Journal of Clinical Epidemiology 109 (2019) 80–89

Journal of Clinical Epidemiology

ORIGINAL ARTICLE

An evidence rating service provided valid correlates of the clinical importance of medical articles and journals

R. Brian Haynes[a,*], Dalton Budhram[b,c], John Cherian[d,e], Emma Iserman[f], Alfonso Iorio[g], Cynthia Lokker[h]

[a]*Professor Emeritus, Health Information Research Unit, Department of Health Research Methods, Evidence and Impact and Department of Medicine, McMaster University Faculty of Health Sciences, Hamilton, Ontario L8S 4K1, Canada*

[b]*Research Assistant, Health Information Research Unit, McMaster University, Hamilton, Ontario L8S 4K1, Canada*

[c]*Medical Student, Faculty of Medicine, Queen's University, Kingston, Ontario, Canada*

[d]*Medical Student on Research Elective, Health Information Research Unit, McMaster University Faculty of Health Sciences, Hamilton, Ontario L8S 4K1, Canada*

[e]*Internal Medicine Resident, Department of Medicine, Michigan State University College of Human Medicine, East Lansing, MI, USA*

[f]*Research Associate, Health Information Research Unit, McMaster University Faculty of Health Sciences, Hamilton, Ontario L8S 4K1, Canada*

[g]*Professor and Chief, Health Information Research Unit, Department of Health Research Methods, Evidence and Impact and Department of Medicine, McMaster University Faculty of Health Sciences, Hamilton, Ontario L8S 4K1, Canada*

[h]*Assistant Professor, Health Information Research Unit, Department of Health Research Methods, Evidence and Impact, McMaster University Faculty of Health Sciences, Hamilton, Ontario L8S 4K1, Canada*

Accepted 23 January 2019; Published online 5 February 2019

Abstract

Objectives: The objective of this study was to determine reliability and validity of McMaster PLUS measures of scientific merit and clinical importance of articles in medical journals.

Study Design and Setting: Analytic survey of peer-reviewed medical journals was carried out. Articles were qualified for inclusion by meeting (1) scientific criteria and (2) a clinical importance rating threshold. Included articles were sent as e-mail alerts to physicians according to their clinical interests. Internal measures included the number of high-quality, clinically important studies published in source journals and response to alerts. For external validation, we correlated internal measures with the Journal Impact Factor (JIF) and citation in DynaMed Plus (DMP).

Results: We evaluated 34,232 articles from 57 journals. Inclusion criteria were met by 2,638 articles (7.71%). The number of qualifying articles per journal was correlated with the number of articles with high clinical importance ratings (r 0.96, $P < 0.001$), article alert clicks (r 0.86, $P < 0.001$), and DMP citations (r 0.99, $P < 0.001$). Correlation was lower with the JIF (r 0.68, $P < 0.01$).

Conclusions: Measures of scientific merit and clinical importance of medical journal articles were strongly correlated with each other, less so with JIFs. Journals varied widely by these measures but, generally, few articles were both scientifically sound and clinically important.

Keywords: Information retrieval; Knowledge translation; Journalology; Journal impact factor; McMaster PLUS database; DynaMed plus

1. Introduction

Health care practitioners are expected to keep pace with important advances in evidence concerning the management of conditions for which they offer care. The medical journal literature is usually the first widely accessible source of new, peer-reviewed evidence from research, but even top medical journals offer a very dilute

Funding: This research was conducted without external funding.

Conflict of interest: McMaster PLUS is owned by McMaster University, a not-for-profit, publicly funded university in Hamilton, ON, Canada. PLUS was created by RBH, is cosupervised by A.I. and R.B.H., and employs E.I. and C.L. PLUS has, or has had, contracts for evidence-based information services with American College of Physicians, British Medical Journal Group, Canadian Medical Association, Canadian Pharmacist Association, EBSCO (owner of DMP), Elsevier, Journal of Bone and Joint Surgery, McGraw Hill, Medscape, and Wolters Kluwer. R.B.H. is a review author with the Cochrane Collaboration and was a member of its Board of Directors in 1994. At the inception of this study, D.B. was a research assistant and J.C. was a medical student on elective at McMaster University, working with R.B.H.

* Corresponding author. Health Information Research Unit, McMaster University Faculty of Health Sciences, 1280 Main Street West, CRL 133, Hamilton, Ontario L8S 4K1, Canada. Tel.: +1-905-525-9140x20152; fax: +1-905-526-8447.

E-mail address: bhaynes@mcmaster.ca (R.B. Haynes).

https://doi.org/10.1016/j.jclinepi.2019.01.010

What is new?

