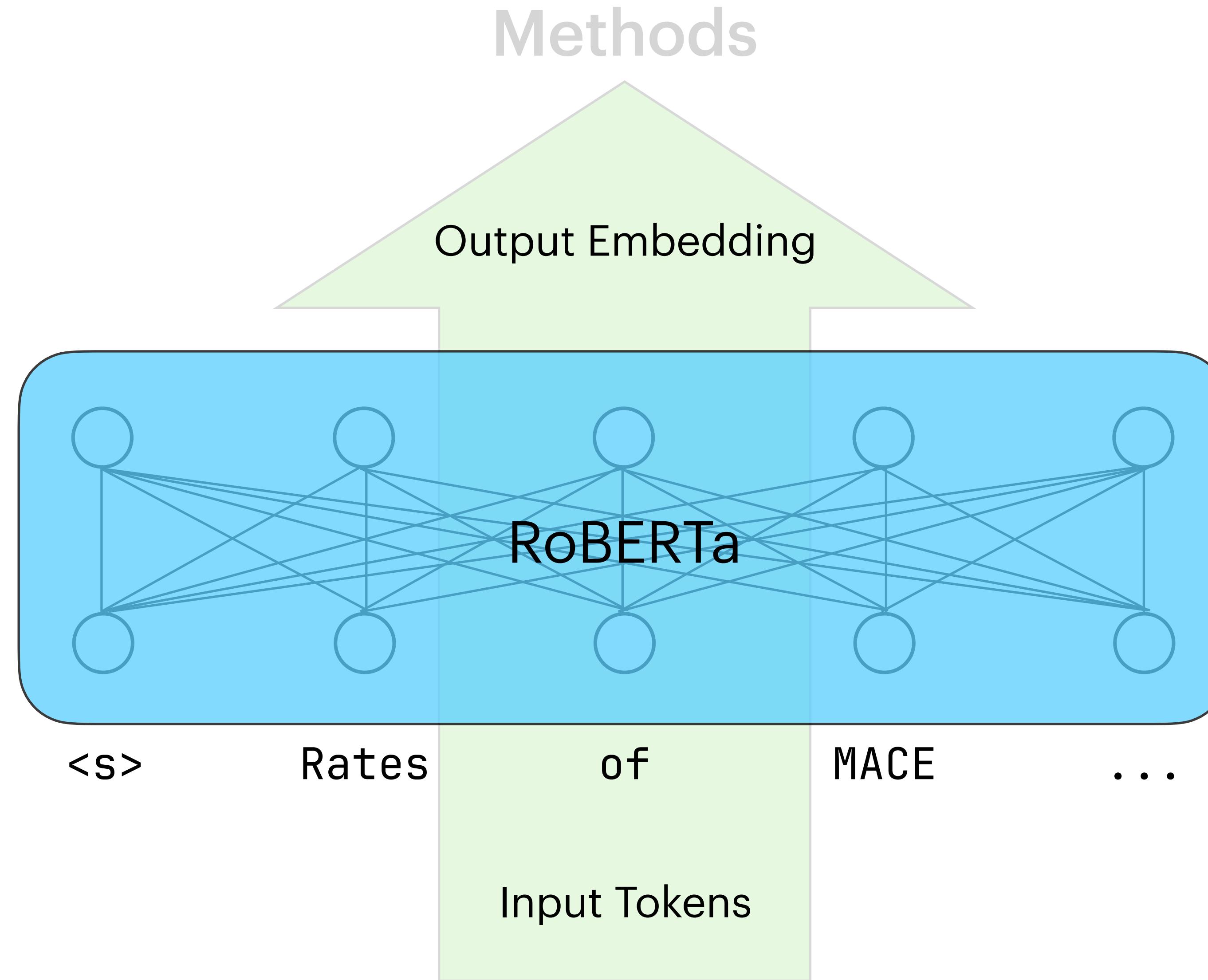
## Methods

- NLP applications usually involve a pre-training step over a large amount of unlabeled data, then a fine-tuning step for your desired task
- RoBERTa is a pre-trained language model that is state-of-the-art in many NLP tasks
- We fine-tune RoBERTa to perform our Entity-Relation Extraction task, using our own dataset



