

Annotation Guidelines

Study title: Qualitative Evaluation of Automated Biomedical Text Summarization

Explanation for the Dimensions of Qualitative Evaluation

You will be asked to rate the summaries along three dimensions mentioned below. First two dimensions — Coherence and Relevance, are to be rated on a 5 point Likert Scale (higher better). The third dimension — Factual Consistency (Faithfulness), is on a binary scale {Consistent, Not Consistent}. To help you better understand the Likert rating scale in our context, we show a negative and a positive example. We also provide justification for our Likert score.

1. Coherence- The summary should be well-structured and well-organized. The summary should not just be a heap of related information, but should build from sentence to sentence to a coherent body of information about a topic.

| | |
|-------------------------|---|
| Query | What is the best way to manage phantom limb pain? |
| Source Document | New studies of the treatment of neuropathic pain have increased the need for an updated review of randomized, double-blind, placebo-controlled trials to support an evidence based algorithm to treat neuropathic pain conditions. Available studies were identified using a MEDLINE and EMBASE search. One hundred and five studies were included. Numbers needed to treat (NNT) and numbers needed to harm (NNH) were used to compare efficacy and safety of the treatments in different neuropathic pain syndromes. The quality of each trial was assessed. Tricyclic antidepressants and the anticonvulsants gabapentin and pregabalin were the most frequently studied drug classes. In peripheral neuropathic pain, the lowest NNT was for tricyclic antidepressants, followed by opioids and the anticonvulsants gabapentin and pregabalin. For central neuropathic pain there is limited data. NNT and NNH are currently the best way to assess relative efficacy and safety, but the need for dichotomous data, which may have to be estimated retrospectively for old trials, and the methodological complexity of pooling data from small cross-over and large parallel group trials, remain as limitations. |
| Negative Example | |
| Summary | a meta-analysis of 5 randomized controlled trials (nnt) found that tricyclic antidepressants and the anticonvulsants were the most frequently studied drug classes . the quality of each trial was assessed . the quality |
| Coherence | 2 |
| Explanation | The third sentence of the summary is incomplete. The second sentence is not well-formed and is weakly linked to the first sentence. |
| Positive Example | |
| Summary | Anticonvulsants , including gabapentin , have documented benefits in neuropathic pain modalities and are often used for phantom limb pain . However , their value in reducing phantom limb pain is still under investigation . |
| Coherence | 5 |
| Explanation | Unlike above, here we see that both the sentences in the summary is well formed and are well-organized and focused on the same topic of phantom limb pain which was not the case with negative example. |

2. Relevance - This dimension evaluates how well the summary captures important content from the source. The summary should include only important information from the source document. In case of a query, you must also judge how relevant is the summary to the query based on the given source document as the context.

| | |
|-------------------------|---|
| Source Document | KNEE OSTEOARTHRITIS. Good morning about 20 years ago I suffered ruptured anterior cruciate ligament and removal of domestic law meniscus, was operated and made me clancy, at present unfortunately my knee is totally affected and I have arthritis and severe pain, according to a dr traumatologo commented me I need a knee prosthesis my question is can you treat me, or turn me can recommend doctors or hospitals to treat in the U.S. |
| Negative Example | |
| Summary | What are the treatments for ruptured anterior cruciate ligament and meniscus? |
| Relevance | 2 |
| Explanation | In the source document the patient is asking for recommendation on hospital or doctors who specializes in the treatment for the topic knee osteoarthritis |
| Positive Example | |
| Summary | How can I find physician(s) or hospital(s) who specialize in knee osteoarthritis? |
| Relevance | 5 |
| Explanation | Unlike above, here we can clearly see the summary is focused on knee osteoarthritis and asking for how to find physicians or hospitals who specialize in it. |

The third dimension, **Factual Consistency**, is measured on a binary scale- **Consistent or Not Consistent**, which we explain below.

3. Factual Consistency- This dimension (also defined as faithfulness) evaluates the factual alignment between the summary and the source document (and query whenever present). A factually consistent summary contains only statements that are entailed by the source document. You may also penalize summaries that contain facts not supported (or can not be verified) in the source document (termed as *hallucinated facts*).

| | |
|------------------------|---|
| Query | Can Panitumumab cause Trichomegaly? |
| Source Document | Xerosis was present in two cases, and paronychia, pyogenic granuloma, trichomegaly, and madarosis were observed in one patient each. Eyelash trichomegaly is an uncommon drug-associated sequelae experienced during treatment with epidermal growth factor receptor (EGFR) inhibitors. Elongation of the eyelashes induced by these agents has predominantly been observed in oncology patients with either colorectal or lung cancer. It is most frequently associated with cetuximab and erlotinib; however, it has also been described in individuals treated with gefitinib or panitumumab. Trichomegaly of the eyelashes during therapy with epidermal growth factor receptor inhibitors: report of 3 cases. Trichomegaly of the eyelashes is a rare adverse effect of EGFR inhibitor therapy and is characterized by a paradoxical overgrowth of eyelashes. |

| Negative Example | |
|------------------|---|
| Summary | No . it is most frequently associated with cetuximab |
| Consistency | Not Consistent |
| Explanation | As mentioned in the source document (highlighted in blue and bolded) panitumumab can cause trichomegaly. But the summary says no and presents only half of what is the whole truth. |
| Positive Example | |
| Summary | yes , trichomegaly is an uncommon drug-associated sequelae observed during treatment with epidermal growth factor receptor (egfr) inhibitors like cetuximab, erlotinib, gefitinib, and panitumumab. |
| Consistency | Consistent |
| Explanation | Unlike above, here we see that summary is not contradicting the source document and also captures the entire truth as discussed in the source. |