

Cochrane-auto: An Aligned Dataset for the Simplification of Biomedical Abstracts

Jan Bakker, Jaap Kamps
University of Amsterdam

Background

Training neural models to simplify biomedical documents requires large, high-quality training datasets.

However, biomedical text simplification corpora are often limited in size.

To this end, Devaraj et al. (2021) introduced the Cochrane corpus, derived from the Cochrane Database of Systematic Reviews.

The CDSR

The Cochrane Database of Systematic Reviews (CDSR) comprises systematic reviews of research in health care and health policy.

A systematic review attempts to identify, appraise and synthesize all empirical evidence that is relevant to a specific research question.

Cochrane reviews are internationally recognized as the highest standard in evidence-based health care.

They are written according to a comprehensive set of guidelines.

The CDSR

Each review includes a technical abstract, which is targeted at healthcare decision makers, and a plain language summary, which should be understandable for a wide range of non-expert readers.

Abstract

Available in [English](#) | [Español](#)

Background

Fracture of the distal radius ('broken wrist') is a common clinical problem. It can be treated conservatively, usually involving wrist immobilisation in a plaster cast, or surgically. A key method of surgical fixation is external fixation.

Objectives

To evaluate the evidence from randomised controlled trials comparing external fixation with conservative treatment for fractures of the distal radius in adults.

Search methods

We searched the Cochrane Bone, Joint and Muscle Trauma Group Specialised Register (September 2006), the Cochrane Central Register of Controlled Trials, MEDLINE, EMBASE and other databases, conference proceedings and reference lists of articles. No language restrictions were applied.

Selection criteria

Randomised or quasi-randomised controlled clinical trials involving adults with a fracture of the distal radius, which compared external fixation with conservative treatment.

Plain language summary

Available in [English](#) | [Español](#)

External fixation versus conservative treatment for distal radial fractures in adults

In older people, a 'broken wrist' (from a fracture at the lower end of one of the two forearm bones) can result from a fall onto an outstretched hand. Treatment usually involves reduction (putting the broken bone back into position) and immobilising the wrist in a plaster cast (conservative treatment). Surgery may be considered for more seriously displaced fractures. One type of surgery is external fixation, in which metal pins are driven into bone, generally via small skin incisions, on either side of the fracture. These pins are then fixed externally by incorporation into a plaster cast or securing into the frame of an external fixator. The external component holds the bony fragments in position while the bone heals. This review looked at the evidence from randomised controlled trials comparing external fixation with conservative treatment.

Fifteen trials, involving 1022 adults with potentially or evidently unstable fractures, were included. While all trials compared external fixation versus plaster cast immobilisation, there was considerable variation in their characteristics especially in terms of patient characteristics and the method of external fixation. Weak methodology, such as using inadequate methods of randomisation and outcome assessment, means that the possibility of serious bias can not be excluded.

The review found that external fixation reduced fracture redisplacement that prompted further treatment and generally improved final anatomical outcome. It appears to improve function too but this needs to be confirmed. The complications, such as a pin tract infection, associated with external fixation were many but were generally minor. Serious complications occurred in both groups. The review concludes that there is some evidence to support the use of external fixation for these fractures.

The Cochrane Corpus

Devaraj et al. (2021) compiled a corpus of (complex, simple) texts comprising parallel subsets of abstracts and lay summaries from the CDSR.

Each complex text in the corpus covers the *Main Results* and *Authors' Conclusions* sections of the technical abstract.

Each simple text comprises a lay summary, starting at the approximate location of the first section, paragraph or sentence (depending on the structure) describing the studies and results.

Analysis of the Cochrane Corpus

We randomly select ten paired (complex, simple) texts from the corpus.

Next, we manually align sentences between these pairs that are equivalent or partially equivalent in meaning.

As a result, we obtain 79 alignments.

