



# Words don't come easy: analysis of words in titles of the publications with the highest Altmetric Attention Score



Kalmer Lauk  
Bibliometrics Specialist  
University of Tartu

## Introduction

No matter how a user searches for relevant scientific information, title is the most important element in any scientific form of publication. Therefore, the title and its quality can have an affect on the impact of the publication. Titles should clearly indicate the subject and arouse interest and an ideal title should be rather short, informative and attractive (Jamali, Nikazad, 2011).

Non-traditional altmetrics has risen next to the traditional citation based metrics as a way to analyze the online impact of published research. But highly attractive or amusing titles seem to gather more attention.

From a researchers point of view they are writing important articles which they want to be read and shared and cited as widely and as much as possible. And not just for pride or glory, but since research funding is largely based on the impact of the research previously published, marketing titles correctly can mean a great deal. This situation, however, can lead to clickbaiting. (Lockwood, 2016)

The synthesized analysis of words in titles was developed in recent years, and has proved to be useful information in revealing the research focuses (Ho et al., 2010) or even how the field has changed through the years (Lauk, 2016).

## Objectives and methods

Following analysis was done to see if publications with high Altmetric Attention Score (Altmetric Score) (AAS) tend to use words that can be related to academic clickbait or buzzwords and do these publications also have academic credentials in the form of citations in Web of Science.

For this analysis the list of the top 1000 publications with the highest AAS was received from altmetric.com on 7<sup>th</sup> June 2016. Since the top 1000 publications with the highest AAS are mostly either articles or news items these two publication types will be looked at more closely. The citation data from Web of Science was received during July 2016. The word analysis was made on a freeware word density analyzer and stop words were not included.

Table 1. 50 most frequently used words in the top 1000 publications with the highest Attention Score.		Table 2. 50 most frequently used words in the top 1000 most cited publications in Web of Science.		Table 3. 50 most frequently used words in the news titles with the highest Attention Score		Table 4. 50 most frequently used words in the most cited news items in Web of Science		Table 5. 50 most frequently used words in the article titles with the highest Attention Score.		Table 6. 50 most frequently used words in the most cited articles in Web of Science.	
Word	#	Word	#	Word	#	Word	#	Word	#	Word	#
study	49	analysis	54	science	17	gene	16	study	44	analysis	37
human	47	protein	46	human	12	cancer	12	risk	43	method	35
risk	43	cancer	45	scientists	11	therapy	12	global	35	molecular	27
global	40	human	41	research	8	materials	11	disease	35	data	25
disease	37	method	41	physics	8	report	10	human	35	cancer	22
mortality	34	molecular	36	brain	7	cells	9	mortality	34	human	22
systematic	33	gene	35	cells	7	annual	8	systematic	33	cells	21
review	33	cells	32	scientific	6	poison	8	review	31	dna	20
virus	30	dna	31	world	6	research	7	virus	28	protein	20
metaanalysis	28	data	26	study	5	review	7	metaanalysis	28	gene	20
cancer	27	sequence	25	global	5	trial	7	association	27	sequence	18
association	27	structure	25	life	5	human	7	health	26	methods	17
science	26	patients	24	quantum	5	control	6	adults	26	development	17
adults	26	system	24	neuroscience	5	disease	6	zika	25	software	17
health	26	expression	22	worlds	5	carbon	6	states	23	structure	16
zika	25	quantum	22	days	5	american	6	cancer	22	system	15
change	25	efficient	20	cancer	5	cell	6	analysis	21	calculations	14
states	24	rapid	19	papers	5	data	6	change	21	efficient	14
united	22	novel	18	just	4	association	5	united	21	simple	14
associated	21	software	18	change	4	taxonomy	5	consumption	21	patients	13
consumption	21	simple	17	reproducibility	4	imaging	5	associated	21	approach	12
analysis	21	proteins	17	giant	4	science	5	women	18	model	12
research	20	cell	17	embryos	4	development	5	trial	17	survey	12
women	20	improved	16	researchers	4	energy	5	cohort	16	statistics	11
trial	20	receptor	16	artificial	4	system	5	evidence	16	theory	11
evidence	19	graphene	16	may	4	global	5	prospective	16	improved	11
cells	19	functional	15	breakthrough	4	medicine	4	social	15	networks	11
social	18	development	15	theory	4	chemistry	4	cardiovascular	15	high	10
scientific	17	diabetes	15	access	3	virus	4	early	15	rna	10
brain	17	based	15	publishing	3	future	4	dietary	14	graphene	10
prospective	16	heart	15	woman	3	carcinogenicity	4	national	14	quantum	10
climate	16	breast	15	gas	3	green	4	weight	14	effects	10
early	16	methods	15	google	3	protein	4	cognitive	13	criteria	10
cohort	16	genome	15	prize	3	stem	4	climate	13	expression	10
humans	15	multiple	14	phd	3	goes	4	humans	13	models	10
cardiovascular	15	calculations	14	bad	3	scientists	4	randomized	13	treatment	9
life	14	activation	14	evidence	3	death	4	infection	12	sets	9
national	14	model	14	system	3	chemical	4	cells	12	plus	9
dietary	14	theory	14	largest	3	npds	4	heart	12	identification	9
weight	14	mutations	13	crispr	3	centers	4	obesity	11	effect	9
time	13	statistics	13	real	3	carcinogenspart	4	physical	11	density	9
food	13	characterization	13	teaching	3	alzheimers	4	research	11	program	9
heart	13	control	13	time	3	nanotubes	4	diet	11	detection	9
data	13	alignment	13	vaccine	3	national	4	scientific	11	clinical	8
obesity	13	disease	13	discovered	3	graphene	4	data	11	accurate	8
cognitive	13	effect	13	history	3	target	4	studies	11	systems	8
randomized	13	program	13	natural	3	nanotechnology	4	food	11	atoms	8
scientists	13	survival	12	take	3	little	3	sleep	11	cell	8
world	12	rna	12	government	3	promotion	3	death	11	basis	8
infection	12	application	12	gender	3	organic	3	diabetes	10	based	8

