

# TREC 2022 Health Misinformation Track Assessing Guidelines

Version: 23 Aug 2022

## Overview

This year the track will judge topics in two phases. The first phase involves judging documents for their usefulness and their contained answer to a topic's question. The second phase involves preference judging the useful documents.

## Phase 1 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 documents that support either an answer of "yes" or "no" to the question.

Your job is to assess documents on:

- 1) How useful is this document for answering the question?
- 2) For documents that are useful, what answer do they support?

## Document Collection

The document collection consists of web pages that have been converted to plaintext (text free of formatting and images). Assessors are encouraged to use the plaintext to judge the page, but if in doubt, the assessor may 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 question and background. 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?". Assessors do not need to know the topic's answer to do assessing. Assessors do **not** judge correctness of documents.

## **1) 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.

**Please note that the usefulness grades have changed in 2022 and are different from 2021.**

The usefulness grades are:

0. Not-useful. A not-useful document either fails to address the question, or fails to address all parts of a question. For example, if the question is "Does yoga improve the management of asthma?" and the document only talks about yoga without talking about asthma or vice versa talks about asthma but not yoga, then the document is not-useful.
1. Useful: The user would find the document useful because it either directly answers the question or provides enough information for the user to determine an answer. Some questions ask about the effectiveness of a specific treatment for a health issue, and merely mentioning the health issue or treatment of the question is not-useful. **To be useful, a document must address all of the parts of a question and help the user make a yes/no decision for the question.**
2. Very-useful: In addition to helping the user make a decision about the question's answer, the document is high quality either because of the detail with which the question is addressed and/or the document appears to be from a highly credible source. **This document is something that you think deserves to be in the top 10 results of a web search for this topic's question.** While both *useful* and *very-useful* documents address the entire question and help the user determine an answer to the question, *very-useful* documents are of "top 10 web search" quality because of their answer quality and/or apparent credibility of their source. You can find more than 10 *very-useful* documents for a given topic, i.e. you are trying to determine *candidates* for being in the top 10 results. In the preference judging phase, your preference judgments will determine the top-10 ranking.

**For a useful or very-useful document, it does not matter whether the assessor believes the information provided in the document is correct or incorrect.** The assessor is judging whether or not a search user would be likely to find the information useful regardless of the document's correctness. For example, two very-useful documents could have different answers to the same question, but both would be viewed by the average user to be high quality results from credible sources suitable as top 10 web search results.

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

- Are written in a language other than English. A multilingual document (e.g., Spanish/English) should be judged on the basis of the information in English.
- Contain adult material.
- Are garbled, empty, unreadable or otherwise broken.

## **2) Document's Answer: Judging what a useful or very-useful document says is the answer to the question.**

For all *very-useful* and *useful* documents, the assessor should then judge what the document says is the answer to the question. All questions are written as "yes/no" questions, i.e., the answer to the question should be yes or no.

Judgment choices for "document's answer" are:

- **Yes:** The document says the answer to the question is "yes" or provides strong support that would lead a user to conclude that the answer is "yes".
- **No:** The document says the answer to the question is "no" or provides strong support that would lead a user to conclude that the answer is "no".
- **Unclear:** The document addresses the question, but a reasonable user would not be able to conclude the answer was "yes" or "no" given the document.

## **Phase 2 Assessor Instructions (Preference Judging)**

In Phase 1, you identified very-useful (top-10 web search quality) and useful documents for each topic's question. For these very-useful and useful documents, you also identified the document's answer to the question (yes, no, or unclear). In phase 2, we will tell you what the correct answer is to each topic question, and you will use a preference judging system to find the best 10 documents for a topic from the documents that contain the correct answer, or in some cases from also the documents that have an unclear answer.

Preference judging systems work by asking you to compare pairs of documents and select the preferred document.

When comparing two documents, you are to **prefer the document that would best help the searcher reach a correct decision**. If you believe both documents to be the same or near duplicates of each other, you may say that the documents are "equal". When making

preference judgments, you are encouraged to take into consideration all factors that you think would matter to a searcher and **influence the searcher to make a correct decision**.

