

# TREC 2021 Health Misinformation Track Assessing Guidelines

Version: 2.0, 23 August 2021

## Assessor Instructions

Before any judging takes place, assessors should be giving the following instructions.

Assume there is a search user who is looking to answer a medical question (e.g., "Does yoga improve the management of asthma?"). The user is searching the document collection for answers to this question. Your job is to assess documents on:

- 1) Does this document contain material that the search user might find **useful** in answering the question?
- 2) Does the document contain information that **supports/dissuades** the use of the treatment in the question?
- 3) How **credible** is the document?

## Document Collection

The document collection consists of web pages that have been converted to plaintext (text free of formatting and images). If an assessor believes that they can correctly judge the page given the plaintext, they may do so, but if in doubt, the assessor should attempt to view the actual webpage via the page's URL. If the URL is not viewable, the assessor must use just the document given its plaintext.

## Search Topics

Assessors should be provided with each search topic's description and narrative. The topic description is always presented as the question the search user wants to answer. In general, topics concern the use of some treatment for a given health issue. For example, "Does yoga improve the management of asthma?" or "Is dexamethasone a good treatment for croup?". The topics as distributed to participants contain additional information that assessors do not need to view. The <stance> and <evidence> fields are for evaluation and participants' use, and are not needed for assessing. Assessors do not need to know the topic's stance to do assessing. Assessors do **not** judge correctness of documents.

## Usefulness: Judging a document as not-useful, useful, or very-useful

Each document is judged based on the degree to which a user would find the document useful for helping make a decision about the search topic's question.

The narrative for each topic provides a specific explanation of what is considered very-useful or useful

The usefulness grades are:

0. Not-useful.

1. Useful: The user would find the document useful for the topic because it provides useful information about the health issue, the treatment, either separately or together that would help the user make a decision about the topic's question.

2. Very-useful: The user would find the document very useful because it specifically talks about the use of the treatment for the health issue, or provides strong guidance about the health treatment regardless of the health issue.

An assessor should judge a document as *very-useful* if the document specifically addresses the topic's question. For example, for the topic of yoga and asthma, a *very-useful* document would contain a direct answer to whether or not yoga improves the management of asthma, or the effects of yoga on asthma.

The document should be judged as *useful* if it provides useful information that helps the user make a decision about the topic's question. A document being about or mentioning the health issue or the treatment are not by default helpful. For example, there are likely a large number of documents about yoga, but for a document about yoga to be useful to the topic's question, it would need to be providing pertinent information that might help the user make a decision about the topic's question.

The difference between a very-useful and useful document is that the very-useful document specifically tells the user about the effectiveness of the treatment for the health issue, while a useful document will not.

**For a useful document, it does not matter whether the assessor believes the information provided in the document is incorrect or could be of harm to the searcher.** The assessor is judging whether or not a search user would be likely to find the information useful regardless of its quality.

As the Health *Misinformation* track, the problem with misinformation is that search users do not recognize it as misinformation and instead find it to be useful information. We

need to record if a user would likely find this document useful for answering the topic's question independent of the document's correctness and credibility.

Documents are automatically **Not Useful** if they:

- Are written in a language other than English,
- contain adult material, or
- are unreadable or broken.

## **Supportiveness: Judging if a document supports the treatment in the question**

For all *very-useful* and *useful* documents, the assessor should then judge if the document supports the use of the treatment in the question. If a document is not useful, assessors should not determine supportiveness, for by definition of being not useful, the document will be ignored by the user and not affect their decision.

Assessor should answer the following question: **Does the document provide information that supports/dissuades the use of the treatment in the question?**

Judgment choices are:

- **Supportive:** The document would support a decision to use the treatment.
- **Neutral:** The document neither supports or dissuades use of the treatment.
- **Dissuades:** The document would dissuade a user to not use the treatment.

## **Credibility: Judging a document's Credibility<sup>1</sup>**

Understanding the purpose of a document should be the first step to judging credibility. The assessor's opinion of the purpose of a document and the correctness of information matters in judging credibility. Credibility is only judged if a document is judged *useful* or *very-useful*.

When assessing credibility, assessors should not confuse it with *usefulness* of a document and judge credibility **independently** from usefulness.

---

<sup>1</sup> The idea of understanding the purpose of a website before judging its quality, determining the amount of expertise, authoritativeness, and trustworthiness (E-A-T), and the cdc.gov example of high E-A-T are ideas based on Google's General Guidelines for search evaluators: <http://static.googleusercontent.com/media/www.google.com/en//insidesearch/howsearchworks/assets/searchqualityevaluatorguidelines.pdf> . Last Accessed: 17/12/2018)

For a document that supports or dissuades the use of the treatment in the question, a credible document provides support for its answer and that support is deemed to be credible by the assessor.

To help determine if a document is credible,

- Use the document by itself. Assessors should not click links or research information presented in the document. The document and its source should provide the evidence for its credibility.
- It may be useful to view the url of the live webpage if it is difficult to determine the credibility from the text of the document. If the url does not work, the credibility will have to be judged based on the document's text and the source of the document.
- Try to determine the amount of expertise, authoritativeness, and trustworthiness of the document. Some websites (e.g. [www.cdc.gov](http://www.cdc.gov)) have high amounts of expertise, authoritativeness, and trustworthiness.
- Check for an author or an institute that published the web document.
- Look for author credential and qualification and determine if the person has the qualification to write this document.
- Certain documents can be written by a non-expert from a well-known newspaper/news channel website and citing the information they provide in their articles. Such documents can be considered credible.
- Certain documents may not contain author information. Check whether there exist any citations of credible sources such as universities research/clinics or government websites.
- Check where the document is published. Medical publishers or hospitals/clinics or government websites or online newspapers with wide circulation are usually credible sources.
- Check for references and sources. Credible sources often cite medical publications and/or lab studies.
- Check how well-written, well-researched and organized is the information

At the other end of the credibility spectrum are documents such as:

- The document is actually for advertising or marketing purposes. If so, the website might be biased or a scam designed to trick people into fake treatments or into buying medical products that do not live up to their claim.
- The information posted is from a personal blog or a forum, or by a non-expert person providing a medical product review or providing medical advice. Such subjective personal opinions or one point-of-view are considered not credible.

- The website provides or states claims that go against well-known medical consensus (e.g. smoking cigarettes does not cause cancer).

Credibility is to be judged on the following scale::

2: Excellent: The type of document or website that exemplifies the highest quality and most credible information source. These are the types of documents we'd want in the topmost search results. It is important that documents with *excellent* credibility distinguish themselves from merely *good* credibility. For health topics, these should be documents that are unquestionably trustworthy and authoritative.

1: Good: While not the highest quality and credibility, these pages are not low credibility.

0: Low: Lacking credibility for the topic. There is little evidence to believe or trust the information source.

## Overall Important Notes

1. The assessors are not the topic authors or the search users. Assessors are making judgments in the context of a search user who does not know the correct answer.
2. Usefulness is with respect to a search user with the question and not based on supportiveness nor credibility. For example, a search user can find non-credible documents useful, for the search users cannot always judge credibility accurately.
3. In contrast to usefulness, the assessor's expert opinion on credibility is used to judge credibility. An assessor should only assess the credibility of the document if the document is judged as useful or very-useful.