## Results

It can be seen from Tables 1 and 2 that the 50 most used words in the titles with the highest AAS (marked red) are also in the titles of the most cited publications in Web of Science (1/5 out of 100 are the same). The publications with high AAS tend to have longer titles (6,78) than the highly cited publications (6,45). The publications with the highest AAS have much more unique words (3444 and 3094).

Tables 3 and 4 show the words used in the 271 news items that are in the top 1000 publications with the highest AAS. The 271 news items with high AAS have a higher average AAS than articles of the same kind (Table 6) (1551,4 and 1412,3). The most cited news items are on average a bit longer (5,1 words and 4,6 words) and have more unique words (997 and 924). This and the fact that only 14 words out of 100 most used words are the same can be related to the average age of the news items with high AAS being an average of ten years younger (average date of publishing 2014) than the most cited news items (average date of publishing 2004) – most cited news cover more subjects.

In Tables 5 and 6 the 722 articles that are in the top 1000 publications with the highest AAS have more similar words with the 722 most cited article titles (21 out of 100). Articles with high AAS are over a word longer than the most cited (7,5 and 6,7) and have a lot more unique words (2612 and 2844). On average the 722 articles with the highest AAS are 155 times less cited than the most cited articles (54 and 8390).

## Discussion and conclusion

There are quite a few titles in the top 1000 publications with the highest AAS that have some kind of a pop cultural reference, linguistic attribute or structure attribute in them. Most of them are academic news titles that are written to attract attention and encourage reading like regular news titles with the titles and coverage being misleading at times (Sumner et al., 2014). However, there are also few academic publications with these references or attributes. Since the average citation per publication for the articles with high AAS is already 54,3 (39% of them were published in 2016), which is over twice as much as for all the articles in Web of Science from 1980 – 2015, it is safe to say that most of these top 1000 publications have also some academic credentials and as Sagi and Yachiam found, highly amusing titles receive fewer citations (Sagi, Yachiam, 2008).

One crucial difference between the titles is that high AAS titles use much more words with significant meaning at the moment of publishing like *zika* or *ebola*. This shows that alternative metrics work and they show the impact far before citations can.

Overall word usages shows that buzzwords are not overused in the titles of academic articles with high ASS and most of them have citations in major databases to reinsure that these articles also have an academic impact, meaning that they are correctly marketed publications with both a social and academic impact. This also shows that 20<sup>th</sup> century and 21<sup>st</sup> century metrics must be used together for a comprehensive view into science.

## References

- Ho, Y. S., Satoh, H., & Lin, S. Y. (2010). Japanese lung cancer research trends and performance in Science Citation Index. *Internal Medicine*, 49(20), 2219–2228.
- Jamali, H.R. & Nikzad, M. (2011). Article title type and its relation with the number of downloads and citations. *Scientometrics*, 88: 653. DOI: 10.1007/s11192-011-0412-z
- Lauk, K. (2016). A bibliometrical analysis of reserch published in Oil Shale. *Oil Shale*, 33(3), 290-297 DOI: 10.3176/oil.2016.3.07
- Lockwood, G. (2016). Academic clickbait: articles with positively-framed titles, interesting phrasing, and no wordplay get more attention online., *The Winnower*. 3:e146723.36330 DOI: 10.15200/winn.146723.36330
- Sagi I.; Yechiam E. (2008). Amusing titles in scientific journals and article citation. *Journal of Information Science*. 34:680–687. DOI: 10.1177/0165551507086261.
- Sumner, P.; Vivian-Griffiths, S.; Boivin, J. et al. (2014) The association between exaggeration in health related science news and academic press releases: retrospective observational study. *BMJ*. 349: g7015 DOI: 10.1136/bmj.g7015

## Contact

Kalmer Lauk, Bibliometrics Specialist at University of Tartu, kalmer.lauk@ut.ee  
Twitter @ KalmerLauk