Key findings

- Even for leading clinical journals, the proportion of articles that are of adequate research merit and high clinical interest is less than 10 percent, and this rate varies substantially across journals.
- Measures of the scientific merit of, and clinical interest in, journal articles are highly correlated with each other and with citation in a point-of-care evidence-based clinical text. Correlations of these measures with journal and article citations are statistically significant but weaker.
- The Cochrane Database for Systematic reviews (CSDR) led in several performance metrics, including the number and proportion of articles entering the McMaster PLUS database, number of articles with high clinical ratings, and number of article clicks in a clinical e-mail alerting service.
- Comparing the journals by citations in the medical literature for articles in the McMaster PLUS database, New England Journal of Medicine generated the most citations, while CSDR placed seventh.

What this adds to what was known?

- An evidence rating service (McMaster PLUS) provides highly reliable and valid measures of quality and clinical importance of published health care research articles and medical journals. Ranking clinical journals based on the frequency of their publication of articles of adequate research quality and high clinical relevance reveals important differences in journal performance. The Cochrane Database of Systematic Reviews performs much better from this perspective than according to journal citation indexes.

What is the implication and what should change now?

- The McMaster Plus journal performance metrics may be useful to guide decisions about journal selection and subscriptions by clinicians, professional organizations, clinical libraries, and others whose purpose is to support evidence-based health care and who have budgets or time constraints that cannot sustain subscribing to all leading clinical journals.

stream of articles that are simultaneously scientifically sound, clinically relevant, and immediately important for clinical practice [1].

A possible option for clinicians to manage this problem could be to limit their reading to the very best journals in their field. However, important problems abound with this approach. For instance, measures of journal performance, such as the Journal Impact Factor (JIF), are based solely on citations in the scientific literature, not practitioner use or patient outcomes in clinical settings [2,3]. Correlations between impact factors and clinical ratings of journal quality are of low order, [4] as is the association between the methodologic quality of studies and citation rates [5]. Also, it has been well documented that articles on a given topic are widely dispersed among journals [6–8]. For example, Garg et al. [8] found that 2,779 references cited in systematic reviews relevant to nephrology were scattered among 466 journals, undermining the notion that a clinician can keep abreast of their field by reading a few journals within their own discipline.

A convenient option for clinicians is to use regularly updated, evidence-informed, ‘‘point of care’’ resources such as Best Practice (BMJ Publishing Group), DynaMed Plus (EBSCO), Evidence-Based Medicine Guidelines (Duodecim), and UpToDate (Wolters Kluwer). These texts provide evidence in the context of contemporary clinical care, but they also have limitations, including variable rates of updating, [9] incomplete topic coverage, [10] few details of studies, proprietary access, and limited or no alerting for subscribers, who may be unaware when practice-changing evidence becomes available for their clinical interests.

A complementary option is for clinicians to subscribe to a medical literature rating service tailored for their specialty. The McMaster Premium Literature Service (PLUS) is a ‘‘health knowledge refinery’’, running continuously since 2003 to alert clinicians to high quality, clinically interesting studies and systematic reviews, according to the clinicians’ self-specified discipline(s) (61 options, from general practice to all major specialties). PLUS metrics of article clinical impact, including number of articles per journal meeting scientific criteria and clinical ratings of article interest, assessed at the time of publication, have previously been shown to predict independently collected article citation counts [11] and JIFs, [12] which only become available many months after publication.

In this investigation, we assessed the reliability of measures that are part of the PLUS appraisal process as well as their validation by indicators of physician interest and research merit.

2. Methods

The methods for the PLUS ‘‘second-order peer review process’’ for clinical journals are described in detail elsewhere [13] and the current journal list, review process, and scientific criteria appear on this website: http://hiru.mcmaster.ca/hiru/HIRU_McMaster_PLUS_projects.aspx. Briefly, validated online journal search filters with high sensitivity (≥99%) for retrieving high quality, clinically relevant articles [14] screen

the surveyed journals daily for articles at the first date of their posting to PubMed. Articles that pass the filter are then assessed by trained research staff to identify those that meet basic inclusion criteria: (1) English language; (2) humans; and (3) relevant topics for health care. Qualifying original studies and review articles are then classified for clinical purpose: therapy and prevention, screening and diagnosis, prognosis, clinical prediction, etiology of adverse effects, quality improvement, and cost-effectiveness.

Research staff then assess each article by applying explicit criteria (summarized at https://hiru.mcmaster.ca/hiru/InclusionCriteria.html) for scientific merit for health care; this step is designated as the critical appraisal process (CAP filter). The reliability of this process has been verified, with kappa values for agreement of 89% (95% confidence interval 78% to 99%) for methodologic criteria applied by multiple research staff doing independent assessments [15]. Every article passing CAP criteria is checked by at least one clinician with a working knowledge of clinical epidemiology, and any remaining questions are addressed by a senior clinical editor.

Articles that pass CAP are then independently rated by at least three members of a world-wide volunteer panel of practicing clinicians for each relevant clinical discipline for the article's topic. For each article they rate, raters must indicate that they are currently in clinical practice and have no conflict of interest for that article. Articles are rated for clinical interest (clinical importance ratings [CIRs]) on anchored 7-point scales (with seven being high; see Fig. 1) for each of clinical relevance (for the rater's main clinical discipline) and newsworthiness (new information for doctors in the same discipline).

