Nature Citation Style:
Attal, K., Ondov, B. & Demner-Fushman, D. A dataset for plain language adaptation of biomedical abstracts. *Sci Data* **10**, 8 (2023). https://doi.org/10.1038/s41597-022-01920-3

DOI:
https://doi.org/10.1038/s41597-022-01920-3

BibTex:
```
@article{attal_dataset_2023,
	title = {A dataset for plain language adaptation of biomedical abstracts},
	volume = {10},
	issn = {2052-4463},
	url = {https://doi.org/10.1038/s41597-022-01920-3},
	doi = {10.1038/s41597-022-01920-3},
	abstract = {Though exponentially growing health-related literature has been made available to a broad audience online, the language of scientific articles can be difficult for the general public to understand. Therefore, adapting this expert-level language into plain language versions is necessary for the public to reliably comprehend the vast health-related literature. Deep Learning algorithms for automatic adaptation are a possible solution; however, gold standard datasets are needed for proper evaluation. Proposed datasets thus far consist of either pairs of comparable professional- and general public-facing documents or pairs of semantically similar sentences mined from such documents. This leads to a trade-off between imperfect alignments and small test sets. To address this issue, we created the Plain Language Adaptation of Biomedical Abstracts dataset. This dataset is the first manually adapted dataset that is both document- and sentence-aligned. The dataset contains 750 adapted abstracts, totaling 7643 sentence pairs. Along with describing the dataset, we benchmark automatic adaptation on the dataset with state-of-the-art Deep Learning approaches, setting baselines for future research.},
	number = {1},
	journal = {Scientific Data},
	author = {Attal, Kush and Ondov, Brian and Demner-Fushman, Dina},
	month = jan,
	year = {2023},
	pages = {8},
}
```