replication_notebook_white_2020

August 18, 2020

1 Review and Replication Report for White et al (2020) submitted to PeerJ

[3]:
```
                                    DOI  \
0                 10.1001/jama.2017.7219
1            10.1001/jamacardio.2017.0175
2            10.1001/jamacardio.2017.2941
3        10.1001/jamapediatrics.2017.1579
4        10.1001/jamapsychiatry.2016.4234
…                                      …
12925              10.19153/cleiej.20.1.2
12926                  10.1037/xge0000268
12927            10.1177/0961463X17701955
12928       10.1080/13676261.2017.1316363
12929              10.5194/tc-11-949-2017

                                    Evidence Licence OA Status  \
0                         open (via free pdf)     NaN    bronze
1        oa repository (via OAI-PMH doi match)     NaN     green
2                                        NaN     NaN    closed
3                         open (via free pdf)     NaN    bronze
4                         open (via free pdf)     NaN    bronze
…                                         …       …         …
12925                  oa journal (via doaj)   cc-by   diamond
12926       oa repository (via pmcid lookup)     NaN     green
12927                                    NaN     NaN    closed
12928                                    NaN     NaN    closed
12929            open (via page says license)   cc-by      gold

                                                  Title  \
0        Effect of Robotic-Assisted vs Conventional Lap…
1        Effect of Monthly High-Dose Vitamin D Suppleme…
2        Vitamin D Supplementation and Cardiovascular D…
3        Association of Neonatal Glycemia With Neurodev…
4        Paternal Depression Symptoms During Pregnancy …
…                                                      …
12925    Comparison of Two Forced Alignment Systems for…
```

```
12926  Once a frog-lover, always a frog-lover?: Infan…
12927  The tyranny of clock time? Debating fatigue in…
12928  Youth studies, citizenship and transitions: to…
12929  How accurate are estimates of glacier ice thic…

                                                Authors  Author count  \
0      Jayne, David; Pigazzi, Alessio; Marshall, Hele…          16.0
1      Scragg, Robert; Stewart, Alistair W.; Waayer, …           9.0
2                    Scragg, Robert; Camargo, Carlos A.           2.0
3      Mckinlay, Christopher J. D.; Alsweiler, Jane M…          15.0
4      Underwood, Lisa; Waldie, Karen E.; Peterson, E…           7.0
…                                                    …           …
12925  Flores Sol\\U00F3Rzano, Sof\\U00Eda; Coto-Sola…           2.0
12926  Martin, Alia; Shelton, Catharyn C.; Sommervill…           3.0
12927                              Snyder, Benjamin H            1.0
12928                          Wood, Bronwyn Elisabeth           1.0
12929  Farinotti, Daniel; Brinkerhoff, Douglas J.; Cl…          37.0

      Author count>20                                    Journal  Year  …  \
0                  No                                       JAMA  2017  …
1                  No                          JAMA Cardiology  2017  …
2                  No                          JAMA Cardiology  2017  …
3                  No                          JAMA Pediatrics  2017  …
4                  No                          JAMA Psychiatry  2017  …
…                 …                                          …   …  …
12925              No                    CLEI electronic journal  2017  …
12926              No  Journal of Experimental Psychology: General  2017  …
12927              No                           Time & Society  2019  …
12928              No                 Journal of Youth Studies  2017  …
12929             Yes                           The Cryosphere  2017  …

      APC charged in DOAJ  DOAJ Currency DOAJ APC  Flourish APC (USD)  \
0                     NaN            NaN      NaN                 NaN
1                     NaN            NaN      NaN                 NaN
2                     NaN            NaN      NaN                 NaN
3                     NaN            NaN      NaN                 NaN
4                     NaN            NaN      NaN                 NaN
…                    …             …       …                  …
12925                  No            NaN      NaN                 0.0
12926                 NaN            NaN      NaN                 NaN
12927                 NaN            NaN      NaN                 NaN
12928                 NaN            NaN      NaN                 NaN
12929                 Yes            EUR   1000.0                 0.0

      Publisher Currency Publisher APC USD APC  \
0                    NaN           NaN     NaN
1                    NaN           NaN     NaN
```

```
2                    NaN           NaN     NaN
3                    NaN           NaN     NaN
4                    NaN           NaN     NaN
…                    …             …       …
12925                NaN           NaN     0.0
12926                NaN           NaN     NaN
12927                USD        3000.0  3000.0
12928                USD        2950.0  2950.0
12929                NaN           NaN  1132.0

                                                 Funders  \
0      Chief Scientist Office in Scotland; National I…
1      Health Research Council of New Zealand [10/400…
2      Health Research Council of New Zealand; Accide…
3      Eunice Kennedy Shriver National Institute of C…
4      New Zealand Ministries of Social Development H…
…                                                    …
12925                                              NaN
12926                                            NICHD
12927  National Science Foundation Doctoral Dissertat…
12928  New Zealand Ministry of Education Teaching and…
12929  Swiss National Science Foundation (SNSF); Nati…

                              Subjects  NZ Reprint author
0            General & Internal Medicine                No
1      Cardiovascular System & Cardiology               Yes
2      Cardiovascular System & Cardiology               Yes
3                           Pediatrics                  Yes
4                           Psychiatry                  Yes
…                                  …                   …
12925                             NaN                   No
12926                       Psychology                  Yes
12927     Social Sciences - Other Topics                Yes
12928     Social Sciences - Other Topics                Yes
12929        Physical Geography; Geology                 No

[12930 rows x 37 columns]
```