To be included in PLUS, articles that pass CAP must have CIRs that average four or more on each scale. CIRs are available for at least 80% of articles within 10 days of passing the critical appraisal process, which is typically within 1 to 30 days of first posting to PubMed (based on the ''create record date'', CRDT).

Qualifying articles are posted to PLUS as soon as at least three raters for at least one clinical discipline have completed their ratings. The PLUS database feeds several evidence-based medicine online information services, including EvidenceAlerts (which was named EvidenceUpdates up to 2016), which serves 61 clinical specialties and subspecialties in primary care, internal medicine, surgery, obstetrics, gynecology, pediatrics, oncology, and mental health. This service contains all PLUS content and is free of charge but users must

1. Relevance to clinical practice in your discipline:

Score		Criterion
	○	Beyond my personal area of expertise but may be of interest to my discipline
7	○	Directly and highly relevant
6	○	Definitely relevant
5	○	Probably relevant
4	○	Possibly relevant: likely of indirect or peripheral relevance at best
3	○	Possibly not relevant
2	○	Probably not relevant: content only remotely related
1	○	Definitely not relevant: completely unrelated content area

2. Newsworthiness to clinical practice in your discipline:

	○	I don't know if this is newsworthy
7	○	Useful information, most practitioners in my discipline definitely don't know this (unless they have read this article)
6	○	Useful information, most practitioners in my discipline probably don't know this
5	○	Useful information, most practitioners in my discipline possibly don't know this
4	○	Useful information, most practitioners in my discipline possibly already know this
3	○	Useful information, most practitioners in my discipline probably already know this
2	○	It probably doesn't matter whether they know this or not
1	○	Not of direct clinical interest

Fig. 1. Clinical importance ratings (CIR) scales for articles that have passed the critical appraisal process (CAP).

register and specify their clinical specialty(ies) of interest to receive alerts. Alerts contain article citation details and URL links to the article's abstract; CIRs and any rater comments; free full-text article, if available; and the article's posting in PubMed. Accessing (''clicking'') any of these links by subscribers is automatically recorded and aggregately used as a measure of subscriber interest. Click data in this report are restricted to users who were registered as physicians in 2015.

For the purposes of the present study, we defined three internal measures and four external measures of article and journal performance (Table 1). Internal measures arise from the PLUS database and external measures are developed and curated by external entities. Alert clicks by users were collected for a 6-month window after an article's release. Internal measures were based on all articles that passed criteria from the 57 top journals in our surveillance list. These 57 journals yield the highest number of qualifying articles per annum from over 370 journals that we have assessed since 1995. The year 2015 was chosen to enable the use of article citation data collected for January 2016 through August 2017 and use of the 2016 JIFs (which are based on citations of articles published in the preceding 2 years). DynaMed Plus (DMP) was chosen as the point of care reference standard based on its comparative performance in updating as shown in a systematic review, and the availability of its citations of articles published in 2015 (see Acknowledgments). Most DMP citations (DMPC) are for prevention, treatment, screening, and diagnosis, so citation rates were based on these same categories in the PLUS database (n = 1,451 articles). All other metrics were based on all 2,237 PLUS-qualifying articles from the 57 journals, including the aforementioned DMP categories as well as prognosis, clinical prediction, etiology of harm, quality improvement, and economics-cost effectiveness.

2.1. Questions

Our primary question was as follows: What is the reliability (consistency) of measures of clinical importance of articles per journal that meet criteria for research merit and clinical interest via the PLUS health knowledge refinery, as measured by (1) number of articles per journal qualifying for the PLUS database; (2) number of articles per journal with mean CIRs (clinical relevance + newsworthiness ratings) ≥11 (see below); (3) number of end-user clickthroughs to a journal's articles in the largest of the PLUS subscriber services, EvidenceUpdates (now named EvidenceAlerts https://plus.mcmaster.ca/evidencealerts/), in the first 6 months after posting in PLUS?

Based on previous studies of individual article ratings, citation counts [11], and JIFs [12], we hypothesized that the CIRs for a journal's articles would be the most credible measure of a journal's performance from a clinical perspective. A cutoff of a total ≥11 (relevance + newsworthiness, of a maximum of 14) was selected as best discriminating among the journals, based empirically on having the fewest number of ties in ranked journal performance when compared with cutoffs of 10 or more, 12 or more, and 13 or more.

Our secondary questions were: (1) For the rankings of journals based on each of the internal measures in the primary question, what is the correlation with the rankings based on the published 2016 JIF for these journals? (2) Based on articles in PLUS published in 2015 and their citations per journal from January 2016 through August 2017, what is the comparative ''journal clinical article citation index'' (JCACI) of the journals? 3. Based on citations of articles published in 2015 in DMP (DMPC), what is the comparative performance of the journals?

