The attached files include the raw data processed to produce the Genuine Index (GI) discussed in this article. This data was obtained from Pubmed (https://www.ncbi.nlm.nih.gov/pubmed) and converted into data tables using a custom program and processing algorithm developed using RapidMiner Studio 6.5 (RapidMiner GmbH. Released 2015. RapidMiner Studio Academia, Version 6.5002).

Descriptions of the relevant, included dataset are as follows

1. File name: 20161129\_GIIndexBetaCorpus-PedsOnco\_OUT.CSV
   1. Contents: 5,907 abstracts dated between 1986 – 2015
   2. Original source:
      1. Pediatric Blood & Cancer (ISSN#: 1545-5017)
      2. Pediatric Hematology and Oncology (ISSN#: 1521-0669)
   3. Search criteria:
      1. *“Pediatr Blood Cancer”[Journal] OR “Pediatr Hematol Oncol”[Journal] AND “has abstract”[text].*
2. File name: 20161129\_GIIndexBetaCorpus-Anesthesiology\_OUT.CSV
   1. Contents: 16,952 abstracts dated between 1986 – 2015
   2. Original source:
      1. Anesthesia and Analgesia (ISSN#: 1526-7598)
      2. Anaesthesia (ISSN#: 13652044)
   3. Search criteria:
      1. *"Anaesthesia"[Journal] OR "Anesthesia and analgesia"[Journal] AND (hasabstract[text] AND ("1986/01/01"[PDAT] : "2015/12/31"[PDAT]))*

Data is provided in CSV format. The following variables are included in each data set:

1. Pubmed ID [PMID]
2. Journal ID [JournalTitle]
3. Article Title [ArticleTitle]
4. First Author Affiliation [Affiliation1]
5. Abstract Text [AbstractText]
6. Publication [Year, GYear]