

## Are citations from clinical trials evidence of higher impact research? An analysis of ClinicalTrials.gov

Item Type	Journal article
Authors	Thelwall, Mike;Kousha, Kayvan
Citation	Thelwall, M. & Kousha, K. Scientometrics (2016) 109: 1341. <a href="https://doi.org/10.1007/s11192-016-2112-1">https://doi.org/10.1007/s11192-016-2112-1</a>
DOI	<a href="https://doi.org/10.1007/s11192-016-2112-1">10.1007/s11192-016-2112-1</a>
Publisher	Springer
Journal	Scientometrics
Download date	2026-05-15 04:40:25
License	<a href="https://creativecommons.org/licenses/by-nc-nd/4.0/">https://creativecommons.org/licenses/by-nc-nd/4.0/</a>
Link to Item	<a href="http://hdl.handle.net/2436/619119">http://hdl.handle.net/2436/619119</a>

# Are citations from clinical trials evidence of higher impact research? An analysis of ClinicalTrials.gov<sup>1</sup>

Mike Thelwall, Kayvan Kousha

Statistical Cybermetrics Research Group, University of Wolverhampton, Wulfruna Street, Wolverhampton, UK.

Email: m.thelwall@wlv.ac.uk TEL/FAX: 01902 321000

An important way in which medical research can translate into improved health outcomes is by motivating or influencing clinical trials that eventually lead to changes in clinical practice. Citations from clinical trials records to academic research may therefore serve as an early warning of the likely future influence of the cited articles. This paper partially assesses this hypothesis by testing whether prior articles referenced in ClinicalTrials.gov records are more highly cited than average for the publishing journal. The results from four high profile general medical journals support the hypothesis, although there may not be a cause-and-effect relationship. Nevertheless, it is reasonable for researchers to use citations to their work from clinical trials records as partial evidence of the possible long-term impact of their research.

**Keywords:** Citation analysis; altmetrics; alternative metrics; clinical trials; medical research.

## Introduction

Medical research has the ultimate goal of improving health in society (e.g., <http://www.mrc.ac.uk/about/>) but it is rarely possible to demonstrate the direct benefits of individual studies due to the long gestation period before related treatments or interventions are approved for patients (e.g., Drolet, & Lorenzi, 2011; for a scientometric approach see: Jones & Hanney, 2016). Thus, it has become common to evaluate medical researchers both formally and informally with the aid of proxy indicators, such as the prestige or impact factors of the journals in which they publish, counts of citations to their articles, or their successful funding bids. A partial exception is the case study component of the UK Research Excellence Framework (REF) (<http://impact.ref.ac.uk/CaseStudies/>). The two citation based methods may undervalue research that is the most directly beneficial because health outcomes do not in themselves produce journal citations. Whilst basic research is important and can lead to long term social benefits (e.g., Smith, Mitchell, & McEwan, 2013), methods are needed to test for and, if necessary correct for, the apparent tendency for citation-based indicators to undervalue more applied medical research (e.g., Lewison & Dawson, 1998). One such indicator is the post-publication peer review scores given to biomedical articles by F1000 (Mohammadi & Thelwall, 2013). Moreover, citations can help to evaluate medical funders (Boyack & Jordan, 2011) and alternative metrics may help with this (Thelwall, Kousha, Dinsmore, & Dolby, 2016). Similar motivations have underpinned many attempts to develop altmetrics and other alternative indicators to reflect the different valid impacts generated by academics in many disciplines (Kousha & Thelwall, 2015b; Thelwall & Delgado, 2015; Wilsdon, J. et al., 2015).

The need for an indicator of the extent to which research informs health practice can be partially addressed by counting citations from medical guidelines (Drew, Pettibone,

---

<sup>1</sup> Thelwall, M. & Kousha, K. (in press). Are citations from clinical trials evidence of higher impact research? An analysis of ClinicalTrials.gov. *Scientometrics*.