1.1 Some basic analysis and confirmation of the results from the spreadsheet

There appear to be 12,930 rows vs 12,016 articles mentioned. It's not clear to me how you would get to 12,016 from the data here. The DOIs appear to be unique.

[4]: 12226

Confirming the approximate levels of OA. Paper says 59% closed which is close to what I get here and the differences almost certainly lie in issue with the count of articles. Broadly speaking this confirms the approximate percentages given in the article and Table 5

[5]:

OA Status	counts	percent
bronze	1102	9.013578
closed	7253	59.324391
diamond	265	2.167512
gold	1704	13.937510
green	1269	10.379519
hybrid	633	5.177491

[6]: 12226

Confirm the figures as a percentage of the open articles

[7]:

OA Status	counts	percent	percent_of_open
bronze	1102	9.013578	22.159662
closed	7253	59.324391	145.847577
diamond	265	2.167512	5.328775
gold	1704	13.937510	34.265031
green	1269	10.379519	25.517796
hybrid	633	5.177491	12.728735

Repeat the analysis for those cases where there is an NZ corresponding author

[8]:

OA Status	counts	percent	percent_of_open
bronze	421	7.855943	23.337029
closed	3555	66.337003	197.062084
diamond	95	1.772719	5.266075
gold	696	12.987498	38.580931
green	438	8.173167	24.279379
hybrid	154	2.873670	8.536585

Average citations by OA type I seem to have a quite significant disagreement with the article. Am I missing something about the way this is calculated? Presumably 'average' is the mean?

[9]:

OA Status	counts	av_cites
bronze	1102	5.183303
closed	7253	4.443816
diamond	265	1.788679
gold	1704	5.142019
green	1269	6.921986
hybrid	633	7.917852

Repeat for NZ reprint Authors

[10]:

OA Status	counts	av_cites
bronze	421	4.976247
closed	3555	3.650070
diamond	95	1.452632
gold	696	4.744253
green	438	5.082192
hybrid	154	6.785714

1.2 Calculating APCs

The paper only reports on APCs for those articles with a NZ author. It might be interesting to look at various implementations of the CAUL approach for APC calculation to see how much difference that makes.

There are minor variations here that presumably relate to the same slight issues with numbers as above.

[11]:

OA Status	counts	av_cites	known_apcs	total_apcs	av_apcs
gold	696	4.744253	696	1171529.0	1683.231322
hybrid	154	6.785714	110	281378.0	2557.981818

1.3 Embargo Periods

[12]:

Archive accepted manuscript	counts	percent
12 months embargo	2116	60.113636
18 months embargo	318	9.034091
2 years embargo	171	4.857955
24 months embargo	41	1.164773
3 months embargo	3	0.085227
36 months embargo	1	0.028409
6 months embargo	73	2.073864
No information	143	4.062500
Permission must be obtained from the publisher	1	0.028409
Upon funder agreement with publisher	2	0.056818
can	580	16.477273
cannot	43	1.221591
unclear	5	0.142045
unknown	23	0.653409

1.4 Looking at Funders

First split out the table and determine the funder names and then create some new True/False columns for each funder.