Complex text

{0: 'We included two trials involving 54 participants with CVI.', 1: 'Many of our review outcomes were not reported or reported by only one of the two studies.', 2: 'The intensity of disease signs and symptoms was measured in both studies but using different scales; we were therefore unable to pool the data.', 3: 'One study reported no difference between the exercise and control groups whereas the second reported a reduction in symptoms in the exercise group.', 4: 'In one study, increases in change in ejection fraction compared with baseline (mean difference (MD) 4.88%, 95% confidence interval (CI) 3.16 to 6.60; 30 participants; $P < 0.00001$), half venous refilling time (MD 4.20 seconds, 95% CI 3.28 to 5.12; 23 participants; $P < 0.00001$) and total venous refilling time (MD 9.40 seconds, 95% CI 7.77 to 11.03; 23 participants; $P < 0.00001$) were observed in the exercise group compared with the control group.', 5: 'One study reported no difference between the exercise and control groups with regard to quality of life or ankle range of motion.', 6: 'Although muscle strength assessed by dynamometry at slow speed did not differ between the two groups in this study, variable peak torque at fast speed was lower in the control group than in the exercise group (2.8 ± 0.9 compared with -0.3 ± 0.6 , $P < 0.03$).', 7: 'The incidence of venous leg ulcers, incidence of surgical intervention to treat symptoms related to CVI and exercise capacity were not assessed or reported in either of the included trials.', 8: 'We rated both included studies as at high risk of bias; hence, these data should be interpreted carefully.', 9: 'Due to the small number of studies and small sample size, we were not able to verify indirectness and publication bias.', 10: 'Therefore, we judged the overall quality of evidence as very low according to the GRADE approach.', 11: 'There is currently insufficient evidence available to assess the efficacy of physical exercise in people with CVI.', 12: 'Future research into the effect of physical exercise should consider types of exercise protocols (intensity, frequency and time), sample size, blinding and homogeneity according to the severity of disease.'}

Simple text

{0: 'This review included two clinical trials, involving a total of 54 participants, that compared directly the effects of physical exercise and a control intervention (evidence current until May 2016).', 1: 'One study reported no difference between the exercise and control groups whereas the second reported a reduction in symptoms in the exercise group.', 2: 'At the end of the study, an improvement in venous blood return was observed in the exercise group compared with the control group.', 3: 'The included studies did not report on new cases of venous leg ulcers.', 4: 'No difference between the exercise and control groups was observed with regard to participants' quality of life, the range of motion of the ankle joint or overall muscle strength.', 5: 'The overall finding of an improvement in venous blood return in the exercise group favours the idea that physical exercise improves blood flow conditions in people with CVI, but we found the risk of bias due to blinding or randomisation to be high for both studies.', 6: 'We therefore consider that there is currently not enough information to determine whether physical exercise is effective in the management of CVI.', 7: 'Quality of the evidence We judged the overall quality of evidence as very low: the two included studies were small (54 participants in total) and were at high risk of bias based on their methods of blinding or randomisation.'}

Analysis of the Cochrane corpus

Of the total of 98 simple sentences in the selected texts, 68 are aligned to at least one of the 139 complex sentences.

Thus, 30 out of 98 simple sentences are not aligned.

While 2 of them are elaborations, the remaining 28 contain information that is present in the full review but not in the complex text.

This is largely because the plain language summaries are written directly from the full review, instead of being simplified versions of the technical abstracts.

Consequently, around 29% of the sentences in the simple reference texts cannot be generated from the complex source text.

Cochrane-auto

We automatically align the simple and complex sentences in the corpus, and we exclude all simple sentences that are not aligned from our references.

(We use from an updated version of the Cochrane corpus, which is based on systematic reviews that were published in the CDSR up until March 14, 2024.)

Alignment model

We make use of the neural CRF alignment model proposed by Jiang et al. (2020).

When provided with a (complex, simple) text pair as input, this model automatically aligns each sentence in the simple text to either one or zero corresponding sentences in the complex text.

In doing so, it leverages the similar order of sentences in parallel texts and utilizes a fine-tuned BERT model to capture the semantic similarity between sentence pairs.

Alignment model

Jiang et al. (2020) applied their model to two simplification corpora: Newsela and Wikipedia.

More specifically, they first created Newsela-manual and Wiki-manual by manually aligning 50 article groups from Newsela and 500 article pairs from Wikipedia.

Then they fine-tuned BERT and trained their alignment models on train splits of these datasets.

Finally, they applied their trained models to the remaining data to create the automatically aligned Newsela-auto and Wiki-auto datasets.

Alignment procedure

We create Cochrane-auto by applying the sentence alignment model that was pretrained on Wiki-manual to the updated Cochrane corpus.

We apply our alignment model to the full text pairs.

When sentences from different paragraphs in a simple text are automatically aligned to the same sentence in the parallel text, we only keep the alignments with the simple paragraph in which the aligned sentences have the highest similarity to that sentence.

Alignment results

	TP	FP	FN	F1
BERT <i>finetune</i>	52	32	27	63.8
CRF Aligner	53	6	26	76.8
+ merge	56	8	23	78.3

Table 1: Performance of sentence alignment methods on 10 annotated text pairs from the Cochrane corpus.

