Entity-based Claim Representation Improves Fact-Checking of Medical Content in Tweets

Amelie Wührl and Roman Klinger
Institut für Maschinelle Sprachverarbeitung, University of Stuttgart, Germany
{amelie.wuehrl,roman.klinger}@ims.uni-stuttgart.de

Abstract

False medical information on social media poses harm to people’s health. While the need for biomedical fact-checking has been recognized in recent years, user-generated medical content has received comparably little attention. At the same time, models for other text genres might not be reusable, because the claims they have been trained with are substantially different. For instance, claims in the SciFact dataset are short and focused: “Side effects associated with antidepressants increases risk of stroke”. In contrast, social media holds naturally-occurring claims, often embedded in additional context: “If you take antidepressants like SSRIs, you could be at risk of a condition called serotonin syndrome’ Serotonin syndrome nearly killed me in 2010. Had symptoms of stroke and seizure.” This showcases the mismatch between real-world medical claims and the input that existing fact-checking systems expect. To make user-generated content checkable by existing models, we propose to reformulate the social-media input in such a way that the resulting claim mimics the claim characteristics in established datasets. To accomplish this, our method condenses the claim with the help of relational entity information and either compiles the claim out of an entity-relation-entity triple or extracts the shortest phrase that contains these elements. We show that the reformulated input improves the performance of various fact-checking models as opposed to checking the tweet text in its entirety.

1 Introduction

People use social media platforms like Twitter to discuss medical issues. This can expose them to false health-related information and poses immediate harm to people’s well-being (Suarez-Lledo and Alvarez-Galvez, 2021). While the necessity for fact-checking biomedical or scientific information has been recognized and addressed in recent years, naturally occurring arguments and claims as they are shared by social media users have received less attention.

Unfortunately, systems trained on datasets from other domains might not be reusable: The datasets that underly existing pretrained models work with atomic, edited or summarized claims (e.g., from datasets like SciFact, Wadden et al., 2020), cover claims that have been selected to be well-formed (COVID-Fact, Saakyan et al., 2021), or contain editorial content such as news headlines (Zuo et al., 2020). Examples 1–5 in Table 1 convey complex biomedical processes, they are relatively short and coherently worded. In addition, they make statements covering only one claim or fact. On the other hand, medical statements as they organically occur for example on Twitter are complex, wordy, imprecise and often ambiguous (Example 6). This makes them substantially different to the claims in established fact-checking datasets for the medical domain. To address the limitations of using only

<table>
<thead>
<tr>
<th>Id</th>
<th>Source</th>
<th>Claim</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>SciFact</td>
<td>A mutation in HNF4A leads to an increased risk of diabetes by the age of 14 years.</td>
</tr>
<tr>
<td>2</td>
<td>PubHealth</td>
<td>Scientists find clues to why binge-drinking causes binge-eating.</td>
</tr>
<tr>
<td>3</td>
<td>Zuo et al.</td>
<td>Scientists discover gene mutation involved in paraplegia and epilepsy</td>
</tr>
<tr>
<td>4</td>
<td>COVID-Fact</td>
<td>Baricitinib restrains the immune dysregulation in covid-19 patients</td>
</tr>
<tr>
<td>5</td>
<td>HealthVer</td>
<td>Frequent touching of contaminated surfaces in public areas is therefore a potential route of SARS-CoV-2 transmission.</td>
</tr>
<tr>
<td>6</td>
<td>CoVERT</td>
<td>So, they die from lung failure caused by extreme pneumonia or heart failure from sludgy blood but the root cause is #COVID19 (which can be confirmed post-mortem) so the death is counted as due to the #coronavirus &amp; NOT due to natural causes of pneumonia or heart attack</td>
</tr>
</tbody>
</table>

Table 1: Claims from different fact-checking datasets.
well-formed claims, Sarrouti et al. (2021) propose a custom dataset and fact-checking model. Their analysis indicates that naturally occurring claims contain multiple, inter-related facts compared to claims in other fact-verification datasets. Along with Zuo et al. (2022), they show that real-world medical claims in user-generated and news content are more complex and longer. In addition, Kim et al. (2021) show that fact-checking systems do not transfer robustly to colloquial claims.