Finch, Giles, & Jordan, 2016; Grant, Cottrell, Cluzeau, & Fawcett, 2000; Thelwall & Maflahi, 2016; for patents, see: Chiou, Magazzini, Pammolli, & Riccaboni, 2016; Kousha & Thelwall, 2015a). Nevertheless, it can take decades for medical research to translate into proven and safe health policies or medications and so citations from guidelines are unlikely to reflect recent research. Guideline citations may also tend to cite research that proves the efficacy or safety of a treatment rather than the primary research that invented the treatment or that initially demonstrated its promise. Citations from clinical trials records may fill both of these gaps by being substantially quicker and more likely to cite key innovations. There are now regulations mandating the public registration of clinical trials in some countries, such as the USA, where the responsible parties must register them in ClinicalTrials.gov prior to starting (Tse, Williams, & Zarin, 2009) and report results in a timely fashion (Prayle, Hurley, & Smyth, 2012). Moreover, some journals require published studies to have been pre-registered in an appropriate site (Laine, Horton, DeAngelis, et al., 2007; Palma, & Zietman, 2015). ClinicalTrials.gov contains information about a large number of public and commercial trials (Califf, Zarin, Kramer, Sherman, Aberle, & Tasneem, 2012; Hirsch, Califf, Cheng, et al., 2013; Ross, Mulvey, Hines, Nissen, & Krumholz, 2009; Stockmann, Sherwin, Koren, et al., 2014; Zarin, Tse, & Ide, 2005; Zarin, Tse, Williams, Califf, & Ide, 2011) and so it is now possible to count citations from clinical trials records. These citations can therefore be evaluated as a new source of evidence about the health benefits promise of medical research. Even though few trials based on published ideas lead to new treatments (Contopoulos-Ioannidis, Ntzani, & Ioannidis, 2003), the mere existence of a trial based on a publication suggests the potential practical value of the new ideas, if even fewer basic research findings lead to beneficial medical outcomes (e.g., Ioannidis, 2006). Nevertheless, the requirement for responsible parties to register clinical trials publicly is not internationally universal and there is no requirement to cite relevant prior work systematically. Hence, it is not clear whether there will be enough citations from clinical trials documents to be worth counting. It is also not clear whether citations from clinical trials are effectively random, biased or dominated by self-citations.

This article introduces a method to count citations from clinical trials and assesses whether they are likely to be useful as impact evidence. In order to investigate the value of clinical trials citations, articles in four high profile general medical journals were compared based upon whether they had been referenced in ClinicalTrials.gov records. It is important to compare articles from the same journal because the prestige of a journal can affect the number of times that an article is cited (Larivière & Gingras, 2010) and journals have different disciplinary foci, leading to different natural citation rates for their articles (Seglen, 1997). Selecting high profile journals ensures that the citation counts will tend to be high and the coverage of the journal will be large, both of which tend to maximise the statistical power of the analysis. The registry ClinicalTrials.gov was chosen despite the existence of many other registries internationally (Ogino, Takahashi, & Sato, 2014) due to its large downloadable collection of about 216,000 clinical studies from the United States and 193 countries since 2000 (<https://clinicaltrials.gov/ct2/resources/trends>; 37% were US-only trials in May 2016). Although published clinical trials tend to be more highly cited than average for the publishing journal (Romero, Cortés, Escudero, López, & Moreno, 2009), it is not known whether the same is true for articles cited by clinical trial records. The following two questions are addressed. A positive answer to the first question would give evidence that citations from ClinicalTrials.org are not random and also that they could be used for impact indicators (van Raan, 1998), but would not demonstrate that they are unbiased or

unaffected by self-citations. This is therefore a first, but not final, step in validating clinical trials citation counts.

- Do articles in The BMJ, the New England Journal of Medicine, The Lancet and JAMA tend to attract more citations than average for the publishing journals when they are referenced in ClinicalTrials.gov records?
- How does the answer to the above question vary over time?

## Data and Methods

Citations to journal articles in trials registered in ClinicalTrials.gov were identified by crawling all records in the site and extracting the cited references from the relevant sections of each document. The website permits crawling (<https://www.clinicaltrials.gov/robots.txt>) and makes comprehensive crawling without duplication possible with a crawler-friendly version of the site (<https://www.clinicaltrials.gov/ct2/crawl>). This was crawled by the free web crawler SocSciBot (socscibot.wlv.ac.uk) during April 19-24, 2016. A program was written to extract the cited publications (and other information) from each record. This is available in the free software Webometric Analyst (lexiurl.wlv.ac.uk: the ClinicalTrials.gov options in the Services and Citations menus).

Citations in trial records can be entered by the uploader or automatically inserted by ClinicalTrials.gov from trials that cite the study by its reference number. The automatically added citations are unlikely to have influenced the trials since they must have been published after the trial was registered in the site and hence all such citations were removed. In addition, the remaining citations are at the discretion of the uploader and may also refer to articles that publish the trial results, are influenced by the trial, or are otherwise related to the trial without influencing it. To reduce the possibility of these happening, only articles that had been published at least a year before the study start date were retained. Given normal publishing delays, these articles are likely to refer to research that was conducted at least 1-2 years prior to the start of the study and seem unlikely to be about the study rather than influencing it. This is a conservative approach given that articles publishing clinical trials results tend to appear about two years after the trial data collection has finished (Riveros, Dechartres, Perrodeau, Haneef, Boutron, & Ravaud, 2013).

