

**Grant Number**: N/A

**Sponsor:** IOE Seed Funding Scheme 2017, UCL Institute of Education

**Project title**: Corpora of Patient Information Sheets and Consent Forms for UK Cancer Trials, 2007-2017

The following files have been archived:

|  |  |
| --- | --- |
| File name | File description (Short description of content, sample size, format, any linking between different types of data, i.e. survey and interviews/focus groups) |
| [Information\_Sheet\_Corpus\_for\_UKDS.txt](http://reshare.ukdataservice.ac.uk/853933/2/Information_Sheet_Corpus_for_UKDS.txt) | An anonymised corpus of 27 collated patient information sheets (PIS) in a single .txt file (324 kB), drawn from an original PIS corpus consisting of 62,030 words. Research notes and referencing information are embedded in this text-based file using angled brackets. The original ethical documents, which were converted from their original file format (.doc/.docx/.pdf) to .txt, were freely obtained from the public domain via web searches using key databases for randomised controlled trials (RCTs) targeting cancer conducted in the UK. Twenty-two PIS and CF in our corpora were drawn from the same RCT. Five PIS do not have a matching CF. The text for two PIS files has been removed due to copyright restrictions, and a link to the original data source has been provided instead. |
| [Consent\_form\_corpus\_for\_UKDS.txt](http://reshare.ukdataservice.ac.uk/853933/1/Consent_form_corpus_for_UKDS.txt) | An anonymised corpus of 23 collated CF in a single .txt file (53 kB), drawn from an original CF corpus consisting of 8118 words. Research notes and referencing information are embedded in this text-based file using angled brackets. The original ethical documents, which were converted from their original file format (.doc/.docx/.pdf) to .txt, were freely obtained from the public domain via web searches using key databases for RCTs targeting cancer conducted in the UK. Twenty-two PIS and CF in our corpora were drawn from the same RCT. One CF does not have a matching PIS. The text for two CF files has been removed due to copyright restrictions, and a link to the original data source has been provided instead. |

**Publications**: (based on this data, if any)

Forthcoming