This mismatch motivates extracting a check-worthy main claim from user-generated content before continuing with fact-checking. This claim detection task, which is also a central task in argument mining, can be addressed as a sequence labeling problem (Zuo et al., 2022, i.a.). While this approach requires dedicated annotated data, we propose an alternative that requires an entity annotation and relation detection system – something that has been developed for various purposes across domains (Yepes and MacKinlay, 2016; Giorgi and Bader, 2018; Scepanovic et al., 2020; Lamurias et al., 2019; Doan et al., 2019; Akkasi and Moens, 2021, i.a.). We hypothesize that the main information relevant to a claim is encoded in entities and their relations, because they convey the key semantic information within a statement and describe how they interact with each other. For our approach we propose to use that information to either find the claim token sequence or to generate a sentence representation based on entity and relation classes. Our results show that entity-based claim extraction supports fact-checking for user-generated content, effectively making it more accessible to MultiVerS (Wadden et al., 2022), an architecture recently suggested for scientific claim verification.

2 Related Work

2.1 Biomedical & Scientific Fact-Checking

The task of fact-checking is to determine the truthfulness of a claim (Thorne and Vlachos, 2018). This has been addressed for various domains (Guo et al. (2022) provide a comprehensive review). For the general domain, some work has explored judging the truthfulness of claims based on its linguistic features (Rashkin et al., 2017) or using the knowledge stored in language models as evidence (Lee et al., 2020). Fact-checking for biomedical and scientific content typically leverages external evidence sources. In biomedicine this is vital as novel research that might change or overturn an existing view on a medical claim can only be taken into account if we tap into up to date, external evidence. In other fact-checking contexts (e.g., in a political context), this requirement is not as strong since the veracity of a statement made at a particular point in time is relatively stable. In the biomedical context, given a claim, fact-checking is typically modeled as a two-step process: evidence retrieval (on document and/or sentence-level) and predicting a verdict. This verdict either determines the veracity of the claim or indicates if the evidence supports or refutes the claim. We can group existing approaches by the genre of text from which claims and evidence stem. Wadden and Lo (2021) formalize scientific claim verification in the SCIVER shared task, in which evidence and claims both originate from expert-written text. Pradeep et al. (2021) approach this task with a pipeline model, while Li et al. (2021a); Zhang et al. (2021) propose modeling one or multiple subtasks in a multi-task learning setup. Recently, Wadden et al. (2022) showed that providing more context, i.e., by representing the claim, full evidence abstract and title in a single encoding, is beneficial for inferring a final verdict.

Moving away from expert-written text, Kotonya and Toni (2020) explore verdict prediction for public health claims and use fact-checking and news articles as evidence. Hossain et al. (2020) classify a tweet into predefined categories of known misconceptions about COVID-19. Mohr et al. (2022) automatically verify tweets with COVID-19-related claims with the help of excerpts from online sources. Finally, some studies explore settings in which the claim and evidence texts originate from different genres. Zuo et al. (2020) investigate retrieving scientific evidence for biomedical claims in news texts. Sarrouti et al. (2021) check user-generated, online claims against scientific articles and Saakyan et al. (2021) explore this task for COVID-19-related claims from Reddit.

2.2 Datasets & Their Claim Characteristics

Various datasets have been proposed to facilitate scientific and medical fact-checking. One common characteristic lies in the claims contained in these datasets: they are typically well-formed and sometimes synthetic. This attribute presents a misalignment with the type of data as it occurs
on social media.

In SciFACT (Wadden et al., 2020) claims are synthetic. They are atomic summaries of claims within scientific articles. As evidence, the dataset provides abstracts from scientific literature as well as sentence-level rationales for the claims within those abstracts. PubHealth (Kotonya and Toni, 2020) and the dataset released by Zuo et al. (2020) include claims from editorial content. Kotonya and Toni (2020) provide claims and evidence texts from health-related news and fact-checking articles while Zuo et al. (2020) identify the headlines of health news articles as claims and provide the scientific papers referenced in the news article as evidence. While this genre of claims and content is targeted towards non-experts, it undergoes journalistic editing and can therefore not be characterized as occurring naturally.

We are aware of three datasets that cover user-generated claims, all with a focus on COVID-19. COVID-Fact (Saakyan et al., 2021) contains medical claims shared on a COVID-19-specific Sub-Reddit. They use the scientific articles that the users reference as evidence documents. The claims have been filtered to retain only well-formed statements. Sarrouti et al. (2021) contribute the HealthVer corpus of real-world statements from online users. To find relevant claims, they query a search engine with COVID-19 questions and use the resulting texts as claims. The provided evidence consists of abstracts from scientific articles. Similar, but exclusively focused on COVID-19 information on Twitter, CoVERT (Mohr et al., 2022) provides fact-checked tweets along with evidence texts from online resources. To the best of our knowledge, only HealthVer and CoVERT cover naturally occurring medical claims from a broad audience.

