BMJ Open

Here's something I prepared earlier: A systematic review of the time to publication of cross-sectional reviews of smartphone health apps

SCHOLARONE™
Manuscripts

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Here's something I prepared earlier:
A systematic review of the time to publication of cross-sectional reviews of smartphone health apps

Mark E. Larsen, DPhil[1]
Jennifer Nicholas, PhD[2,3, 4]
Jin Han, BSc, PhD[1]
Christopher Lemon, MPhil[5,6]
Kelsi Okun[7]
Michelle Tye, PhD[1]
David Wong, DPhil[8]
Iana Wong, MSc[1]
Quincy J. J. Wong, PhD[9]
Kit Huckvale, PhD[1]

[1] Black Dog Institute, UNSW Sydney, Australia
[2] Orygen, Victoria, Australia
[3] Centre for Youth Mental Health, The University of Melbourne, Australia
[4] Center for Behavioral Intervention Technologies, Northwestern University, Chicago, IL, United States
[5] St Vincent's Hospital, 390 Victoria Street, Sydney, Australia
[6] Faculty of Medicine, UNSW Sydney, Australia
[7] School of Humanities and Sciences, Stanford University, Palo Alto, CA, United States
[8] Division of Informatics, Imaging & Data Sciences, University of Manchester, United Kingdom
[9] School of Psychology, Western Sydney University, Australia

Corresponding author:
Mark Larsen
Black Dog Institute, UNSW Sydney, Hospital Road, Randwick NSW 2031, Australia
+61 (0)2 9382 8508
mark.larsen@blackdog.org.au

Abstract

Objectives: Across a range of health conditions, apps are increasingly valued as tools for supporting the delivery and co-ordination of healthcare. Research-led cross-sectional reviews of apps are a potential resource to inform app selection in face of uncertainties around content quality, safety, and privacy. However, these peer-reviewed publications only capture a snapshot of highly dynamic app stores and marketplaces. To determine the extent to which marketplace dynamics might impact the interpretation of app reviews, the current study sought to quantify the lag between the reported time of app assessment and publication of the results of these studies.

Design: Searches were conducted on MEDLINE, Embase and PsycINFO to identify published cross-sectional reviews of health, fitness, or wellness apps. Publication timeline metadata were extracted, allowing the primary outcome measure, the delay between app store search and manuscript publication, to be calculated. A secondary measure, the time between search and manuscript submission, was also calculated where possible.

Results: After screening, 136 relevant cross-sectional app review studies were analysed. The median time to publication was 431 days (approximately 14 months, range: 42-1054 days). The median time to submission was 269 days (approximately 9 months, range: 5-874 days). Studies which downloaded apps typically took longer to publish (p=0.010), however the number of apps reviewed did not impact the time to publication (p=0.964). Studies which recommended specific apps were not published more rapidly (p=0.998).

Conclusions: Most health app reviews present data that is at least a year out of date at the time of publication. Given the high rate of turnover of health apps in public marketplaces, it may not be appropriate, therefore, for these reviews to be presented as a resource concerning specific products for commissioners, clinicians and the public. Alternative sources of information may be better calibrated to the dynamics of the app marketplace.

Keywords

Strengths and limitations of this study

- This review considers over a decade of published cross-sectional reviews of health apps.
- The age of the review findings, at the time of publication, was determined and compared to the observed rate of change of the app stores.
- The time to journal submission was also calculated, where possible, providing an indication of the quickest possible time for results to be made publicly available to inform decisions.
- Heterogeneity across reviewed clinical and technical domains may impact publication time.

Introduction

Rationale

Smartphone applications (apps) are increasingly valued as tools for supporting the delivery and coordination of healthcare. Across a range of health conditions, there is growing evidence that app-based self-care interventions can be effective at reducing symptoms,[1] supporting self-management,[2,3] and promoting health behaviour change.[4,5] In 2017, half of surveyed Australian primary care doctors reported recommending apps to their patients at least once a month.[6] Across both physical[7] and mental health,[8] consumers either indicate interest in using health apps or report having already attempted to integrate apps into their health management. At a systems level, there is growing interest in the potential for digital health to enable value-based care that offers potential resource savings compared to face-to-face therapies. Examples of established initiatives include the Australian Federal Government's e-Mental Health Strategy,[9] which seeks to increase the accessibility and reach of mental health support while decreasing load on traditional services, and the state of California's Technology Suite Collaborative, which is harnessing digital technology to expand the capacity and capability of the county mental health systems, again to decrease burden on traditional care pathways.[10] Most recently, the National Health Service in England stated its intention for 'digitally-enabled primary and outpatient care [to] go mainstream' across the entire health system as part of its long term plan.[11]