The citation counts of all regular articles in the four journals were extracted from Scopus using queries for the journal ISSN, the document type article, and each year from 2008 to 2016. Multiple ISSNs for a journal in Scopus (e.g., electronic and print) were searched separately. For example, one JAMA query was:

ISSN(1538-3598) AND DOCTYPE(ar) AND PUBYEAR IS 2015

The range 2008 to 2016 was chosen because 2008 was the first year in which the journal articles (almost) all had DOIs in Scopus.

Articles mentioned in ClinicalTrials.gov records can be referenced in many different formats and with many different variants of journal names. To ensure that articles referenced in trials records were correctly matched to their Scopus record, the matching process used article DOIs, when present. The remaining articles were matched by journal name (standardised across common variants), publication year and title (converted to lower case). Most articles contained DOIs and so this was the dominant matching mechanism. Articles referenced in trials that were not matched to a Scopus record despite having been published in one of the journals and years covered were ignored. These were predominantly non-article types of document (e.g., editorials, errata, images, comments, letters).

For each year, the average citation counts of articles referenced in ClinicalTrials.gov were compared against the average citation counts of the remaining articles in the four journals examined. Individual years were analysed separately because older articles can be expected to have higher citation counts. Articles published early in the year have longer to be cited compared to articles published later in the year but this advantage seems unlikely to affect either of the two groups more than the other and so this factor was ignored. Although there is a statistical technique to deal with this issue, it does not allow effect sizes to be estimated (Thelwall, Haustein, Larivière, & Sugimoto, 2013) and so it was not used. For each journal/year combination, citation averages were calculated using the geometric mean with a 1 offset because citation data is highly skewed and so the arithmetic mean is inappropriate (Fairclough & Thelwall, 2015; Zitt, 2012). All citation counts were normalised by dividing by the geometric mean citation count for the journal and year of publication. This is an extension of the standard citation normalisation technique (Waltman, van Eck, van Leeuwen, Visser, & van Raan, 2011ab). This also allows 95% confidence intervals to be calculated for the (geometric) mean using the normal distribution formula both within individual years and across years.

## Results

In each of the four journals, the average number of citations attracted by prior articles referenced in ClinicalTrials.gov records is substantially higher than the average number of citations attracted by other articles in the same journals (Table 1). This supports the hypothesis that citation by a clinical trial record is evidence of the likelihood of an article having above average impact, at least for its publishing journal (RQ1).

**Table 1.** Journal and year geometric mean normalised citation counts for prior articles referenced in ClinicalTrials.gov and other articles in the four journals (geometric mean, 95% confidence interval and sample size).

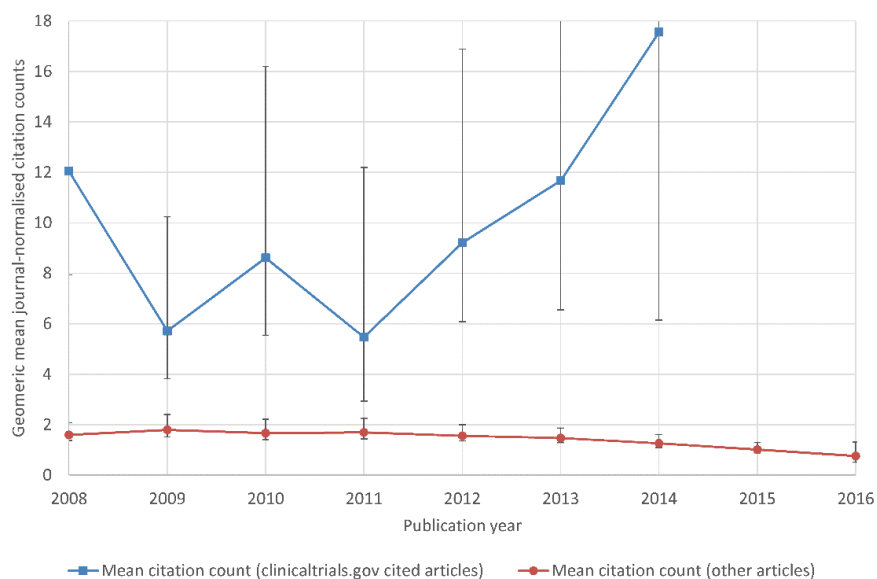
Journal	Mean citation count of articles referenced in ClinicalTrials.gov	Mean citation count of articles not referenced in ClinicalTrials.gov
BMJ	8.5 (7.1, 10.2) N=123	1.4 (1.3, 1.5) N=3937
JAMA	3.5 (3.1, 4.0) N=212	1.3 (1.2, 1.4) N=1681
The Lancet	8.1 (7.1, 9.3) N=291	1.3 (1.3, 1.4) N=2817
NEJM	12.1 (11.2, 13.1) N=489	1.5 (1.5, 1.7) N=3025