2.3 Detecting, Extracting & Generating Claims

The task of claim detection is relevant to the field of fact-checking as well as the area of argument mining. From an argument mining perspective, claim detection requires identifying the claim as the core component within the argument structure (Daxenberger et al., 2017). While mainly rooted in the political domain and social sciences (Lawrence and Reed, 2019; Vecchi et al., 2021, i.a.), some work has explored claim detection in scientific text. Achakulvisut et al. (2019); Mayer et al. (2020); Li et al. (2021b, i.a.) extract claims from clinical and biomedical articles, Wührl and Klinger (2021) classify tweets that contain medical claims.

At the same time, detecting a checkable and check-worthy claim is considered the first task within a fact-checking pipeline (Guo et al., 2022). The task of claim-check-worthiness detection is to determine if a given claim should be fact-checked. Typically this is framed as a document, sentence or claim-level classification or ranking task: Gencheva et al. (2017); Jaradat et al. (2018); Wright and Augenstein (2020, i.a.) study this task for general domain claims, in the CLEF-CheckThat! shared task (Nakov et al., 2022) participants are tasked to identify tweets that contain check-worthy claims about COVID-19. To the best of our knowledge, Zuo et al. (2022) are the first to explore this on the token level by extracting check-worthy claim sequences from health-related news texts. This shows that identifying biomedical claim sequences in longer documents for the purpose of fact-checking is understudied. The focus in fact-checking datasets and shared tasks (e.g., FEVER (Thorne et al., 2018) or SciVER (Wadden and Lo, 2021)) is typically to infer the relationship between a claim-evidence pair or on retrieving evidence for a given claim.

While in the studies described above the original phrasing of a document or claim is kept intact, some work has proposed extracting relevant semantic information to reconstruct the content that is being conveyed. Recently, Magnusson and Friedman (2021) show that fine-grained biomedical information within scientific text can be extracted into a knowledge graph to model claims. Related to our work is Yuan and Yu (2019) who extract triplets from health-related news headlines to capture medical claims. Their focus is on classifying the triples as claim or non-claim which leaves fact-checking for future work. Our objective is to extract a concise claim representation and to explore its impact on fact-checking.

Moving even further away from the original text, Wright et al. (2022) suggest generating claims from scientific text to address the data bottleneck for the downstream fact-checking task. They report comparable performances for models trained on automatically generated claims compared to a model trained on the manually labeled SciFACT claims. Their work is related to Pan et al. (2021) who generate claims to facilitate zero-shot fact
We are seeing from people who get the vaccine.

The S1 spike protein used in the Covid vaccine may be toxic to endothelial cells. Likely the cause of the clotting problems.

We propose two variants: 

We assume we have a sequence of tokens $t = (t_1, \ldots, t_n)$. In addition, we have a set of $m$ annotations $A = \{(e^{a_1}_\text{subj}, r^{a_1}, e^{a_1}_\text{obj}), \ldots, (e^{a_m}_\text{subj}, r^{a_m}, e^{a_m}_\text{obj})\}$, which encode entity and relation information, respectively $e$ and $r$. The entities are located within the token sequence $t$ and identified by their character-level onset $k$ and offset $\ell$ such that $e = (k, \ell)$, with $1 \leq k, \ell \leq n$. The relation $r$ is a string representing the relation type (e.g., “cause of”).

Building on data of this type, we suggest two claim condensation methods of the form

$$\text{condense}(t, a) \rightarrow c$$

that transform the sequence $t$ along with its annotation $a$ into a claim-like token sequence $c$. We propose two variants:

**Representing Claims as Triples.** We reduce the claim to what we hypothesize to be its core components: two medical entities and the relation between them. We hypothesize that the entities express the most relevant information with regard to the claim.

In this representation the claim is a concatenation of the subject entity tokens, the name of the relation $r$ and the object entity tokens:

$$\text{condense}_{\text{triple}}(t, a) = t^{e^{a}_\text{subj}} \circ r^a \circ t^{e^{a}_\text{obj}}$$

This approach ignores tokens that are not part of the relation or entity annotation.

**Extracting Claim Sequences.** Alternatively, we extract a subsequence from the original text. For each annotation $a$ in $t$, we apply

$$\text{condense}_{\text{seq}}(t, a) = t^{e^{a}_\text{subj}} \cdots \circ \cdots t^{e^{a}_\text{obj}}.$$ 

This retains the way the author of the original text chose to express the relation, including all tokens that are mentioned between the entities. Commonly, this also involves words that indicate the relation class, but we do not ensure that.

Figure 1 shows examples of both condensation methods. The example is taken from Mohr et al. (2022).