Performance measures are defined in Table 1. Alerts were derived from EvidenceUpdates for articles published in 2015. Citation counts for January 2016 through August 2017 for PLUS articles for the 57 journals and JIFs for 2016 for these journals were accessed from Clarivate Analytics (https://clarivate.com/). Based on the distribution of mean values for article CIRs, we also rank-ordered journals based on the number of their articles that had average combined ratings for relevance + newsworthiness of at least 11, 12, or 13.

Table 1. Internal and external performance measures

Type of factor	Performance factor	Definition
Internal to PLUS	Articles per journal in the PLUS database	Number of articles that (1) passed the automated journal filter; and (2) entered the Premium LiteratUre Service (PLUS), including research assistant's (RA) assessment, clinical editor's review, and physicians' ratings for clinical relevance and newsworthiness (CIRs, average rating of ≥4/7 on each rating scale).
	Articles with mean combined CIRs ≥11/14	Number of articles per journal in PLUS with a mean combined relevance + newsworthiness score ≥11 for all CIRs.
	Alert clicks	Number of user clicks on e-mail alerts for articles from a given journal during the first 6 months after dissemination from PLUS, as measured in EvidenceAlerts.
External to PLUS	2016 Journal Impact Factor (JIF)	The JIF from 2016 as imported from InCites Journal Citation Reports. The 2016 JIF is the average number of citations of articles published in a journal during the 2 preceding years.
	Total citations (TC)	Number of citations in 2016 through August 2017 to a journal's PLUS articles published in 2015.
	Journal clinical article citation index (JCACI)	Number of citations in 2016 through August 2017 per journal for PLUS articles published in 2015 divided by the number of PLUS articles for that journal published in 2015.
	DMP citations (DMPC)	Number of PLUS articles per journal published in 2015 and cited in DynaMed Plus (DMP).

Table 2. Performance of journals (measures are defined in Table 1)

	Internal to PLUS measures of clinical performance			External to PLUS measures			
Journal	Articles in PLUS	Articles with mean relevance + news score ≥11/14	Alert clicks	Total citations (TC)	Journal clinical article citation index (JCACI)	Journal Impact Factor (2016)	Articles cited in DynaMed (DMP)
Acad Emerg Med	13	1	2,138	108	8.308	2.537	2
Am J Cardiol	49	3	4,889	358	7.306	3.154	14
Am J Gastroenterol	17	2	1,699	254	14.941	10.383	9
Am J Kidney Dis	15	5	983	206	13.733	6.269	6
Am J Med	19	2	2,856	163	8.579	5.610	4
Am J Obstet Gynecol	35	13	2,497	196	5.600	4.681	13
Am J Psychiatry	18	3	2,533	374	20.778	13.505	8
Am J Respir Crit Care Med	22	5	2,566	517	23.500	13.118	4
Am J Sports Med	23	6	889	231	10.043	4.517	16
Ann Emerg Med	18	3	4,131	128	7.111	5.008	7
Ann Intern Med	49	17	7,572	1,554	31.714	16.593	34
Ann Rheum Dis	34	7	2,508	617	18.147	12.384	8
Ann Surg	40	19	2,128	329	8.225	8.569	6
Arch Dis Child Fetal Neonatal Ed	12	4	674	43	3.583	3.969	10
Arch Phys Med Rehabil	43	1	1,735	245	5.698	3.045	17
BJOG	29	3	1,639	164	5.655	4.096	12
BMJ	66	27	10,387	1,514	22.939	19.697	34
Br J Psychiatry	12	2	1,184	194	16.167	7.06	5
Br J Surg	39	8	1,625	290	7.436	5.596	19
Chest	27	4	2,714	242	8.963	6.136	7
Circulation	26	6	4,221	682	26.231	17.202	13
Clin J Pain	22	3	1,479	57	2.591	2.712	3
Clin Rehabil	54	1	2,614	150	2.778	2.403	26
CMAJ	10	5	1,974	106	10.600	6.724	2
Cochrane Database Syst Rev	703	103	55,898	2,122	3.018	6.103	558
Crit Care Med	24	6	2,554	296	12.333	7.422	3
Diabet Med	28	2	1,657	127	4.536	3.152	10
Diabetes Care	41	7	4,097	824	20.098	8.934	18
Diabetes Obes Metab	29	4	2,857	360	12.414	6.198	5
Eur Heart J	47	15	5,612	1,369	29.128	15.064	18
Eur Respir J	14	1	795	195	13.929	8.332	7
Gastroenterology	28	12	3,295	845	30.179	18.187	13
Gut	19	7	1,913	270	14.211	14.921	8
Health Technol Assess	70	19	6,652	234	3.343	4.058	45
Heart	19	3	2,230	203	10.684	5.693	10
Int J Clin Pract	18	2	2,191	57	3.167	2.226	6
J Am Coll Cardiol	43	23	5,794	1,849	43.000	17.759	14
J Am Geriatr Soc	26	1	1,937	132	5.077	3.842	4
J Bone Joint Surg Am	14	8	886	131	9.357	5.163	9