The temporal trend in results varies by publishing journal (Figs. 1-4; see Table 2 for sample sizes). The BMJ and JAMA results are broadly consistent with no changes over time, whereas the results from The Lancet suggest lower evidence of impact for more recently published articles and the NEJM results are consistent with both. The current year (2016) is missing from the graphs because no trial recording a future start date (2017 or later) cited articles from any of the four journals published in 2016. Overall, then, the data does not give strong evidence of a substantial time lag between being referenced in a clinical trials record and becoming highly cited (RQ2).

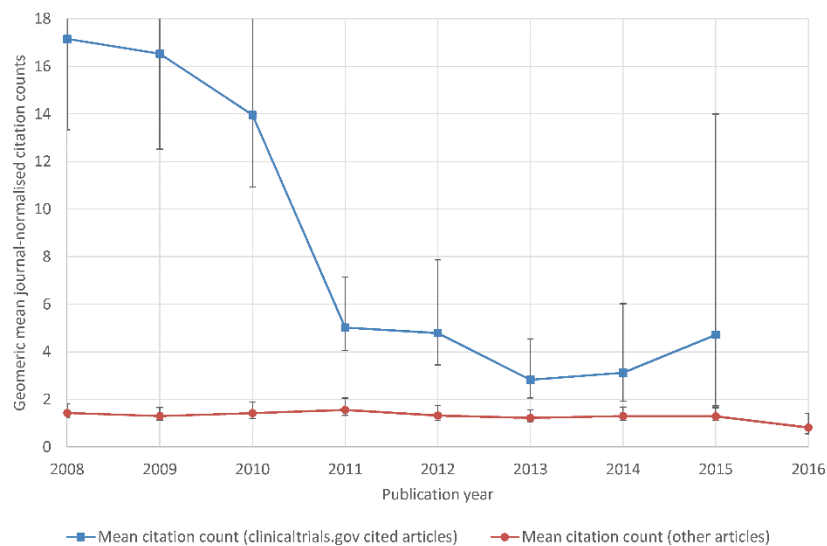
Because all citation counts are normalised by dividing by the journal geometric mean, only the relative size difference between the two sets is important. For example, the graphs mask (as irrelevant for this paper) the fact that NEJM is more highly cited than The BMJ.

**Table 2.** The number of articles in each of the four journals analysed that were referenced or not referenced by ClinicalTrials.gov records.

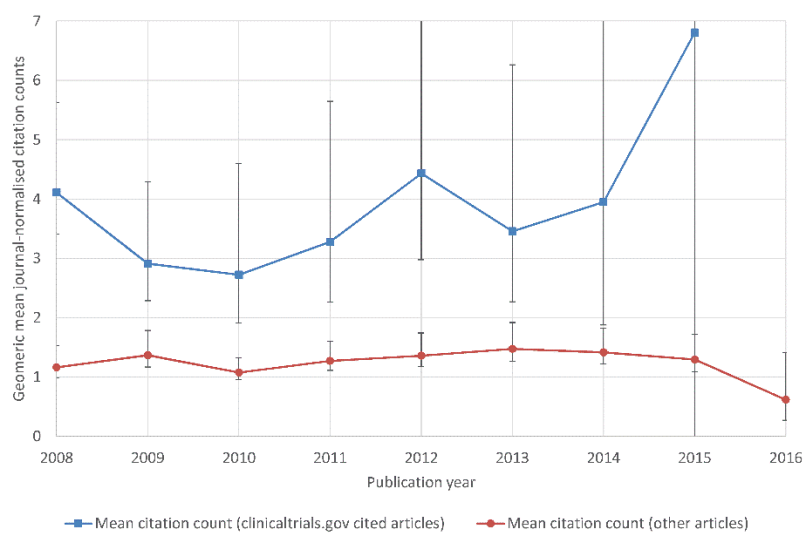
Publication year	BMJ cited	BMJ uncited	Jama cited	Jama uncited	Lancet cited	Lancet uncited	NEJM cited	NEJM uncited
2008	8	430	44	154	37	581	71	370
2009	25	350	51	171	25	572	64	394
2010	31	339	42	176	18	431	61	375
2011	22	326	35	174	18	279	60	365
2012	28	494	38	211	10	252	52	398
2013	15	577	41	213	6	241	39	392
2014	14	661	37	225	4	232	24	389
2015	5	624	26	196	2	304	22	350
2016	1	110	4	55	2	94	8	80
All	149	3911	318	1575	122	2986	401	3113