3 Methods

With this work we investigate if knowledge about biomedical entities allows us to extract a concise claim representation from user-generated text that enables fact-checking systems to predict a verdict. To explore this, we suggest two methods to extract and construct entity-based, claim-like statements. We assume we have a sequence of tokens $t = (t_1, \ldots, t_n)$. In addition, we have a set of $m$ annotations $A = \{(e^{a_1}_\text{subj}, r^{a_1}, e^{a_1}_\text{obj}), \ldots, (e^{a_m}_\text{subj}, r^{a_m}, e^{a_m}_\text{obj})\}$, which encode entity and relation information, respectively $e$ and $r$. The entities are located within the token sequence $t$ and identified by their character-level onset $k$ and offset $\ell$ such that $e = (k, \ell)$, with $1 \leq k, \ell \leq n$. The relation $r$ is a string representing the relation type (e.g., “cause of”).

Building on data of this type, we suggest two claim condensation methods of the form

$$\text{condense}(t, a) \rightarrow c$$

that transform the sequence $t$ along with its annotation $a$ into a claim-like token sequence $c$. We propose two variants:

**Representing Claims as Triples.** We reduce the claim to what we hypothesize to be its core components: two medical entities and the relation between them. We hypothesize that the entities express the most relevant information with regard to the claim.

In this representation the claim is a concatenation of the subject entity tokens, the name of the relation $r$ and the object entity tokens:

$$\text{condense}_{\text{triple}}(t, a) = t^{e^{a}_\text{subj}} \circ r^a \circ t^{e^{a}_\text{obj}}$$

This approach ignores tokens that are not part of the relation or entity annotation.

**Extracting Claim Sequences.** Alternatively, we extract a subsequence from the original text. For each annotation $a$ in $t$, we apply

$$\text{condense}_{\text{seq}}(t, a) = t^{e^{a}_\text{subj}} \cdots \circ \cdots t^{e^{a}_\text{obj}}.$$ 

This retains the way the author of the original text chose to express the relation, including all tokens that are mentioned between the entities. Commonly, this also involves words that indicate the relation class, but we do not ensure that.

Figure 1 shows examples of both condensation methods. The example is taken from Mohr et al. (2022).

4 Experiments

We investigate whether we can reduce the complexity of user-generated claims in order to make the information that they convey accessible to pretrained “off-the-shelf” fact-checking models and circumvent the necessity of custom training data and specialized models. We specifically explore the use of entity information to formulate a condensed version of a claim (see Section 3). More concretely, we compare how the claim representation impacts the performance of a fact-checking model.

4.1 Experimental Setting

4.1.1 Data

To test our claim condensation methods as outlined in Section 3, we assume the availability of entity and relation information. This is not an unrealistic assumption: Entity and relation extraction systems exist (Yepes and MacKinlay, 2016; Giorgi and Bader, 2018; Scepanovic et al., 2020; Lamurias et al., 2019; Doan et al., 2019; Akkasi and Moens, 2021, i.a.). For our experiments, we build on top of data that has such annotations to focus the evaluation on the extraction method instead of evaluating the quality of a NER/RE system. To the best of our knowledge, the only dataset that provides both fact checking as well as entity and relation information is the CoVERT corpus (Mohr...
et al., 2022). The dataset consists of fact-checked medical claims in tweets about COVID-19 and includes evidence texts that the annotators provided to substantiate their verdicts (SUPPORT, REFUTE, NOT ENOUGH INFORMATION). Importantly, the dataset also contains the span and type of medical entities and type of relations for each Twitter post. The entity classes cover Medical Condition, Treatment, Symptom/Side-effect and Other. Each tweet is also labeled with causative relations (not_cause_of and causative_agent_of) between a subject and an object entity. We use these annotations to formulate the condensed claims.

CoVERT includes 300 tweets with a total of 722 entities and 300 relations. In instances where multiple objects have been annotated for an entity, we choose the triple which appears first in the document under the assumption that the first claim is the main claim of the statement. For short texts, such as tweets, we hypothesize that people will mention their central, main claim at the beginning of their statement. Additionally, this emulates the atomic nature of claims in SCIFACT. CoVERT is crowd-annotated and provides three evidence texts per claim. From those, we choose the first snippet that is in line with the majority fact-checking verdict as the gold evidence. While SCIFACT assigns the NOT ENOUGH INFO (NEI) label if a given abstract does not provide enough information to come to a verdict, in CoVERT a tweet is labeled as NEI if annotators were not able to find any evidence or if there was no majority w.r.t. the verdict. We therefore drop the 36 tweets labeled NEI for our experiments, as there is no agreement w.r.t. the verdict class or no available evidence. This leaves us with 264 extracted claims (198 SUPPORT, 66 REFUTES).