It is not clear how the funders in Table 8 were filtered out. It seems like they might have focussed

only on precise string matching rather than looking for potential duplicates? There are some differences between what appear to be duplicated funders under different names and this should be investigated more fully. There are also some funders missing from Table 8 (or is the Marsden fund collected up under the Royal Society of New Zealand?).

There seems to be a trend for medical funders to have higher levels of OA than for others? It might be of value to examine this across the broader dataset as well. It might be interesting to look at a comparison between the more international articles that are excluded in general from this analysis to compare OA levels.

[15]:

	funder	count
4139	University of Otago	415
1335	Health Research Council	376
1912	Marsden Fund	304
1360	Health Research Council of New Zealand	292
4009	University of Auckland	291
3384	Royal Society of New Zealand	276
2171	Ministry of Business Innovation and Employment	231
4302	Victoria University	134
4306	Victoria University of Wellington	121
2019	Massey University	106
2879	New Zealand Ministry of Business Innovation an…	89
1829	MBIE	79
3500	Rutherford Discovery Fellowship	68
3398	Royal Society of New Zealand Marsden	67
2852	New Zealand Marsden Fund	67
1306	HRC	67
4093	University of Canterbury	58
203	Auckland Medical Research Foundation	58
3401	Royal Society of New Zealand Marsden Fund	57
2258	Ministry of Health	53
1491	Heart Foundation	51
3133	Otago Research Grant	48
1861	MacDiarmid Institute	47
4195	University of Otago Research Grant	46
1799	Lottery Health	45
2101	Medical Research Council	45
227	Auckland University	44
281	Australian Research Council	43
2787	New Zealand Government	43
4252	University of Waikato	43

```
/Users/266883j/anaconda3/lib/python3.7/site-packages/ipykernel_launcher.py:2:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
```

docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

[18]:

	funder	bronze	closed
0	0	0.0	0.0
1	University of Otago	42.0	231.0
2	Health Research Council	47.0	173.0
3	Marsden Fund	31.0	169.0
4	Health Research Council of New Zealand	38.0	136.0
5	University of Auckland	29.0	190.0
6	Royal Society of New Zealand	26.0	154.0
7	Ministry of Business Innovation and Employment	21.0	155.0
8	Victoria University	10.0	88.0
9	Victoria University of Wellington	10.0	78.0
10	Massey University	5.0	65.0
11	New Zealand Ministry of Business Innovation an…	8.0	58.0
12	MBIE	6.0	50.0
13	Rutherford Discovery Fellowship	8.0	35.0
14	Royal Society of New Zealand Marsden	8.0	35.0
15	New Zealand Marsden Fund	8.0	35.0
16	HRC	4.0	30.0
17	University of Canterbury	6.0	34.0
18	Auckland Medical Research Foundation	11.0	28.0
19	Royal Society of New Zealand Marsden Fund	8.0	28.0
20	Ministry of Health	6.0	25.0
21	Heart Foundation	10.0	26.0
22	Otago Research Grant	6.0	25.0
23	MacDiarmid Institute	9.0	27.0
24	University of Otago Research Grant	4.0	25.0
25	Lottery Health	6.0	23.0
26	Medical Research Council	6.0	15.0
27	Auckland University	1.0	26.0
28	Australian Research Council	2.0	22.0
29	New Zealand Government	3.0	34.0
30	University of Waikato	4.0	23.0