Quality remains a key concern for healthcare providers seeking to integrate health apps into routine care. Content quality, safety, and privacy deficits have been identified in a wide range of health app categories.[12-15] Despite recent progress in clarifying regulatory requirements around 'software as a medical device' in the USA and Europe, only a small fraction of available apps either fall into a category that requires formal regulatory review or have been subject to experimental evaluation. Indeed, the number of health apps evaluated through randomised studies within the research literature is dwarfed by the numbers available to consumers.[16] These apps are typically made available within the same commercial marketplaces as apps for navigation, social media, and finance. This combination of prevalent quality issues and potentially large numbers of options presented without technical differentiation represents a major challenge for healthcare systems, clinicians, and consumers trying to select high quality, clinically-appropriate apps.

Research-led cross-sectional reviews of published apps ('app reviews') that critically appraise aspects of app quality and safety are a potential resource for consumers and practitioners, to inform rational app selection. Indeed, many app reviews either state, as aims, an intention to guide health professionals and consumers to the best apps for a given health condition or make recommendations targeting clinicians in discussion.[17] There is now a substantial collection of such reviews; our searches identified at least 149 such studies published between 2008 and 2019. Evidence of the potential impact on practice and policy of these cross-sectional studies include citations in clinical guidelines,[18] professional guidance concerning health app use[19,20] and design,[21] health system policy documents,[22,23] and expert[24] and intergovernmental[25] reports. Tools commonly used in app reviews, such as the Mobile App Rating Scale

(MARS),[26] were developed with the explicit goal of providing an app evaluation resource for use by health professionals (as well as researchers.)

However, there is a critical, and widely acknowledged limitation of these reviews: they are static snapshots of a dynamic environment. Within app stores, app updates, additions, removals, and search result list changes are common and unpredictable. In 2016, the dynamic nature of the two leading commercial app environments was quantified.[27] Tracking the search results for depression, bipolar disorder, and suicide prevention apps each day over 9 months, findings indicated that half of the Google Play search results change approximately every 4 months. Moreover, across both platforms, an app for depression became unavailable to download every 2.9 days.[27] These dynamic changes highlight the potential for information contained in cross-sectional reviews to become out-of-date, limiting its validity if used for the purposes of selecting and recommending specific health apps.

Objective

In order to explore the extent to which marketplace dynamics might impact the interpretation of research-led app reviews, the current study sought to quantify the lag between the reported time of app assessment and publication of the results of these studies. We assessed the impact on time-to-publication of specific features of the review process likely to act as a proxy for researcher workload, such as whether assessment involved downloading and reviewing app content. Finally, given that some app reviews explicitly state their intention to influence professional and patient behaviour, for example by recommending specific apps for use, we tested the hypothesis that these studies would be published more rapidly.

Methods

Literature search

We aimed to identify reported studies that performed cross-sectional analysis, assessments, or reviews of smartphone health and wellbeing apps. To identify studies, we developed a bespoke literature search strategy. Working separately, two reviewers (KH and JN) first performed exploratory searches of articles published in 2018 indexed by the MEDLINE citation database. Each reviewer used these searches to try to devise, respectively, a specificity-maximising and sensitivity-maximising search strategy (detailed in Supplementary File 1.)

In order to evaluate the performance of these alternative strategies, the results of each search (specificity-maximising n=78, sensitivity-maximising n=220) were independently screened by two reviewers (ML and JN). After reconciling any differences, screening yielded a binary partition of relevant/non-relevant studies for each search strategy. Subsequent comparison of these results indicated the overall suitability of a specificity-maximising approach. Individual discrepancies in included/excluded studies were also reviewed, yielding qualitative judgements about the likely contributions of different search terms to the observed results.