4.1.2 Fact-checking Models

We use the MultiVerS architecture which has recently been suggested for evidence-based scientific fact verification (Wadden et al., 2022). At the time of writing, this approach ranks first for the shared task SCIVER. It takes as input a claim-evidence pair and represents both in a single encoding to predict a fact-checking label and identify rationales with the evidence. Claim, title and evidence abstract sentences are concatenated using separator tokens and assigned global attention during training. The model subsequently uses a classifier over the separator token that identifies the claim to predict the fact-checking verdict and an additional classification head over the separator tokens between the evidence sentences.

Based on this architecture, Wadden et al. (2022) provide various models. fever is trained on the FEVER dataset for general domain fact-checking. fever_sci is trained on a combination of FEVER data and weakly-labeled biomedical fact-checking data. The other models build on top of fever_sci and are subsequently fine-tuned on gold-labeled, in-domain data for verdict prediction and rationale selection using scifact, covidfact and healthver.

In order to test the impact of the claim representations, we do not adapt the fact-checking model, but alter the input claims.

4.1.3 Baseline: Predicting Claim Sequences

To provide a baseline and gauge the impact of entity-based claim representation as opposed to predicting a claim sequence without relying on entities, we compare to the model by Zuo et al. (2022). They train a Bi-LSTM-CRF sequence labeling model to detect check-worthy claims in medical news articles. Such articles are similar to tweets in that they are also non-expert-written text conveying medical information. Using their code base and provided training data, we recreate their best performing model which encodes the input with a combination of BioBERT and FLAIR embeddings. We use the resulting model to predict claim sequences in the CoVERT tweets. For tweets where the model predicted more than one claim sequence in a tweet we use the prediction with the highest confidence score. Note that for 6 tweets the model does not predict any claim. This leaves us with 258 claims.

4.1.4 Evaluation

We evaluate the claim condensation techniques on the downstream task of predicting a fact-checking
we report the Label-Only approach, which represents the models receive as input. The methods that we propose are:

- **condense**\textsubscript{triple} \: Claim represented by an entity–relation triple.
- **condense**\textsubscript{seq} \: Shortest token sequence which contains all entities.

Table 2 reports the results. The columns indicate which type of claim the models receive as input. For each claim type and model we report precision, recall and $F_1$ as well as the difference Δ in $F_1$ to the prediction performance for the full tweet. The table rows denote which model is used for prediction. The models (fever, fever\textsubscript{sci}, scifact, covidfact, healthver) are based on the MultiVerS architecture and vary w.r.t. the type of data they were trained on.

Overall, we observe three major patterns from the results: (1) All models show limited performance when presented with the full tweet. (2) Delimiting the claim sequence always improves verdict prediction. (3) Representing the claim based on the entities and relations is highly beneficial and leads to the most successful predictions. In the following, we discuss the results in more detail.

**Fact-checking models struggle to predict verdicts for full tweets.** In the first block of Table 2, we see that the performance is generally low (avg. $F_1$ =12.4) when the models are tasked to check the full tweet. The fever model fails to predict fact-checking verdicts for this type of input. The healthver model is the most successful ($F_1$ =45.2), presumably because its training data fits the CoV\textsc{ert} data best.

**Delimiting the claim sequence is beneficial.** Using the claim sequence prediction obtained with the Zuo et al. (2022) model as claim input shows an improved performance across all models (increases between 2.3 and 13.8pp in $F_1$ compared to predictions for the full tweet). healthver remains the most successful model (48.2 $F_1$). Notably, the covidfact model benefits most from the adapted input (Δ 13.8pp in $F_1$).

**Entity-based claim condensation improves verdict prediction.** Across all models, one of the entity and relation-based claim representations achieves the best results. For three out of five models, condense\textsubscript{triple} claims facilitate the best prediction compared to other input types. fever, fever\textsubscript{sci} and covidfact achieve $F_1$-scores of 6.5, 32.8 and 41.3, respectively. For the scifact and healthver model, using the condense\textsubscript{seq} extracted claims leads to the most reliable predictions: we observe 14.0 $F_1$ for scifact and 62.0 $F_1$ for healthver. healthver’s prediction for the condense\textsubscript{seq} claims is the most successful across all models and settings.

Across the board, the covidfact model benefits the most from delimiting the claim sequence. Here, we observe increases in $F_1$ of 13.8, 33.4 and 29.7pp when comparing the performance on the full tweet with that for a Zuo et al. (2022) claim, condense\textsubscript{triple} and condense\textsubscript{seq} claim, respectively.