	diamond	gold	green	hybrid	total	pc_closed	pc_open
0	0.0	0.0	0.0	0.0	0.0	NaN	NaN
1	8.0	89.0	25.0	20.0	415.0	55.662651	44.337349
2	3.0	113.0	30.0	10.0	376.0	46.010638	53.989362
3	4.0	41.0	47.0	12.0	304.0	55.592105	44.407895
4	2.0	85.0	23.0	8.0	292.0	46.575342	53.424658
5	2.0	47.0	13.0	10.0	291.0	65.292096	34.707904
6	2.0	50.0	31.0	13.0	276.0	55.797101	44.202899
7	2.0	34.0	12.0	7.0	231.0	67.099567	32.900433
8	2.0	14.0	14.0	6.0	134.0	65.671642	34.328358
9	2.0	12.0	13.0	6.0	121.0	64.462810	35.537190

```
10      1.0   27.0    4.0     4.0  106.0  61.320755  38.679245
11      1.0   12.0    7.0     3.0   89.0  65.168539  34.831461
12      0.0   16.0    6.0     1.0   79.0  63.291139  36.708861
13      0.0    9.0    8.0     8.0   68.0  51.470588  48.529412
14      1.0   11.0    6.0     6.0   67.0  52.238806  47.761194
15      1.0    9.0    8.0     6.0   67.0  52.238806  47.761194
16      1.0   25.0    5.0     2.0   67.0  44.776119  55.223881
17      1.0   10.0    6.0     1.0   58.0  58.620690  41.379310
18      1.0   14.0    3.0     1.0   58.0  48.275862  51.724138
19      1.0    9.0    6.0     5.0   57.0  49.122807  50.877193
20      1.0   15.0    3.0     3.0   53.0  47.169811  52.830189
21      0.0    6.0    6.0     3.0   51.0  50.980392  49.019608
22      1.0   11.0    5.0     0.0   48.0  52.083333  47.916667
23      0.0    7.0    3.0     1.0   47.0  57.446809  42.553191
24      1.0   11.0    5.0     0.0   46.0  54.347826  45.652174
25      0.0   10.0    4.0     2.0   45.0  51.111111  48.888889
26      1.0   12.0    8.0     3.0   45.0  33.333333  66.666667
27      0.0    8.0    7.0     2.0   44.0  59.090909  40.909091
28      0.0    9.0    8.0     2.0   43.0  51.162791  48.837209
29      0.0    4.0    2.0     0.0   43.0  79.069767  20.930233
30      0.0    7.0    8.0     1.0   43.0  53.488372  46.511628
```

1.5 Quick validation against COKI dataset

For a further validation I pull data from an internal dataset for a quick comparison of OA and classes of OA for 2017 publications from the relevant universities. Query based on work by Rebecca Handcock.

```
Downloading: 100%|      | 9292/9292 [00:01<00:00, 5308.57rows/s]
```

[21]:

```
                                  doi  year  is_oa gold_doaj  green  \
0            10.1186/s40638-017-0061-7  2017   True     False   True
1            10.1007/s00234-017-1816-0  2017  False     False  False
2         10.1016/j.ecoser.2016.10.013  2017  False     False  False
3            10.1016/j.joi.2017.10.004  2017   True     False   True
4     10.1016/j.jvolgeores.2017.01.009  2017  False     False  False
…                                  …     …      …         …      …
9287        10.1109/m2vip.2017.8211452  2017  False     False  False
9288        10.1016/j.xphs.2016.10.017  2017  False     False  False
9289         10.1109/peds.2017.8289250  2017  False     False  False
9290              10.2196/resprot.8522  2017   True      True   True
9291     10.1016/j.jempfin.2017.04.001  2017   True     False   True

     green_only hybrid bronze  citations
0         False   True  False          4
1         False  False  False         19
2         False  False  False          5
3          True  False  False          7
```

```
4         False  False  False          1
…         …    …    …        …
9287      False  False  False          0
9288      False  False  False          8
9289      False  False  False          0
9290      False  False  False          0
9291       True  False  False         11

[9292 rows x 9 columns]
```

We get fewer overall articles which is not suprising as the local data collection should be better overall. Basic check on levels of different categories and citation counts. Note that for this analysis the terms 'gold_doaj' corresponds to the use of gold in the paper, and 'green_only' corresponds to the use of green in the paper.

```
is_oa 3522 37.90357296599225 7.424190800681431
gold_doaj 1178 12.67757210503659 7.077249575551782
hybrid 485 5.2195436934997845 10.245360824742267
bronze 815 8.770985794231597 6.411042944785276
green 2772 29.832113646147224 8.293650793650794
green_only 1044 11.235471373224279 7.295977011494253
closed 5770 62.09642703400775 4.151299826689774
```