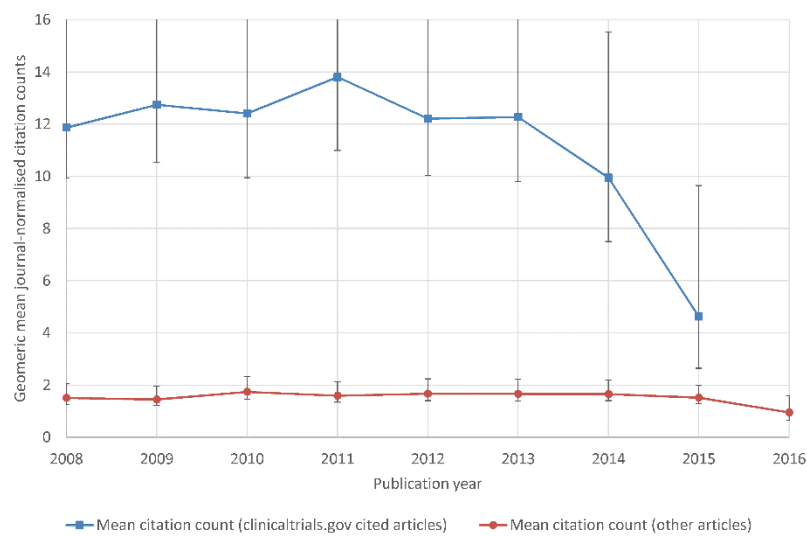
**Figure 1.** Journal and year normalised average numbers of citations for articles published between 2008 and 2016 in The BMJ: articles referenced in ClinicalTrials.gov records (and published before the study start date) against the remaining articles.



**Figure 2.** As above for The Lancet.



**Figure 3.** As above for JAMA.



**Figure 4.** As above for NEJM.

## Discussion and conclusions

An important limitation is that ClinicalTrials.gov is a US-based site and therefore its citations can be expected to have a bias towards the US, as is the case with journal citations (Lancho-Barrantes, Bote, Vicente, Rodríguez, & de Moya Anegón, 2012). The methodological restriction to a small number of high impact journals is another important limitation because these account for a small minority of medical articles and it is not clear that the patterns found in them would extend to the generality of medical research. For example, a high proportion of publications in high impact journals about a new topic is a statistically significant indirect indicator of a drug eventually being marketed (Kissin, 2010). Another weak point is that citation counts do not seem to reflect the quality of clinical trials (Akcan, Axelsson, Bergh, Davidson, & Rosén, 2013) and so citation-based metrics may not be useful indicators for trial-related publications. Finally, citations from clinical trials records may be biased in some way, may have a high proportion of self-citations, may disproportionately cite high impact types of documents (e.g., randomized control trials, systematic reviews or meta analyses) rather than primary research. Moreover, papers may also be more referenced by clinical trials if they have many authors or more funding, which may also affect their Scopus citation counts (e.g., Glänzel & Schubert, 2001). Thus, a more detailed analysis of the individual citations would be needed to validate the use of clinical trials citations for evaluation purposes.

The results give strong statistical evidence that articles from four leading medical journals that are cited with DOIs in ClinicalTrials.gov have higher citation counts than typical articles from the journal, at least for those that are 2-6 years old, depending on the journal, with the difference probably being due to statistical power differences rather than journal differences. There are at least four potential reasons for the higher citation impact found.

- Articles referenced in ClinicalTrials.gov lead to medical or health care innovations by **motivating** (at least in part) the trials, and this value leads to them being more recognised and cited (including in articles derived from the trials).
- Articles referenced in ClinicalTrials.gov do **not** lead to medical or health care innovations but become more cited due to (a) being cited in articles derived from unsuccessful trials and/or (b) publicity due to the presence of the citation in the trial reference list.
- Articles referenced in ClinicalTrials.gov are disproportionately often **self-citations** from the clinical trials scientists, who tend to be highly cited successful researchers because they attract funding for clinical trials.
- Older articles referenced in ClinicalTrials.gov are **selected because of their high citation counts**, either directly (as prestigious articles) or indirectly (due to ranking boosts in digital library search engines or because their high contribution to research has already been recognised and led to high citations).
- Articles referenced in ClinicalTrials.gov tend to be **highly cited types of articles**, such as randomised control trials, systematic reviews or meta-analyses.