**RoBERTa is pre-trained on 160GB of  
Wikipedia entries, News articles,  
Reddit links, and Internet Stories**

# Entity-Relation Extraction Model Architecture

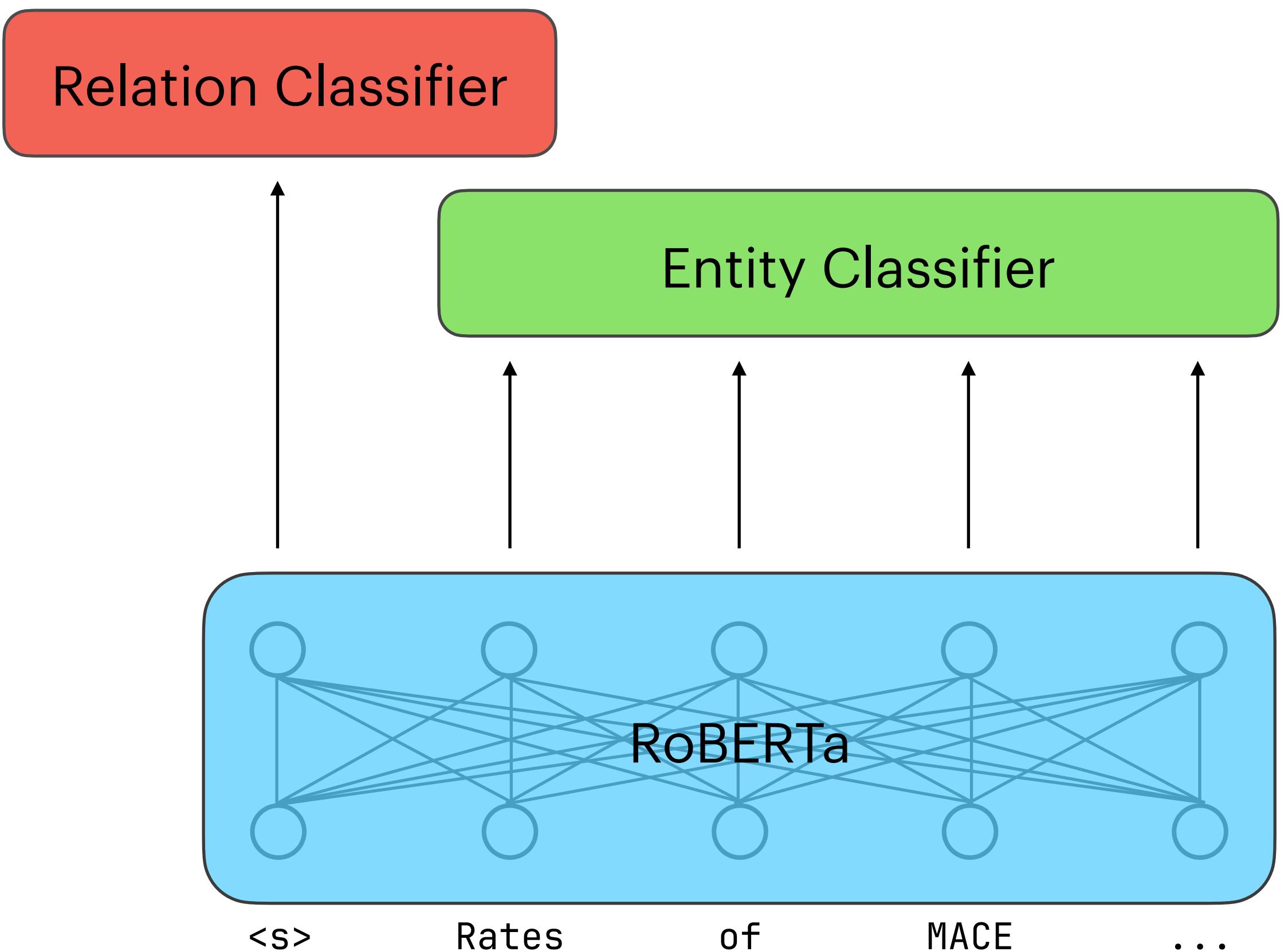


# Entity-Relation Extraction Model Architecture

## Methods

- **Entity Classifier:** Predict *Risk Estimates* and *Study Variables* for each input token
  - Uses Inside, Outside, Beginning (IOB2) labeling
- **Relation Classifier:** Predict the Relation (True or False) between a *Risk Estimate* and *Study Variable* pair
  - Query by inserting special tokens around entities, for example:

Rates of \$ MACE \$ (RR ^ 0.91 ^ 95% ...)