(*Continued*)

Table 2. Continued

Journal	Internal to PLUS measures of clinical performance			External to PLUS measures			
	Articles in PLUS	Articles with mean relevance + news score ≥11/14	Alert clicks	Total citations (TC)	Journal clinical article citation index (JCACI)	Journal Impact Factor (2016)	Articles cited in DynaMed (DMP)
J Clin Nurs	16	2	782	24	1.500	1.384	0
J Clin Oncol	90	28	1,941	2,762	30.689	20.982	31
J Infect Dis	13	0	385	132	10.154	6.344	3
J Pediatr	29	9	2,280	213	7.345	3.89	14
J Rheumatol	18	2	937	119	6.611	3.236	0
J Vasc Surg	20	3	319	148	7.400	3.454	8
JAMA	74	37	13,113	3,638	49.162	37.684	52
Lancet	110	58	12,460	6,493	59.027	44.002	78
Lancet Oncol	69	30	2,530	4,040	58.551	26.509	45
N Engl J Med	121	87	17,751	20,880	172.562	59.558	101
Neurology	41	10	4,263	553	13.488	8.166	13
Obesity (Silver Spring)	18	0	983	81	4.500	3.614	4
Obstet Gynecol	43	9	2,424	349	8.116	5.656	33
Pediatrics	69	14	7,321	626	9.072	5.196	38
Rheumatology (Oxford)	18	4	1,845	89	4.944	4.524	4
Spine (Phila Pa 1976)	24	1	834	104	4.333	2.439	4
Stroke	34	2	2,556	263	7.735	5.787	13
Thorax	16	3	2,123	99	6.188	8.121	3

Abbreviations: CMAJ, Canadian Medical Association Journal; PLUS, Premium LiteratUre Service.

2.2. Statistical analysis

Descriptive data were prepared including frequencies (e.g., for number of articles passing criteria for each journal) and ranked data (for journal performance, to permit comparison with the 2016 JIF). Using SPSS (version 24, IBM), we created a correlation matrix between the internal measures in question 1 and between these measures and citation counts, JIFs, and citations in DMP for the journals. Pearson product moment correlations (r) were applied to continuous data, with 95% 2-tailed confidence intervals. Journal ranks were expected to vary according to the various performance measures, with no ''gold standard'', and no statistical comparisons between individual journals were preplanned; accordingly, overall journal ranks were ordered in quintiles rather than as individual journal rankings.

3. Results

Using the definitions in Table 1, journal performance data for the 57 journals are presented in Table 2 and rank groupings by quintiles in Table 3. Of 34,232 articles published in 2015, 2,707 (7.9%) passed critical appraisal (CAP), and 2,638 (7.7%) had acceptable CIRs of relevance (mean $\geq$ 4/7) and newsworthiness ($\geq$4/7), qualifying for entry to the PLUS database.

3.1. Internal performance measures

Journal performance, judged by the number of articles from each journal entering the PLUS database in 2015 (Table 2), ranged by orders of magnitude, from the Cochrane Database of Systematic Reviews (CDSR) contributing 703 articles, to the Canadian Medical Association Journal with 10 articles.

Of the 2,638 articles qualifying for PLUS in the publication year 2015, 662 (25.1%) articles had combined mean CIRs of at least 11 (Table 2). CDSR again led the way with 103 articles at or above the cutoff, followed by New England Journal of Medicine (N Engl J Med) (87 articles), Lancet (58 articles), Journal of the American Medical Association (JAMA, 37), and Lancet Oncology (30). Just 143 articles (5.4%) had averaged CIRs of at least 12/14; only 11 articles (0.4%) averaged at least 13, and no article had averaged CIRs of 14.

For the number of article click-throughs by users (Table 2), CDSR articles led with 55,898 alert clicks, then N Engl J Med (17,751), JAMA (13,113), Lancet (12,460), and BMJ (10,387). Summing the journal rankings for the three internal performance measures in Table 1, treating each measure equally, the five highest ranked journals were CDSR, N Engl J Med, Lancet, Pediatrics, and JAMA. The performance ranks for all journals according to our primary

Table 3. Journal performance ranks by quintile, ordered on numbers of articles from each journal in 2015 with mean combined clinical importance ratings (CIRs) at least 11