Citation in ClinicalTrials.gov would be most useful as an alternative impact indicator in the first case only because in the third and fourth cases, the indicator would tend to confirm what other indicators would show and in the second case the citation does not reflect a valuable contribution of the cited research. The higher citation impact is in addition to the more direct impact that articles can have on health outcomes through supporting or motivating clinical trials.

The temporal results were inconclusive but do not give strong evidence for the value of using citation in a clinical trial record as evidence of likely future higher citation impact. Logically, however, the opposite is almost certainly true. Although articles published in the same year as the start of a trial were excluded for methodological reasons, when they are not derived from a trial then it would be valid to count contemporary trial citations to them. Thus, it would be possible to use citations from trials in the current year as evidence of likely future impact in practice. Such articles would be likely to have low numbers of journal article citations due to their newness and so the trial citation would be a valuable impact indicator.

A practical application of the findings is that researchers can use the presence of citations to their work in clinical trials records to support a claim for its potential long-term impact. Such citations can be found easily by searching the site using their last name (e.g., Google *site:clinicaltrials.gov/ct2/show/Whooley*) or the titles of their articles (e.g., Google *site:clinicaltrials.gov/ct2/show/ "Prevalence of symptomatic pelvic floor disorders in US women"*), and manually checking for correct matches. Given the possibility of irrelevant citations, as discussed above, authors might accompany such claims by brief contextual justification of the value of the cited research to the study (e.g., "informed study methods"). From a wider research policy perspective, the potential of researchers to use evidence of citations from clinical trials as an indicator of likely future impact may provide a mild incentive to them to encourage the translation of their ideas into practice, which would presumably be a public good.

## References

- Akcan, D., Axelsson, S., Bergh, C., Davidson, T., & Rosén, M. (2013). Methodological quality in clinical trials and bibliometric indicators: no evidence of correlations. *Scientometrics*, 96(1), 297-303.
- Boyack, K. W., & Jordan, P. (2011). Metrics associated with NIH funding: a high-level view. *Journal of the American Medical Informatics Association*, 18(4), 423-431.
- Califf, R. M., Zarin, D. A., Kramer, J. M., Sherman, R. E., Aberle, L. H., & Tasneem, A. (2012). Characteristics of clinical trials registered in ClinicalTrials.gov, 2007-2010. *Jama*, 307(17), 1838-1847.
- Chiou, J. Y., Magazzini, L., Pammolli, F., & Riccaboni, M. (2016). Learning from successes and failures in pharmaceutical R&D. *Journal of Evolutionary Economics*, 26(2), 271-290.
- Contopoulos-Ioannidis, D. G., Ntzani, E. E., & Ioannidis, J. P. (2003). Translation of highly promising basic science research into clinical applications. *The American Journal of Medicine*, 114(6), 477-484.
- Drew, C. H., Pettibone, K. G., Finch Iii, F. O., Giles, D., & Jordan, P. (2016). Automated Research Impact Assessment: a new bibliometrics approach. *Scientometrics*, 106(3), 987-1005.
- Drolet, B. C., & Lorenzi, N. M. (2011). Translational research: understanding the continuum from bench to bedside. *Translational Research*, 157(1), 1-5.
- Fairclough, R., & Thelwall, M. (2015). More precise methods for national research citation impact comparisons. *Journal of Informetrics*, 9(4), 895-906. doi: 10.1016/j.joi.2015.09.005
- Glänzel, W., & Schubert, A. (2001). Double effort= double impact? A critical view at international co-authorship in chemistry. *Scientometrics*, 50(2), 199-214.