The results show that both condense methods improve the performance of the fact-checking models. While the $F_1$-scores and the improvements (Δ values) vary across models, we observe the same pattern across our experiments: providing a concise claim as input leads to a more reliable verdict prediction. We also see that claims from both condense methods are more successfully checked than the predicted claim sequence identified by the Zuo et al. (2022) model. This shows that entities and relations do capture the core information of a claim relatively well. It is important to note that the condense claims are constructed using gold annotated entities and relations from CoV\textsc{ert}, while the predicted sequence is not. This needs to be taken into account when comparing the results for those claim representations.

**5 Analysis and Discussion**

We aim to understand in which cases condensing the claim is helpful and when it harms the
performance. We therefore conduct an error analysis where we compare the predictions of the best model (healthver) with the full tweet as input with predictions of that model using the claims from the most successful condensation method condense\textsubscript{eq}. The examples mentioned in the following section are displayed in Table 3. For the sake of brevity, we provide the relevant evidence documents in the Appendix, Table 4.

In total, there are 54 instances in which both claim inputs lead to a correct label. In those instances, the tweet itself tends to be fairly short (see Ex. 1a) or relatively well-formed (see Ex. 1b).

There are 74 instances in which the condensed claim sequence produces a correctly predicted label while the check based on the full tweet input does not lead to a correct result. For 66 out of 74, we observe that the label flips from NEI to the correct label (see Ex. 2a). This shows that the condensation can make the evidence more accessible to the fact-checker. In addition, Ex. 2b shows how a condensed claim is assigned a correct label, while the full tweet is not. This might be the case because the claim is presented as a question in the tweet.

In 28 cases condensing the claim leads to an incorrect prediction while checking the full tweet leads to a correct output. In 20 cases, condensing the claim changes the predicted label from the gold verdict to NEI (see Ex. 3a). This indicates that condensation can render evidence unusable. In Example 3b the condensation actually misrepresents the statement because it cuts of the phrase ‘no evidence’ before the claim. We recognize that this is a potential pitfall of the claim extraction methods.

There are 108 instances where both claim types lead to incorrectly predicted labels. In 90 out of 108 cases, both are labeled with NEI. The evidence did not provide sufficient information to check the claim. Example 4a exemplifies that, to a certain degree, the NEI label makes sense. The evidence (see Table 4) does not specifically mention long-term consequences of mRNA (vaccines). To conclude that the claim is supported by the evidence, we need to infer that long-term effects are improbable, because the mRNA does not stay in the body or affect the DNA. Similarly, in 4b, the evidence (see Table 4) requires reasoning, because ‘pneumonia’ and ‘flu-like symptoms’ which the tweet claims are primary causes of death in COVID-19 patients are not mentioned directly in the evidence. In addition, the comparative statement in the evidence of septic shock and multi-organ failure being the more prevalent causes of death as opposed to respiratory failure might pose difficulties for the model.

### 6 Conclusion and Future Work

Based on the substantial mismatch between the biomedical claims as they are most typically expected by existing fact-checking models and the nature of real-world, user-generated medical statements made on Twitter, we propose to extract entity-based claim representations. We use the entities as the core information relevant to the claim, and extract condensed claims from tweets. When presented with the adapted claim input, the fact-checking models we experiment with are able to verify the claims more reliably as opposed to when they are tasked to infer a verdict for a full
Actually wearing masks causes bacterial pneumonia which people can die from NOT covid19. Most people do not know how to don/doff PPE properly. Follow the science Big Guy!

Up to half of hospitalized COVID patients have elevated levels of antiphospholipid antibodies, or antibodies that cause blood clots to form. Patients with these antibodies are much more likely to have severe respiratory disease and kidney injury. #COVID19

“It’s unclear if his death was related to the virus.” This is why we perform autopsies. There is a significant likelihood that #COVID19 played a role in that it is known to affect endothelial cells & has been shown to cause neurological symptoms including stroke.

Are you aware that the vaccines could cause miscarriage? The real data regarding covid is that there are tiny numbers, percentage wise, of generally healthy people under the age of 60 that die from COVID or that get admitted into ICU. Are you worried about cancer too?

The predominant symptoms of ‘long COVID’ are psychological in nature, with anxiety and depression being most common. But those of course are also exactly the conditions which have been caused in, literally, millions of people, especially young people, by the lockdowns.

Know the facts! There is no evidence that #COVID19 #vaccines cause #infertility, says @username @username & @username #NIAW2021 #InfertilityAwareness

Covid is no joke, this is why we need the vaccine. We know that mRNA doesn’t cause long term effects since it decomposes in your body within 1-2 hours. Please everyone, get vaccinated as soon as you can!