# Experimental Design

## Methods

- Three RoBERTa variations: RoBERTa-Large (340M), RoBERTa-Base (110M), BioMed-RoBERTa-Base (110M)
- Two architectures tested for model training:
  - **Separate models:** one trained on **NER only**, the other trained on **RE only**
  - **Joint model:** model is trained on **both NER & RE task** (Joint Entity-Relation)
- Each model category was trained with 10 different seeds
- Top 3 models on validation set were then evaluated on held-out test set

Background

Objectives

Dataset

Methods

**Results**

# Test Metrics

## Results

- Now that we have trained a model, we can see how well it performs on our test dataset (annotated examples)
- **Named Entity Recognition:** Precision, Recall, F1-Score
- **Relation Extraction:** Matthew's Correlation Coefficient, Accuracy
- **Entity-Relation Extraction:** Jaccard Set Similarity
- In simpler terms, the closer the metric is to 1.0, the better

# Test Metrics

## Results

		Predicted Class
		Yes      No
Actual Class	Yes	TP      FN
	No	FP      TN

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\text{F1 Score} = \frac{TP}{TP + 0.5(FP + FN)}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}$$

Matthew's Correlation Coefficient

$$= \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

$$\text{Jaccard Similarity} = \frac{|\text{Predicted} \cap \text{Actual}|}{|\text{Predicted} \cup \text{Actual}|}$$

# Named Entity Recognition (NER) Task

## Results

Model	Training Method	Precision	Recall	F1 Score
biomed-roberta-base	Separate	<b>0.774</b>	<b>0.789</b>	<b>0.781</b>
	Joint Entity Relation	0.745	0.789	0.767
roberta-base	Separate	<b>0.720</b>	<b>0.775</b>	<b>0.746</b>
	Joint Entity Relation	0.665	0.734	0.698
roberta-large	Separate	0.771	0.818	0.793
	Joint Entity Relation	<b>0.778</b>	<b>0.838</b>	<b>0.807</b>

# Relation Extraction (RE) Task

## Results

---

<b>Model</b>	<b>Training Method</b>	<b>MCC</b>	<b>Accuracy</b>
biomed-roberta-base	Separate	<b>0.929</b>	<b>0.967</b>
	Joint Entity Relation	0.927	0.967
roberta-base	Separate	<b>0.926</b>	<b>0.966</b>
	Joint Entity Relation	0.834	0.924
roberta-large	Separate	0.929	0.967
	Joint Entity Relation	<b>0.938</b>	<b>0.972</b>

---

# Entity-Relation Extraction Task

## Results

---

<b>Model</b>	<b>Training Method</b>	<b>Similarity</b>
biomed-roberta-base	Separate	<b>0.412</b>
	Joint Entity Relation	0.379
roberta-base	Separate	<b>0.340</b>
	Joint Entity Relation	0.265
roberta-large	Separate	0.433
	Joint Entity Relation	<b>0.443</b>

---

# Discussion

## Results

- RoBERTa-Large single model was the best for all tasks, achieving **0.81** F1 Score on NER, **0.93** MCC on RE, and **0.44** Jaccard Similarity on Entity-Relation Extraction
  - Impressive performance from just 95 training examples (<100KB of data)
- BioMed-RoBERTa-Base with separate models (NER-only + RE-only) had slightly worse performance than RoBERTa-Large single model (JER)
- Both BioMed-RoBERTa-Base and RoBERTa-Base two model (NER-only + RE-only) pipelines performed much better than their respective single model (JER) pipelines

# Applications

## Results

- Researchers can semi-automate their literature reviews and meta-analysis to keep up with the increasing number of published articles
- This research will provide a framework for other Entity-Relation Extraction applications in medical fields (code will be open-sourced)
- Sets the foundation for a clinical knowledge database
  - Let users instantly aggregate, appraise, and summarize medical articles
  - Next steps involve extracting study population, methodologies, and performing quality appraisal