We used this information to devise a unified search strategy based on the original specificity-maximising approach. This strategy was then re-run on the original sample of 2018 citations to confirm that no relevant citations were omitted. In a final step, we broadened the search to include all years and selectively removed terms from the strategy to ascertain their impact on the final result set. Terms that did not alter the overall result count were discarded. Search results were also reviewed to ensure that studies already known to the reviewers were captured by the strategy. The final search strategy is detailed in Box 1.

(apps or applications or (app adj development)).ti.

AND

(smartphone? or mobile? or cell? or cellular? or (smart adj phone?) or iphone?).ab,ti.

AND

(review or (cross adj sectional) or content or quality or survey or assessment).ti. or (mobile adj2 rating adj scale?).ab. or (google adj play).ab.

AND

(appstore? or store? or marketplace? or (market adj place?) or apple or google or android or download$ or (app$ adj rating adj scale?)).ab.

Box 1. Optimised specificity-maximising search strategy.

Searches were run on MEDLINE, Embase and PsycINFO on 30 April 2019 and included all studies published between 2008 (on the basis that this was the year in which the first commercial app store was launched) and the search date. Search results were combined and deduplicated before screening.

Eligibility criteria

Search result titles and abstracts were reviewed against a standard set of inclusion criteria. Studies were retained if they i) focussed on a topic relating to health, fitness, or wellness (irrespective of whether the intended app users were consumers, carers, clinicians, researchers or some combination of these); ii) involved a cross-sectional search of an app store or library intended to generate a set of apps for subsequent examination; and iii) applied one or more methods to this set to evaluate either the metadata associated with each app (such as app store descriptions), the contents of each app, or both.

Study selection

Each study was screened by two out of three reviewers (of KH, ML and JN), working independently, with any discrepancies resolved by the third reviewer. Inter-rater agreement during initial screening of studies returned by the original search strategy (n=78) was calculated using Fleiss' Kappa at 0.78, indicating substantial agreement between reviewers.

Data extraction

Following screening, the full text of each included study was obtained for data extraction. Table 1 details the data elements that were extracted, if available. Coding aimed to quantitate the time taken to publish each study and identify proxy measures of the effort required for its completion (e.g. number of apps included in the study, and whether apps were downloaded as part of the review process.) A final parameter concerning whether study authors identified specific apps in their results or discussion (e.g. to recommend for, or caution against, use) was collected to investigate whether inclusion of such recommendations influenced publication speed. The data extraction schema was developed through an initial pilot phase in which n=60 studies were reviewed by three reviewers (KH, ML and JN) to identify relevant data items and confirm the feasibility of extraction.

Category	Item	Description
Publication timeline	Earliest search date	Earliest stated date that app stores were searched.
	Latest search date	If a date range is specified, the latest stated date that app stores were searched.
	Updated search date	Date of any subsequent searches or updates to the original search.
	Submission date	Date of manuscript submission to journal.
	Publication date	Earliest identified date at which the accepted, peer-reviewed manuscript is made available to the public – which may be an online-first/electronic pre-print. Pre-prints prior to manuscript acceptance were not considered.
	Dates imputed	A Boolean variable coded as: FALSE if both the search date and publication date were specified precisely, or TRUE if either date was imputed. Imputation was based on the mid-point of the specified date range, for example if a search month is specified rather than a search date, then the 15th day of the month was the imputed search date.
Review parameters	Number of apps reviewed	Number of apps reported for analysis, after any screening or filtering for relevance.
	Apps downloaded	Ordinal variable coded as 'no apps downloaded' (e.g. analysis was based on only app store metadata), 'some apps downloaded' (e.g. a targeted or random sample), or 'all apps downloaded'.
	Apps recommended	A Boolean variable coded as: TRUE if individual apps were named and described in a manner which suggests or recommends their

7 / 20

| | | use, or FALSE otherwise. |

Table 1. Data extracted from each reported study.

Extraction was completed in a two-phase process. In the first phase, we attempted to automatically extract study metadata, app store search and publication history dates. We used a heuristic text matching strategy to locate and excerpt relevant text from study full text, published study metadata and citation database records. Any matched text was used to pre-populate a standardised data extraction form for subsequent review.