I never said Covid-19 wasn’t a real coronavirus. And deaths linked to Covid-19 are primarily caused directly from pneumonia, or flu-like symptoms. The classifications for influenza and pneumonia reporting changed when Covid-19 appeared. Facts.

<table>
<thead>
<tr>
<th>id</th>
<th>full tweet</th>
<th>cond_{seq}</th>
<th>Pred.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a</td>
<td>Actually wearing masks causes bacterial pneumonia which people can die from NOT covid19. Most people do not know how to don/doff PPE properly. Follow the science Big Guy!</td>
<td>masks causes bacterial pneumonia</td>
<td>F C G</td>
</tr>
<tr>
<td>1b</td>
<td>Up to half of hospitalized COVID patients have elevated levels of antiphospholipid antibodies, or antibodies that cause blood clots to form. Patients with these antibodies are much more likely to have severe respiratory disease and kidney injury. #COVID19</td>
<td>elevated levels of antiphospholipid antibodies, or antibodies that cause blood clots to form</td>
<td>R R R</td>
</tr>
<tr>
<td>2a</td>
<td>“It’s unclear if his death was related to the virus.” This is why we perform autopsies. There is a significant likelihood that #COVID19 played a role in that it is known to affect endothelial cells &amp; has been shown to cause neurological symptoms including stroke.</td>
<td>COVID19 played a role in that it is known to affect endothelial cells &amp; has been shown to cause neurological symptoms</td>
<td>N S S</td>
</tr>
<tr>
<td>2b</td>
<td>Are you aware that the vaccines could cause miscarriage? The real data regarding covid is that there are tiny numbers, percentage wise, of generally healthy people under the age of 60 that die from COVID or that get admitted into ICU. Are you worried about cancer too?</td>
<td>vaccines could cause miscarriage</td>
<td>S R R</td>
</tr>
<tr>
<td>3a</td>
<td>The predominant symptoms of ‘long COVID’ are psychological in nature, with anxiety and depression being most common. But those of course are also exactly the conditions which have been caused in, literally, millions of people, especially young people, by the lockdowns.</td>
<td>long COVID’ are psychological in nature, with anxiety</td>
<td>S N S</td>
</tr>
<tr>
<td>3b</td>
<td>Know the facts! There is no evidence that #COVID19 #vaccines cause #infertility, says @username @username &amp; @username #NIAW2021 #InfertilityAwareness</td>
<td>vaccines cause #infertility</td>
<td>S R S</td>
</tr>
<tr>
<td>4a</td>
<td>Covid is no joke, this is why we need the vaccine. We know that mRNA doesn’t cause long term effects since it decomposes in your body within 1-2 hours. Please everyone, get vaccinated as soon as you can!</td>
<td>mRNA doesn’t cause long term effects</td>
<td>N N S</td>
</tr>
<tr>
<td>4b</td>
<td>I never said Covid-19 wasn’t a real coronavirus. And deaths linked to Covid-19 are primarily caused directly from pneumonia, or flu-like symptoms. The classifications for influenza and pneumonia reporting changed when Covid-19 appeared. Facts.</td>
<td>deaths linked to Covid-19 are primarily caused directly from pneumonia</td>
<td>S S R</td>
</tr>
</tbody>
</table>

Table 3: Example predictions for full tweets vs. condense_{seq} claims. For each error category, we provide two examples (a and b). The predictions are made by the healthver model. F: full tweet as input, C: condensed with method condense_{seq}, G: gold annotation. S: Supports, R: Refutes, N: Not enough information.
benefit evidence retrieval as well (Hanselowski et al., 2018).

7 Ethical Considerations

Unreliable fact-checking evidence and verdicts potentially exacerbate the spread of misinformation because they lend false credibility to harmful health-related information. Therefore, it is greatly important to carefully evaluate and analyze automatic fact-checking systems before their predictions can be used reliably.

It is important to acknowledge that by extracting a claim sequence from a broader statement, we might omit essential context. This could impact the statement’s meaning, its intended gravity or generally misrepresent the claim that the author originally meant to convey. To alleviate this and contextualize an automatically generated verdict, it is important to design applications which are transparent with respect to the input claims and prediction process.

Acknowledgements

This research has been conducted as part of the FIBISS project which is funded by the German Research Council (DFG, project number: KL 2869/5-1). We thank the anonymous reviewers for their valuable feedback.

References


Eva Maria Vecchi, Neele Falk, Iman Jundi, and Gabriella Lapesa. 2021. Towards argument mining...


### A Appendix

Table 4 shows examples from the CoV ERT dataset along with gold evidence documents and fact-checking verdicts.