- Grant, J., Cottrell, R., Cluzeau, F., & Fawcett, G. (2000). Evaluating “payback” on biomedical research from papers cited in clinical guidelines: Applied bibliometric study. *Bmj*, 320(7242), 1107-1111.
- Hirsch, B. R., Califf, R. M., Cheng, S. K., Tasneem, A., Horton, J., Chiswell, K., ... & Abernethy, A. P. (2013). Characteristics of oncology clinical trials: insights from a systematic analysis of ClinicalTrials.gov. *JAMA internal medicine*, 173(11), 972-979.
- Ioannidis, J. P. (2006). Evolution and translation of research findings: from bench to where. *PLOS ONE*, 1(7), e36.
- Jones, T. H., & Hanney, S. (2016). Tracing the indirect societal impacts of biomedical research: development and piloting of a technique based on citations. *Scientometrics*, 107(3), 975-1003.
- Kissin, I. (2010). Can a bibliometric indicator predict the success of an analgesic? *Scientometrics*, 86(3), 785-795.
- Kousha, K. & Thelwall, M. (2015a). Patent citation analysis with Google. *Journal of the Association for Information Science and Technology*. doi: 10.1002/asi.23608
- Kousha, K. & Thelwall, M. (2015b). Web indicators for research evaluation, part 3: Books and non-standard outputs. *El Profesional de la Información*, 24(6), 724-736. doi:10.3145/epi.2015.nov.04
- Laine, C., Horton, R., DeAngelis, C. D., Drazen, J. M., Frizelle, F. A., Godlee, F., ... & Sahni, P. (2007). Clinical trial registration—looking back and moving ahead. *New England Journal of Medicine*, 356(26), 2734-2736.
- Lancho-Barrantes, B. S., Bote, G., Vicente, P., Rodríguez, Z. C., & de Moya Anegón, F. (2012). Citation flows in the zones of influence of scientific collaborations. *Journal of the American Society for Information Science and Technology*, 63(3), 481-489.
- Larivière, V., & Gingras, Y. (2010). The impact factor's Matthew Effect: A natural experiment in bibliometrics. *Journal of the American Society for Information Science and Technology*, 61(2), 424-427.
- Lewis, G., & Dawson, G. (1998). The effect of funding on the outputs of biomedical research. *Scientometrics*, 41(1-2), 17-27.
- Mohammadi, E., & Thelwall, M. (2013). Assessing non-standard article impact using F1000 labels. *Scientometrics*, 97(2), 383-395.
- Ogino, D., Takahashi, K., & Sato, H. (2014). Characteristics of clinical trial websites: information distribution between ClinicalTrials.gov and 13 primary registries in the WHO registry network. *Trials*, 15(1), 428.
- Palma, D. A., & Zietman, A. (2015). Clinical Trial Registration: A Mandatory Requirement for Publication in the Red Journal: *International Journal of Radiation Oncology*, 91(4), 685-686.
- Prayle, A. P., Hurley, M. N., & Smyth, A. R. (2012). Compliance with mandatory reporting of clinical trial results on ClinicalTrials.gov: cross sectional study. *Bmj*, 344, d7373.
- Riveros, C., Dechartres, A., Perrodeau, E., Haneef, R., Boutron, I., & Ravaut, P. (2013). Timing and completeness of trial results posted at ClinicalTrials.gov and published in journals. *PLoS Med*, 10(12), e1001566.
- Romero, A., Cortés, J., Escudero, C., López, J., & Moreno, J. (2009). Measuring the influence of clinical trials citations on several bibliometric indicators. *Scientometrics*, 80(3), 747-760.