In the second phase, each study was reviewed manually to verify automatically identified data and populate coding items not suitable for automation. Only studies which reported at least one search date (i.e. earliest search date) and a publication date were retained for analysis. Each study was reviewed by a pair of reviewers (from JH, KH, CL, ML, JN, KO, MT, IW, QW), each working independently. Any discrepancies were resolved by a third reviewer (from KH, ML and JN) not involved in the original review. Because data extraction included items with non-categorical assignments, we assessed inter-rater agreement using raw agreement (the proportion of scored data items where each reviewer pair assigned the same value). Overall agreement was 0.79 (n=1273/1618 data items) versus 0.83 (n=538/646) for those items extracted in the pilot phase. Considering extraction of study publication dates alone, being the data items intended to inform the primary analysis of study, agreement was 0.84 (n=478/570) versus 0.90 (n=207/230) extracted in the pilot phase.

Data analysis

The primary reported outcome is time to publication (TTP), calculated as the difference in days between the earliest search date and the date of publication. This window was justified on the basis that it reflects a conservative upper bound on the 'staleness' of information contained in any review at the earliest time it becomes accessible to a public audience. A secondary measure, the time to submission (TTS), was calculated as the difference between the earliest search date and the submission date. Descriptive statistics are reported for both TTP and TTS.

Data concerning review parameters (number of apps reviewed, whether apps were downloaded, and whether apps were recommended) are presented descriptively. Due to non-normalcy, non-parametric tests were used to assess the relationship between these parameters and the primary outcome, TTP. The correlation between TTP and the number of apps reviewed was measured using Spearman's correlation coefficient (ρ). The impact of downloading some or all apps on TTP was assessed using a Kruskal-Wallis test, with follow-up tests to identify differences between specific groups. Whether studies which recommended specific apps were published more rapidly was assessed using a Wilcoxon rank sum test.

Two sensitivity analyses were specified *a priori*. We anticipated that studies would exist where date information was reported only partially, for example, reporting only the month and year in which app searches were performed. In these cases (n=93), we imputed the date as the 15th of the stated month

or, if the authors reported a range of actual dates, selected a single date lying in the middle of this range. The first sensitivity analysis assessed the consequences of partial date reporting by comparing TTP for those studies where date imputation was and was not required.

A second sensitivity analysis aimed to explore the consequences of assuming that no important app changes occurred between the earliest and most recent app search/update dates. To do this, TTP was recalculated using the *last available* search date for each study (i.e. the latest reported value of 'earliest search date', 'latest search date', and 'updated search date') and compared to the original TTP measure. For both sensitivity analyses, TTP values were compared using Wilcoxon rank sum tests.

Finally, a *post hoc* exploration of the relationship between the two effort-related review parameters (number of apps reviewed and whether apps were downloaded) was conducted. The number of apps reviewed was compared across the sub-groups based on whether none, some, or all of the apps were downloaded in the review process and tested with a Kruskal-Wallis test. An additional metric, the review time per app, was defined as the time to publication divided by the number of apps reviewed. This metric was again compared across downloaded sub-groups, and compared using a Kruskal-Wallis test. All analyses were conducted using MATLAB 8.6. As this was an analysis of previously published literature, patients or the public were not involved in the design, or conduct, or reporting, or dissemination plans of our research.