<table>
<thead>
<tr>
<th>id</th>
<th>full tweet</th>
<th>evidence</th>
<th>Gold</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a</td>
<td>Actually wearing masks causes bacterial pneumonia which people can die from NOT covid19. Most people do not know how to don/doff PPE properly. Follow the science Big Guy!</td>
<td>There’s no evidence that mask-wearing causes bacterial pneumonia.</td>
<td>R</td>
</tr>
<tr>
<td>1b</td>
<td>Up to half of hospitalized COVID patients have elevated levels of antiphospholipid antibodies, or antibodies that cause blood clots to form. Patients with these antibodies are much more likely to have severe respiratory disease and kidney injury. #COVID19</td>
<td>The NIH-supported study, published in Science Translational Medicine, uncovered at least one of these autoimmune antiphospholipid (aPL) antibodies in about half of blood samples taken from 172 patients hospitalized with COVID-19. Those with higher levels of the destructive autoantibodies also had other signs of trouble. They included greater numbers of sticky, clot-promoting platelets and NETs, webs of DNA and protein that immune cells called neutrophils spew to ensnare viruses during uncontrolled infections, but which can lead to inflammation and clotting. These observations, coupled with the results of lab and mouse studies, suggest that treatments to control those autoantibodies may hold promise for preventing the cascade of events that produce clots in people with COVID-19.</td>
<td>S</td>
</tr>
<tr>
<td>2a</td>
<td>“It’s unclear if his death was related to the virus.” This is why we perform autopsies. There is a significant likelihood that #COVID19 played a role in that it is known to affect endothelial cells &amp; has been shown to cause neurological symptoms including stroke.</td>
<td>Some people with COVID-19 either initially have, or develop in the hospital, a dramatic state of confusion called delirium. Although rare, COVID-19 can cause seizures or major strokes. Muscular weakness, nerve injury, and pain syndromes are common in people who require intensive care during infections.</td>
<td>S</td>
</tr>
<tr>
<td>2b</td>
<td>Are you aware that the vaccines could cause miscarriage? The real data regarding covid is that there are tiny numbers, percentage wise, of generally healthy people under the age of 60 that die from COVID or that get admitted into ICU. Are you worried about cancer too?</td>
<td>Miscarriages have been reported following vaccination, but there’s no evidence to show vaccines were the cause. The number of miscarriages reported after vaccination does not appear to exceed the number you would ordinarily expect.</td>
<td>R</td>
</tr>
<tr>
<td>3a</td>
<td>The predominant symptoms of ‘long COVID’ are psychological in nature, with anxiety and depression being most common. But those of course are also exactly the conditions which have been caused in, literally, millions of people, especially young people, by the lockdowns.</td>
<td>This phenomenon has led to short term as well as long term psychosocial and mental health implications for children and adolescents. The quality and magnitude of impact on minors is determined by many vulnerability factors like developmental age, educational status, pre-existing mental health condition, being economically underprivileged or being quarantined due to infection or fear of infection.</td>
<td>S</td>
</tr>
<tr>
<td>3b</td>
<td>Know the facts! There is no evidence that #COVID19 #vaccines cause infertility, says @username @username &amp; @username #NIAW2021 #InfertilityAwareness</td>
<td>There’s no evidence that approved vaccines cause fertility loss. Although clinical trials did not study the issue, loss of fertility has not been reported among thousands of trial participants nor confirmed as an adverse event among millions who have been vaccinated.</td>
<td>S</td>
</tr>
<tr>
<td>4a</td>
<td>Covid is no joke, this is why we need the vaccine. We know that mRNA doesn’t cause long term effects since it decomposes in your body within 1-2 hours. Please everyone, get vaccinated as soon as you can!</td>
<td>It’s important to know that mRNA doesn’t affect your genes in any way because it never enters the nucleus of cells, where your DNA is kept. After the mRNA does its job, it breaks down and is flushed out of your system within hours.</td>
<td>S</td>
</tr>
<tr>
<td>4b</td>
<td>I never said Covid-19 wasn’t a real coronavirus. And deaths linked to Covid-19 are primarily caused directly from pneumonia, or flu-like symptoms. The classifications for influenza and pneumonia reporting changed when Covid-19 appeared. Facts.</td>
<td>We found that septic shock and multi organ failure was the most common immediate cause of death, often due to supplicative pulmonary infection. Respiratory failure due to diffuse alveolar damage presented as immediate cause of death in fewer cases.</td>
<td>R</td>
</tr>
</tbody>
</table>

Table 4: Examples from CoV ERT with gold evidence and fact-checking verdicts.