- Ross, J. S., Mulvey, G. K., Hines, E. M., Nissen, S. E., & Krumholz, H. M. (2009). Trial publication after registration in ClinicalTrials.gov: a cross-sectional analysis. *PLoS Med*, 6(9), e1000144.
- Seglen, P. O. (1997). Why the impact factor of journals should not be used for evaluating research. *BMJ: British Medical Journal*, 314(7079), 498.
- Smith, L. B., Mitchell, R. T., & McEwan, I. J. (2013). *Testosterone: From Basic Research to Clinical Applications*. Berlin: Springer.
- Stockmann, C., Sherwin, C. M., Koren, G., Campbell, S. C., Constance, J. E., Linakis, M., ... & Spigarelli, M. G. (2014). Characteristics and publication patterns of obstetric studies registered in ClinicalTrials.gov. *The Journal of Clinical Pharmacology*, 54(4), 432-437.
- Thelwall, M. & Delgado, M. (2015). Arts and humanities research evaluation: No metrics please, just data. *Journal of Documentation*, 71(4), 817-833. doi:10.1108/JD-02-2015-0028
- Thelwall, M., Kousha, K., Dinsmore, A. & Dolby, K. (2016). Alternative metric indicators for funding scheme evaluations. *Aslib Journal of Information Management*, 68(1), 2-18. doi:10.1108/AJIM-09-2015-0146
- Thelwall, M., Haustein, S., Larivière, V. & Sugimoto, C. (2013). Do altmetrics work? Twitter and ten other candidates. *PLOS ONE*, 8(5), e64841. doi:10.1371/journal.pone.0064841
- Thelwall, M. & Maflahi, N. (2016). Guideline references and academic citations as evidence of the clinical value of health research. *Journal of the Association for Information Science and Technology*, 67(4), 960-966. doi:10.1002/asi.23432
- Tse, T., Williams, R. J., & Zarin, D. A. (2009). Reporting “basic results” in ClinicalTrials.gov. *CHEST Journal*, 136(1), 295-303.
- van Raan, A. F. (1998). In matters of quantitative studies of science the fault of theorists is offering too little and asking too much. *Scientometrics*, 43(1), 129-139.
- Waltman, L., van Eck, N. J., van Leeuwen, T. N., Visser, M. S., & van Raan, A. F. (2011a). Towards a new crown indicator: Some theoretical considerations. *Journal of Informetrics*, 5(1), 37-47.
- Waltman, L., van Eck, N. J., van Leeuwen, T. N., Visser, M. S., & van Raan, A. F. (2011b). Towards a new crown indicator: An empirical analysis. *Scientometrics*, 87(3), 467-481.
- Wilsdon, J. et al. (2015). *The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management*. [http://www.hefce.ac.uk/pubs/rereports/Year/2015/metrictide/Title\\_104463\\_en.html](http://www.hefce.ac.uk/pubs/rereports/Year/2015/metrictide/Title_104463_en.html)
- Zarin, D. A., Tse, T., Williams, R. J., Califf, R. M., & Ide, N. C. (2011). The ClinicalTrials.gov results database—update and key issues. *New England Journal of Medicine*, 364(9), 852-860.
- Zarin, D. A., Tse, T., & Ide, N. C. (2005). Trial registration at ClinicalTrials.gov between May and October 2005. *New England Journal of Medicine*, 353(26), 2779-2787.
- Zitt, M. (2012). The journal impact factor: angel, devil, or scapegoat? A comment on JK Vanclay's article 2011. *Scientometrics*, 92(2), 485-503.

## Appendix: Specific method details

The free web crawler SocSciBot (<http://socscibot.wlv.ac.uk>) crawled the ClinicalTrials.gov index site at <https://www.clinicaltrials.gov/ct2/crawl>) and the link files in the SocSciBot link results folder imported into Excel and used to extract a complete list of URLs. From these URLs, a complete list of standard study URLs was generated.

A second SocSciBot crawl was then started, using a 1-second pause between requests (for politeness) and using a dummy fake startup URL and the above list of URLs as the start.txt initial list (this ensures that only URLs in the list are crawled).

When the crawl had finished, the free Webometric Analyst software (<http://lexiurl.wlv.ac.uk>) was used to extract the relevant information as follows.

- The *Services* menu, *ClinicalTrials.gov: Extract information from SocSciBot crawl* (e.g., *AgesEligible*, *MeshTerms*, *Publications*, *PubsLinkedViaNCT*) menu item was used to extract relevant information from the crawled pages (pointing the program to the root folder for the second crawl). >> **Record list file.**
- The *Services* menu, *ClinicalTrials.gov: Summarise results of the above* (e.g., *study type*, *gender by year: pubs*; *keyword freqs.*) menu item was used to generate summary information (pointing the program to the record list file generated above). >> **Summary tables files.**
- The *Citation* menu, *Clinicaltrials.gov (or similar): Extract citations from |-separated col, extract doi and year, match with record year* menu item was used to generate lists of users-added publications (pointing the program to the record list file generated above, and selecting the user-added publications column). >> **List of user added publications file.**
- Using Scopus, a list of all publications of type journal article was extracted for JAMA, BMJ, Lancet and NEJM. >> **Scopus journal file.**
- The Scopus publication list was matched with each of the above two publications files by exact doi matching (case insensitive) or, for articles without DOI matches, title, journal name and year matching (case insensitive) using the *Citation* menu, *Match DOIs or article title, year, journal from one file with same info from another* menu item (selecting the Scopus journal file first, then the List of user added publications file). >> **List of user added publications and matching Scopus records file.**
- Duplicate user added publication matches (i.e., the same publication cited by different trials records) were removed in Excel. Publications from the same or later year than the trial start year were also removed in Excel. >> **List of unique user added prior publications and matching Scopus records file.**
- Geometric mean normalised citation counts for each file were created using the *Citation* menu, *Calculate geometric mean normalised citation counts for a marked subset of articles, separately by year* menu item (selecting the Scopus journal file first, then the List of user added publications file). >>**geometric mean normalised user added results file.**