Performance quintile	Journal (alphabetical order within quintiles)	
First	Ann Intern Med	J Am Coll Cardiol
	Ann Surg	J Clin Oncol
	BMJ	Lancet
	Cochrane Database Syst Rev	Lancet Oncol
	Health Technol Assess	N Engl J Med
	JAMA	
Second	Am J Obstet Gynecol	J Pediatr
	Ann Rheum Dis	J Bone Joint Surg Am
	Br J Surg	Neurology
	Diabetes Care	Obstet Gynecol
	Eur Heart J	Pediatrics
	Gastroenterology	
Third	Am J Respir Crit Care Med	Circulation
	Am J Kidney Dis	Crit Care Med
	Am J Sports Med	Diabetes Obes Metab
	Arch Dis Child Fetal Neonatal Ed	Gut
	CMAJ	Rheumatology (Oxford)
	Chest	
Fourth	Am J Cardiol	Br J Psychiatry
	Am J Gastroenterol	Clin J Pain
	Am J Med	Heart
	Am J Psychiatry	J Vasc Surg
	Ann Emerg Med	Thorax
	BJOG	
Fifth	Acad Emerg Med	J Clin Nurs
	Arch Phys Med Rehabil	J Infect Dis
	Clin Rehabil	J Rheumatol
	Diabet Med	Obesity (Silver Spring)
	Eur Respir J	Spine (Phila Pa 1976)
	Int J Clin Pract	Stroke
	J Am Geriatr Soc	

Abbreviations: CMAJ, Canadian Medical Association Journal.

measure, number of articles per journal with combined mean clinical score $\geq$11, are presented in quintiles in Table 3.

3.2. External performance measures

Comparing the journals by citations in the medical literature, articles in PLUS published by N Engl J Med in 2015 generated the most citations (20,880), followed by Lancet (6,493), Lancet Oncology (4,040), and JAMA (3,638), with CDSR falling to seventh place (2,122). The JCACI, which divides total citations for PLUS articles per journal, gave identical rankings for the top four journals, but CDSR dropped to 54th place, based on its low citation count per article. DMPC, however, placed CDSR first (558 articles cited in 620 topics), then N Engl J Med (101 cited in 115 topics), Lancet (78 cited in 76 topics), JAMA (52 cited in 65 topics), and Health Technol Assess (45 cited in 50 topics).

3.3. Correlations among measures

The number of articles per journal with high CIRs, which we designated a priori as the leading measure of journal clinical performance, was significantly correlated with each of the other internal measures, qualified for PLUS (r 0.79, $P < 0.001$) and accessed by users (r 0.86, $P < 0.001$) (Table 4). The number of articles per journal with higher CIRs was also correlated with all external measures, including the JIF (r 0.68, $P < 0.01$), total citations for PLUS articles per journal (r 0.73, $P < 0.01$), JCACI (r 0.64, $P < 0.01$), and PLUS articles cited in DMP (r 0.80, $P < 0.001$). Correlations of internal measures with total citations of these articles were similar to those with the JIF. The JCACI and JIF were highly correlated (r 0.92, $2P < 0.001$), but JCACI average citations per journal (17.0) were generally higher than JIF averages (9.9; $P < 0.001$). Journal rankings based on the three internal measures were each highly correlated with article citation in DMP, including number of articles in the PLUS database (Pearson r 0.99, $P < 0.001$), number of articles with high CIRs (r 0.96, $P < 0.001$), and article alert clicks (r 0.86, $P < 0.001$).

4. Discussion

We used data from an ongoing evidence-based clinical literature rating service to assess the quality, quantity, and

Table 4. Pearson correlations (and 95% confidence intervals) among internal measures and between internal and external measures of journal performance

	Internal measures (see Table 1)			External measures (see Table 1)			
Internal measures	**Articles in the PLUS database**	**Articles with high clinical ratings (CIRs)**	**Alert clicks**	**2016 JIF**	**Total citations (TC)**	**JCACI**	**Articles cited in DynaMed**
Articles in the PLUS database	1	0.79*** (0.67 to 0.87)	0.96*** (0.94 to 0.98)	0.15 (−0.11 to 0.40)	0.24 (−0.03 to 0.47)	0.11 (−0.16 to 0.36)	0.995*** (0.991 to 0.997)
Articles with high CIRs		1	0.86*** (0.78 to 0.92)	0.68** (0.51 to 0.80)	0.73*** (0.57 to 0.83)	0.64*** (0.46 to 0.77)	0.80** (0.68 to 0.88)
Alert clicks			1	0.31* (0.05 to 0.53)	0.37* (0.12 to 0.58)	0.26 (0.00 to 0.49)	0.97** (0.94 to 0.98)

*$P < 0.05$, 2-tailed; **$P < 0.01$, 2-tailed; ***$P < 0.001$, 2-tailed.

perceived clinical importance, of articles from 57 medical journals. Measures derived from the rating process and end-user response were highly correlated with one another, and the number of highly rated articles per journal was moderately correlated with the JIF and other citation-based measures. All three internal measures were highly correlated with citation in an evidence-based clinical text, DMP. The high correlations of the internal measures are not surprising as they are successive stages of the same appraisal process, albeit performed sequentially by different actors: research staff, clinical raters, and end-users.