Results

Search and selection

Searches performed on 30 April 2019 yielded 439 study reports. After deduplication, screening and full-text review (summarised in Figure 1), 136 reports were included in the final analysis.
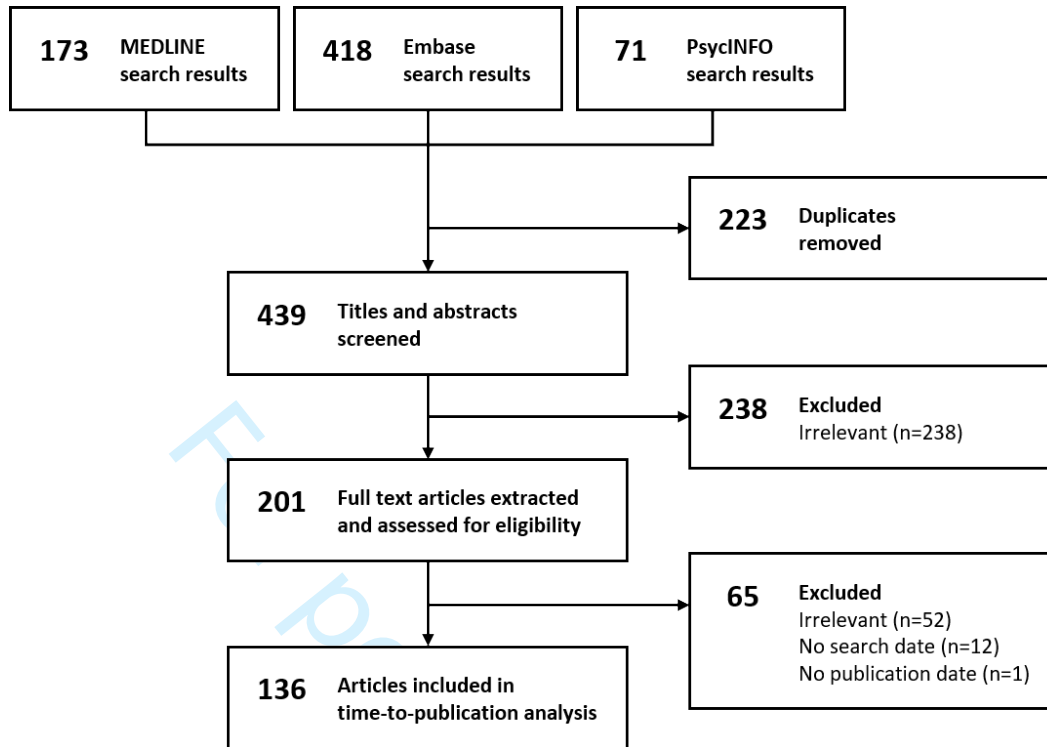
Figure 1. Study selection flowchart.

Time to publication and submission

The median time to publication, TTP, was 431 days (approximately 14 months, range: 42-1054 days) from the earliest search date. 100 papers reported a submission date, however in 8 cases the submission date was prior to the search date, therefore these papers were therefore excluded from the analysis. From the 92 remaining studies, the median time to submission, TTS, was 269 days (approximately 9 months, range: 5-874 days). The distribution of TTP and TTS are shown in Figure 2.
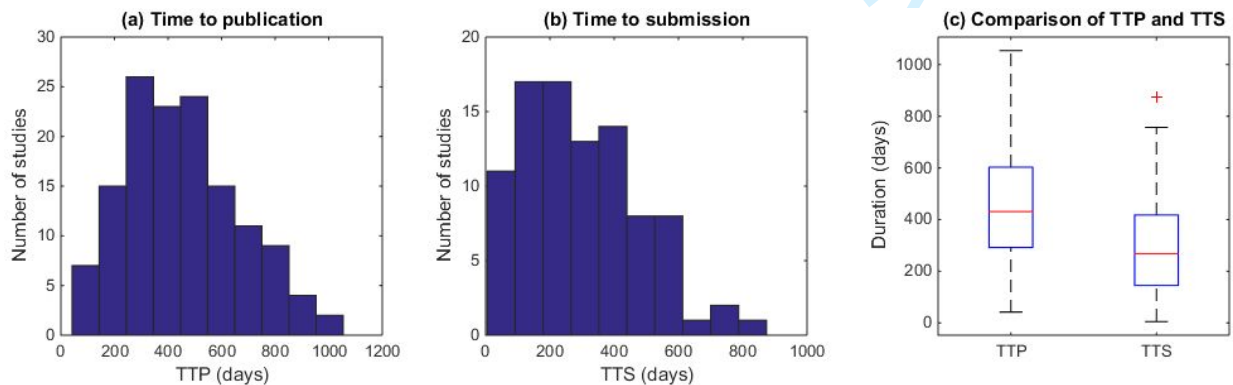

Figure 2. Frequency distributions for (a) the time to publication (TTP), and (b) time to submission (TTS), with (c) a side-by-side comparison.

Review parameters

The median number of apps reviewed in the 136 included studies was 52, although there was large variation between studies (range: 4-1806 apps). A near-zero correlation was found between the number of apps reviewed and TTP (ρ=-0.004, p=0.964).

Authors typically downloaded the apps for review, rather than relying on app store descriptions: 72.1% (n=98/136) papers indicated all apps were downloaded for review vs 17.6% (n=24/136) where no apps were downloaded. A targeted or random sample of apps were downloaded in 9.6% (n=13/136) of studies, and in 1 study it was not possible to determine whether or not apps were downloaded. There was a significant difference in TTP between the subgroups (p=0.010). Figure 3 shows the distribution of TTP for each sub-group. Follow-up tests identified a longer publication time when all apps compared to no apps were downloaded (median TTP: 476 days vs 292 days).