In addition to containing a correct answer, factors may include, but are not limited to the following:

- Quality of explanation for the answer, i.e. searchers may make better decisions when a document has a correct answer with an explanation and reasoning as opposed to simply having the correct answer.
- Presentation quality. Is the answer and document written in a manner that is easy to read and comprehend?
- Some documents will have more expertise, authoritativeness, and trustworthiness<sup>1</sup>. For example, [www.cdc.gov](http://www.cdc.gov) has high amounts of expertise, authoritativeness, and trustworthiness. If two documents seem to contain the information, but one has more credibility, you would assume that the more credible document would influence the searcher more and be preferred.
- An informative document from a credible source would be preferred to a document that is for advertising or marketing purposes.
- Documents written by experts would be preferred to those by non-experts.
- The whole document context should be considered. For example, a single correct sentence embedded in a document filled with scam treatments is less likely to influence a searcher to make a correct answer than a document filled with credible information.

If you come across a document that contains an incorrect answer, this means that in Phase 1 the document was mistakenly judged with the incorrect answer, for in Phase 2 we will only show you documents judged to contain the correct answer or possibly an "unclear" answer. If you find an incorrect document, please prefer the other document with its correct answer.

The preference judging system that will be used for TREC Health Misinformation will look similar to (see next page):

---

<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)

The screenshot shows the Noddy interface with two search results displayed. On the left is a result from "everydayhealth.com" about selenium and esophageal cancer. On the right is a result from "plain.no" about research by Plein. A "Topic Information" button is visible at the top right of the interface.

**Left Result (everydayhealth.com):**

- Title: Does Selenium Prevent Esophageal Cancer? | Everyday Health
- Document ID: en.noclean.c4-train.01158-of-07168.106659
- Content includes links for print/logo, search, log in, following topics, saved items, newsletters, tools, daily health news, my daily RA, my daily diabetes, settings, logout, and subscribe menu.
- Section "Conditions" lists: Atrial Fibrillation, Cold and Flu, Depression, Heart Failure, High Cholesterol, Multiple Sclerosis, Psoriasis, Psoriatic Arthritis, Rheumatoid Arthritis, Type 2 Diabetes, Ulcerative Colitis, View All, Drugs A-Z, and Symptom Checker.
- Section "Healthy Living" is present.

**Right Result (plain.no):**

- Title: Research - Plein
- Document ID: en.noclean.c4-train.01563-of-07168.66097
- Content includes links for print/logo, search, sign in, cart, and a note about being skeptical.
- Section "Filter by ingredient" lists: All, Bone Support, Caffeine, Immune System, Caffeine, and More Info.
- Section "Caffeine" lists: Alertness, Caffeine, and Cognitive functions.
- Section "Found in:" lists: Leaves and fruits of certain plants.
- Section "Do you need it?" contains a note about caffeine's effects.
- Section "Consequences of not getting enough" notes a zombie state.

Clicking on the "Topic Information" button will bring up the topic's question and background information.

The screenshot shows the Noddy interface with a modal dialog box centered over the main content area. The dialog contains the following text:

**Does selenium help prevent cancer?**  
(Answer is unhelpful)

Selenium supplements are claimed to help prevent cancer. A very useful document would discuss the effectiveness of selenium to prevent cancer. A useful document would help a user decide about use of selenium as a cancer preventative by providing information on recommended ways to prevent cancer, the medical use of selenium, or both.

At the bottom of the dialog is an "OK" button.

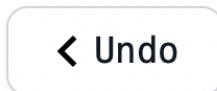
The background of the interface shows the same two search results as the first screenshot, with the "Topic Information" button still visible at the top right.

You record your preference judgment using the preference widget:

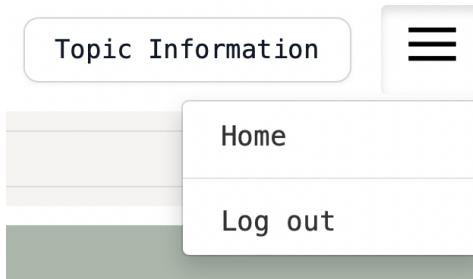
Preference: **Left**    Equal    Right

Thus, if the left document is more likely to influence a searcher to make a correct decision, you would click on "Left" and similarly if the right document is better. If the documents are the same or near-duplicates with the same source, etc. then you should judge them "Equal".

If after making a judgment, you decide it was a mistake, you can go back to the previous judgment pair using the "Undo" button:

A button with a left arrow icon and the word "Undo" next to it.