Consistent with our a priori hypothesis that the number of highly clinically rated articles was the most robust internal measure, it was the only internal measure that was significantly correlated with all four external measures. These findings support the validity of the clinical importance rating process, added on to the critical appraisal process. The moderate correlation of the number of highly rated articles with the citation-based measures also suggests that the internal and external measures assess complementary aspects of journal merit, particularly relating to perceived importance to physicians in clinical practice in comparison with researchers who publish in the medical literature.

Comparing the internal and external measures of individual journals, the most striking contrast is the first rank of CDSR in the numbers of qualified articles for PLUS, high clinical relevance and newsworthiness scores, and physician interest (click counts), compared with its modest JIF (6.103). The higher performance of CDSR from a clinical perspective is at least partly explained by higher clinical ratings for relevance for systematic reviews, compared with original articles (mean clinical relevance ratings, 5.46 vs. 5.27, respectively, absolute difference 0.18, 95% CI 0.10 to 0.27, $2P < 0.001$), perhaps somewhat offset by nonsignificantly lower mean ratings for newsworthiness for reviews vs. original articles (4.75 vs. 4.83, $2P = 0.09$). Furthermore, Cochrane reviews were rated as slightly less relevant on average than reviews published in the other 56 journals (mean 5.30 vs. 5.40, difference −0.09, CI −0.17 to −0.02, $2P < 0.01$) and less newsworthy (mean 4.45 vs. 4.69, difference −0.25, CI −0.32 to −0.18, $2P < 0.001$). Thus, it is review articles in general that clinical raters see as more relevant than original articles on average. It is intriguing that the CDSR publishes only systematic reviews while the second-place journal for measures of clinical relevance, N Engl J Med, has a policy against publishing systematic reviews.

The difference in CDSR between clinical performance and citation findings is at least partly due to the much larger volume of qualifying articles in PLUS published by CDSR (703), compared with other journals (e.g., 121 for N Engl J Med), coupled with the considerably lower citation counts for CDSR articles (three citations per article for CDSR, vs. 172 for N Engl J Med). Citations arise from researchers publishing articles, who, given our findings, are more likely to cite original studies, perhaps because their projects began, or articles were submitted, before the reviews were published. For physicians trying to keep up to date, assuming that they are expressing their needs through clinical importance ratings and click-throughs from e-mail alerts, CDSR in aggregate far out-paced the other journals. For example, CDSR alerts generated a total of 55,895 click-throughs, vs. 17,751 for N Engl J Med, the second-ranked journal.

Of interest, however, if journals sought to increase their JIFs, they could do so by publishing a higher proportion of articles that are both scientifically sound and clinically important. The 2016 JIF for the 57 journals averaged 9.870; the JCACI showed that if the JIF had been based just on articles meeting PLUS criteria, the average JIF would have risen to 16.596, a 72% increase. The obvious problem is that there are not enough higher quality, clinically interesting articles available for publication, hence the need for services that concentrate, then distill such articles across journals that are able to attract them.

There are limitations to our research. First, no ''gold standard'' exists for determining the impact of journal articles on practitioner performance or patient outcomes. We used multiple ratings of clinical relevance and newsworthiness by practicing clinicians as a surrogate, but the relationship of such ratings to observed clinical practice and, most importantly, patient outcomes is unknown. Second, PLUS serves general medical practice and a broad range of specialties but it is not comprehensive for subspecialties. Third, we aggregated all the clinical importance ratings across specialties relevant to a given article; the findings for how journals perform from a clinical perspective could, and likely would, differ by specialty. That said, a high proportion of articles for a given specialty are published in the more general medical journals, rather than in the journals for that discipline [8]. Fourth, research staff, clinical editors, clinical raters, and users were not blinded to the journal of publication for the articles they were perusing, which may have resulted in biases in ratings and click-throughs. However, the selection of articles for rating is governed by explicit criteria, applied by research staff in a highly reproducible manner [15]. Fifth, the choice of a combined average clinical importance rating of at least 11 of 14 to determine the clinical impact of the articles by journal was arbitrary, based on somewhat fewer ties (i.e., better discrimination) in ranking the journals when compared with other cutoffs. Additional research is needed to determine the variability across years of the measures of performance in this study. Sixth, the strength of the correlations of the internal measures and DMPC may have been biased by DMP's authors' use of PLUS article feeds as one of their key sources for updating the DMP text. However, the DMP literature surveillance includes over 500 journals, almost an order of magnitude greater than the 57 journals in our study. Seventh, publishing applied research articles of direct clinical relevance is only one of several goals of medical journals, which may include basic research, news, commentaries, editorials, discussions forums, letters, and other items pertaining to health and health care. However, we would argue that publishing articles linking strong research to clinical practice is the most important function for clinical journals.

5. Conclusions

The McMaster PLUS process provides reproducible and externally valid measures of clinical articles at the time of their publication, and ranking of journals for publication of articles of clinical importance.