Alignment results

	Cochrane- auto	Newsela- auto	Wiki- auto
Domain	Biomedical	News	General
# Doc Pairs	5,585	18,820	138,095
# Sent Pairs	35,800	813,972	685,769

Table 2: Statistics for the automatically aligned Cochrane-auto, Newsela-auto and Wiki-auto datasets.

Preprocessing

We preprocess Cochrane-auto similarly to how Cripwell et al. (2023a) preprocessed Newsela-auto and Wiki-auto.

That is, for each sentence c_i in a complex document, we use the simple sentence s_j to which it is aligned as a reference. If it is aligned to multiple s_j s, we concatenate them; if it is not aligned, we use an empty string.

Next, we create paragraph- and document-level references by concatenating the references for each sentence in a complex paragraph or document.

Preprocessing

We filter out instances where less than 50% of the sentences in the complex document C are aligned to any s_j .

We also remove instances where the length of a document exceeds 1024 tokens, or would exceed 1024 tokens after adding the special tokens needed for the plan-guided simplification approach of Cripwell et al. (2023a).

As a result, the preprocessed Cochrane-auto dataset consists of 894 train, 125 validation and 121 test instances.

Dataset statistics

	Cochrane- auto	Newsela- auto	Wiki- auto
# Doc Pairs	1,085	18,319	85,123
# Para Pairs	4,171	361,964	178,982
# Sent Pairs	14,719	707,776	461,852
Avg. $ c_i $	35.61	22.49	28.64
Avg. $ s_i $	27.75	15.84	21.57
Avg. n	13.57	38.64	5.43
Avg. k	9.01	42.60	4.53
Avg. p	3.53	1.96	2.58

Table 3: Statistics of the datasets after preprocessing, where n is # sentences in C , and k is # sentences in S and p is # sentences per paragraph in C .

Copy	Rephrase	Split	Merge	Delete
8.4	45.3	4.5	6.5	35.3

Table 4: Operation class distribution for Cochrane-auto in percentages.

Example

Complex document

Two randomised trials with a total of 161 participants were included in this review. The studies did not report on mortality and rate of limb loss. One randomised trial with a total of 133 participants showed that there was a significant improvement in ankle brachial index (ABI) in participants who received folic acid compared with placebo (mean difference (MD) 0.07, 95% confidence interval (CI) 0.04 to 0.11, $P < 0.001$) and in participants who received 5-methyltetrahydrofolate (5-MTHF) versus placebo (MD 0.05, 95% CI 0.01 to 0.10, $P = 0.009$). A second trial with a total of 18 participants showed that there was no difference (P non-significant) in ABI in participants who received a multivitamin B supplement (mean \pm SEM: 0.7 ± 0.1) compared with placebo (mean \pm SEM: 0.8 ± 0.1). No major events were reported.

Currently, no recommendation can be made regarding the value of treatment of hyperhomocysteinaemia in peripheral arterial disease. Further, well constructed trials are urgently required.

Simple document

Two trials with 161 participants with PAD were included in this review. None of the other predefined primary outcomes (mortality and rate of limb loss) were assessed in these studies. One trial showed a significant improvement in the ankle brachial index (ABI) in participants treated daily with 400 μg folic acid or 5-methyltetrahydrofolate (5-MTHF). A second trial showed that there was no difference in ABI in participants who received a multivitamin B supplement compared with placebo.

Original reference

We looked at studies where treatments to lower homocysteine were used in people with PAD and hyperhomocysteinaemia. Two trials with 161 participants with PAD were included in this review. One trial showed a significant improvement in the ankle brachial index (ABI) in participants treated daily with 400 μg folic acid or 5-methyltetrahydrofolate (5-MTHF). A second trial showed that there was no difference in ABI in participants who received a multivitamin B supplement compared with placebo. None of the other predefined primary outcomes (mortality and rate of limb loss) were assessed in these studies. More research about the effect of homocysteine lowering therapy on the clinical progression of disease in people with PAD and hyperhomocysteinaemia is needed.

Limitations

- Our automatic alignments are imperfect;
- Simple sentences that are correctly aligned may still contain information that is not present in the source sentence;
- The 'real' alignments may not reside at the sentence-level;
- The deletion and reordering of sentences may impact the discourse structure of the reference document;
- Elaborative sentences are excluded from the references;
- The simplifications exhibit greater variation, which may lead to conservative models.

Conclusion

We present Cochrane-auto: a large aligned dataset for the simplification of biomedical abstracts at the document-, paragraph- and sentence-level.

Our freely available corpus complementing Newsela-auto and Wiki-auto facilitates text simplification research beyond direct lexical and grammatical revisions.

Questions?