To go back to the topic selection page, you can click on the three horizontal lines in the upper right corner and select "Home". You can also log out whenever you want.



You can increase the size of documents by the following feature in the middle of the header. It will be kept during the judgment process.



The progress number is beside the topic information button in the right corner, indicating approximately how percent of judgments are done until all documents are entirely judged. It's not very accurate. More like a wild guess.

A progress bar showing "Progress: 12.5%" in a light grey box. To its right is the "Topic Information" button and the three-line menu icon.

A fast way to find occurrences of a single keyword in the documents is to use the web browser's "find in page" search feature, which is brought up by typing CTRL-F. For example in Google Chrome's browser it will pop up a widget that allows you to enter a keyword and then use up and down arrow buttons to find the next or previous occurrence. The search will first go through the left document, and then move on to the right document. For example:

The screenshot shows a web-based judging system with two document panels. The left panel is from [everydayhealth.com](https://www.everydayhealth.com/esophageal-cancer/does-selenium-prevent-esophageal-cancer.aspx) and the right panel is from [plain.no](https://plein.no/pages/research). Both panels have a search bar at the top with the word 'cancer'. The left panel highlights the word 'cancer' in several sentences. The right panel also highlights 'cancer' and includes additional context about CoQ10 and its role in the immune system.

The judging system also offers the means to enter search keywords and phrases to highlight all occurrences automatically in documents. Keywords can only contain numbers and letters, and they are not case sensitive. In the location that says "Search keywords", you can type a keyword or phrase and then press the "Enter" key to add that as a word to highlight:

The screenshot shows the same judging system interface with multiple search terms entered in the search bar: 'cancer', 'health', 'calcium', 'clock', and 'plants'. The left panel from everydayhealth.com highlights 'cancer' and 'health'. The right panel from plain.no highlights 'cancer', 'health', 'clock', and 'plants'. This demonstrates how the system highlights multiple search terms simultaneously.

Besides highlighting keywords through the search box, in the judging system, you can highlight sentences and paragraphs in the documents by mouse down, drag, and mouse up. The system keeps highlighted part of the documents until the end of the judgment session. You also can remove the highlighted part with the mouse by clicking and dragging over a highlighted section.

The highlighting by the mouse has more priority than the search keywords. For example, in the following picture, “selenium” was entered in the search box, but when the user highlighted the first sentence in the left document, it turned yellow. If the user deletes highlighting, it will reveal any search keywords.

The judging system may present a document to you multiple times, asking you to compare it to different documents each time. Any highlighting you add to a document will be retained by the system for the next time you see it. You can use this highlighting to indicate the parts of a document that were particularly useful for making your decision, potentially speeding things up if you are shown the document again.

Preference: Left Equal Right

A

Progress: 0.13% Topic Information

selenium x health x sulphur x

<https://link.springer.com/article/10.1007/s13167-010-0033-2>

Effects of **selenium** with special emphasis on major human tumours. The unique chemical properties along with metabolism and preventive mechanisms are also discussed.

Keywords

**General cancer prevention Carcinogenesis Redox Selenium Selenium metabolism**

**Selenium**-chemical properties

Selenium was discovered by Berzelius in 1818 [1]. Already in the original publication the striking similarity to sulphur and tellurium was described. However, there are a few important differences between selenium and sulphur explaining the major biological effects that are observed in biomolecules where selenium is in place of sulphur.

<http://www.whfoods.com/genpage.php?dbid=95&tname=nutrient>

Title: **selenium**  
Document ID: en.noclean.c4-train.05449-of-07168.132747

[The World's **Healthiest Foods**] April 22-28, 2019

The World's **Healthiest Foods** are health-promoting foods that can change your life.

Try our WHFoods Meal Plan.

The George Mateljan Foundation is a not-for-profit foundation with no commercial interests or advertising. Our mission is to help you eat and cook the healthiest way for optimal health.

**selenium**

To log into the system, you will be given a username and password by NIST. (You can't create a new account.)



Username\*

Password\*

Remember Me

**Sign In**

Click [here](#) to make a new account.

You will be assigned to one or more topics, which will appear in a dropdown when you log in. You can start with any topic. If you log out or switch topics, and work that you've done will be retained.



✓ Does selenium help prevent cancer? (Answer is unhelpful)

Can evening primrose oil help treat eczema? (Answer is unhelpful)

Start Judgment!