We posit that ranking clinical journals based on the frequency of publication of articles of adequate research quality and high clinical relevance has important advantages. It is based on reproducible ratings of scientific merit by research staff; multiple, specialty-specific, ratings of relevance and interest by practicing clinicians who have declared no conflicts of interest for each of the articles they are assessing; and clinical end-user response. Furthermore, these assessments are available much sooner than JIFs, which lag behind journal publication by 2 years.

Our results may be useful to guide decisions about journal selection and subscriptions by clinicians, professional organizations, clinical libraries, and others whose purpose is to support evidence-based health care and who have budgets or time constraints that cannot sustain subscribing to all 57 journals. Clinical journals that are regularly reviewed but did not make this short-list can be obtained from the McMaster PLUS website (https://hiru.mcmaster.ca/hiru/journalslist.asp), and a list of journals that have been assessed, but did not qualify for regular review can be obtained from the authors. These listings could be helpful in further rationalizing journal selections. Editors of clinical journals could also use the findings to focus their review process and article selection.

CRediT authorship contribution statement

R. Brian Haynes: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing - original draft, Writing - review & editing. **Dalton Budhram:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Validation, Visualization, Writing - original draft, Writing - review & editing. **John Cherian:** Conceptualization, Investigation, Writing - review & editing. **Emma Iserman:** Data curation, Formal analysis, Methodology, Project administration. **Alfonso Iorio:** Conceptualization, Resources, Visualization, Writing - review & editing. **Cynthia Lokker:** Conceptualization, Methodology, Writing - review & editing.

Acknowledgments

The authors are grateful to Dr Ruth Hertzman Miller of DMP, who arranged to provide the PubMed identification numbers of articles published in 2015 and cited in the DMP text. The authors thank the clinical raters who volunteer their time in the McMaster Online Rating of Evidence (MORE) service.

References

[1] Haynes RB. Loose connections between peer reviewed clinical journals and clinical practice. Ann Intern Med 1990;113:724−8.
[2] Schoonbaert D, Roelants G. Citation analysis for measuring the value of scientific publications: quality assessment tool or comedy of errors? Trop Med Int Health 1996;I:739−52.
[3] Ogden TL, Bartley DL. The ups and downs of journal impact factor. Ann Occup Hyg 2007;52(2):73−82.
[4] Saha S, Saint S, Christakis DA. Impact factor: a valid measure of journal quality? J Med Libr Assoc 2003;91:42−6.
[5] Lee KP, Schotland M, Bacchetti P, Bero LA. Association of journal quality indicators with methodological quality of clinical research articles. JAMA 2002;287:2805−8.

[6] Narayan VM, Chrouser K, Haynes RB, Parrish R, Dahm P. Defining the publication source of high quality evidence in urology: an analysis of EvidenceUpdates. BJU Int 2016;117:861−6.
[7] Baier L, Wilczynski NL, Haynes RB. Tackling the growth of the obesity literature: obesity evidence spreads across many journals. Int J Obes (Lond) 2009;34(10):1526−30.
[8] Garg AX, Iansavitchous AV, Kastner M, Walters LA, Wilczynski N, McKibbon A, et al. Lost in publication: half of all renal practice evidence is published in non renal journals. Kidney Internat 2006;70: 1995−2005.
[9] Jeffery R, Navarro T, Lokker C, Haynes RB, Wilczynski NL, Farjou G. How current are leading evidence-based medical texts? An analytic survey of four online textbooks. J Med Internet Res 2012;14(6):e175.
[10] Prorok JC, Iserman EC, Wilczynski NL, Haynes RB. Quality, breadth, and timeliness of content updating of ten online medical texts: an analytic survey. J Clin Epidemiol 2012;65:1289−95.
[11] Lokker C, McKibbon KA, McKinlay RJ, Wilczynski NL, Haynes RB. Data available at the time of publication successfully predict 2-year citation counts for clinical articles: retrospective cohort study. BMJ 2008;336:655−7.
[12] Lokker C, McKibbon KA, McKinlay RJ, Wilczynski NL, Haynes RB. How well are journal and clinical article characteristics associated with the Journal Impact Factor? J Med Libr Assoc 2012; 100:28−33.
[13] Haynes RB, Cotoi C, Holland J, Walters L, Wilczynski NL, Jedraszewski D, et al. McKibbon A for the McMaster Premium LiteratUre Service (PLUS) Project. A second order of peer review: a system to provide peer review of the medical literature for clinical practitioners. JAMA 2006;295:1801−8.
[14] Wilczynski NL, Haynes RB. Electronic culling of the clinical research literature: Filters to reduce the burden of hand searching. International Congress on Peer Review of the Biomedical Literature. Chicago, IL, September 9-10 2013: [abstract].
[15] Wilczynski NL, McKibbon KA, Haynes RB. Enhancing retrieval of best evidence for health care from bibliographic databases: calibration of the hand search of the literature. Stud Health Technol Inform 2001;10(Pt 1):390−3.