**Thank You for Listening! Any Questions?**

# References

- Amin, S., Dunfield, K. A., Vechkaeva, A., & Neumann, G. (2020, July 9). *A data-driven approach for noise reduction in distantly supervised biomedical relation extraction*. ACL Anthology. <https://aclanthology.org/2020.bionlp-1.20/>
- Crowther, M., Lim, W., & Crowther, M. A. (2010, October 28). *Systematic Review and meta-analysis methodology*. American Society of Hematology. <https://doi.org/10.1182/blood-2010-05-280883>
- Eberts, M., & Ulges, A. (2021, June 28). *SPAN-based joint entity and relation extraction with Transformer pre-training*. arXiv.org. <https://arxiv.org/abs/1909.07755>
- George, A., Stead, T. S., & Ganti, L. (2020). *What's the risk: Differentiating risk ratios, odds ratios, and hazard ratios?* Cureus. <https://doi.org/10.7759/cureus.10047>
- Giorgi, J., Wang, X., Sahar, N., Shin, W. Y., Bader, G. D., & Wang, B. (2019, December 20). *End-to-end named entity recognition and relation extraction using pre-trained language models*. arXiv.org. <https://arxiv.org/abs/1912.13415>
- Gururangan, S., Marasović, A., Swayamdipta, S., Lo, K., Beltagy, I., Downey, D., & Smith, N. A. (2020, May 5). *Don't stop pretraining: Adapt language models to domains and tasks*. arXiv.org. <https://arxiv.org/abs/2004.10964>
- Jackson, R. G., Jansson, E., Lagerberg, A., Ford, E., Poroshin, V., Scrivener, T., Axelsson, M., Johansson, M., Franco, L. A., & Papa, E. (2020, July 16). *Ablations over Transformer models for biomedical relationship extraction*. F1000Research. <https://f1000research.com/articles/9-710>
- Lewis, P., Stoyanov†, V., Du, J., & Ott, M. (2020, November 19). *Pretrained Language Models for Biomedical and Clinical Tasks: Understanding and Extending the State-of-the-Art*. ACL Anthology. <https://aclanthology.org/2020.clinicalnlp-1.17.pdf>
- Li, H., Xu, Z., Taylor, G., Studer, C., & Goldstein, T. (2018, November 7). *Visualizing the loss landscape of neural nets*. arXiv.org. <https://arxiv.org/abs/1712.09913>
- Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019, July 26). *Roberta: A robustly optimized Bert pretraining approach*. arXiv.org. <https://arxiv.org/abs/1907.11692>
- Muka, T., Glisic, M., Milic, J., Verhoog, S., Bohlius, J., Bramer, W., Chowdhury, R., & Franco, O. H. (2019, November 13). *A 24-step guide on how to design, conduct, and successfully publish a systematic review and meta-analysis in Medical Research*. European journal of epidemiology. <https://pubmed.ncbi.nlm.nih.gov/31720912/>
- Schmidt, L., Weeds, J., & Higgins, J. P. T. (2020, January 30). *Data mining in clinical trial text: Transformers for classification and question answering tasks*. arXiv.org. <https://arxiv.org/abs/2001.11268>
- U.S. National Library of Medicine. (2020, July 1). *About - PUBMED*. National Center for Biotechnology Information. <https://pubmed.ncbi.nlm.nih.gov/about/>
- Wadhwa, S., Yin, K., Hughes, K. S., & Wallace, B. C. (2020, May 26). *Semi-automating knowledge base construction for cancer genetics*. arXiv.org. <https://arxiv.org/abs/2005.08146>
- Wang, Y., Yu, B., Zhang, Y., Liu, T., Zhu, H., & Sun, L. (2020, October 26). *Tplinker: Single-stage joint extraction of entities and relations through token pair linking*. arXiv.org. <https://arxiv.org/abs/2010.13415>
- Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Scao, T. L., Gugger, S., ... Rush, A. M. (2020, July 14). *Huggingface's transformers: State-of-the-art natural language processing*. arXiv.org. <https://arxiv.org/abs/1910.03771>