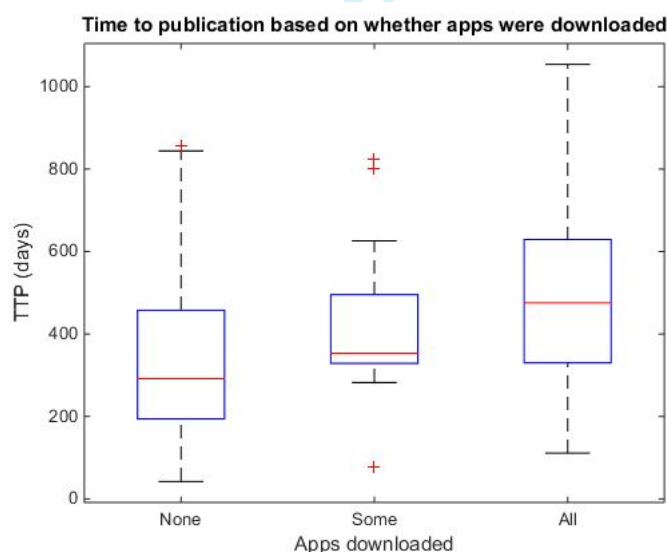Figure 3. Distribution in the time to publication based on whether apps were downloaded as part of the review process.

Specific apps were named and recommended for use in 15.4% (n=21/136) of the reviewed studies. Studies which included recommendations for specific named apps were published marginally more quickly than other studies, but this difference was not significant (median: 425 vs 440 days, p=0.998, see Figure 4.)
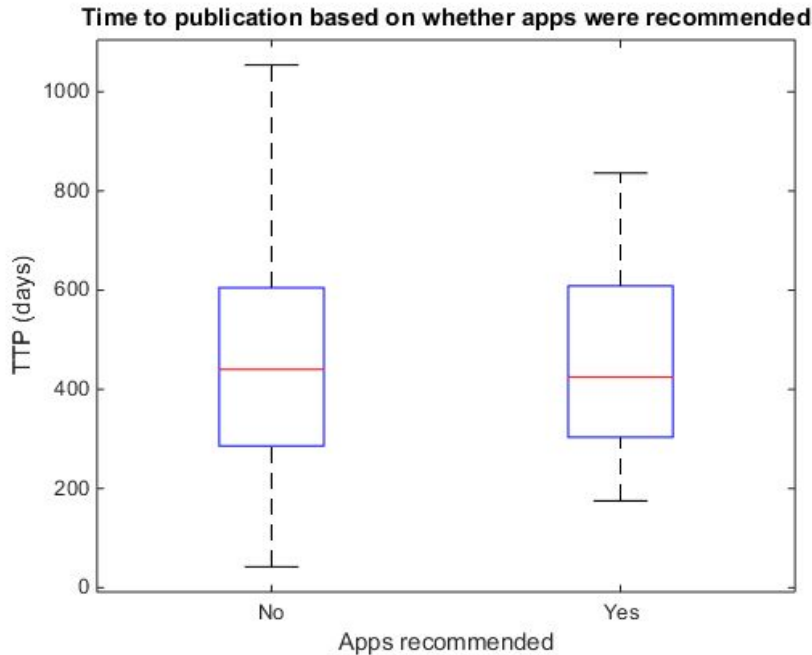
Figure 4. Distribution in the time to publication based on whether specific apps were named and recommended.

Sensitivity analyses

Two-thirds of the studies (92/136, 67.6%) did not specify an exact app store search date, and two (2/136, 1.5%) were not associated with a precise publication date. In combination, dates were imputed in 93 of the studies (68.4%). The difference in time to publication was not significantly different between papers with precise or imputed dates (457 vs 430 days respectively, p=0.648),

Approximately a quarter of the app reviews (37/136, 27.2%) reported a latest search date, and 5.1% (7/136) reported an updated search date. Using the latest of the three reported dates, the median time to publication reduced to 387 days, which did not reach significance for difference from the primary outcome (p=0.063).

Post-hoc analysis

As there appeared to be a relationship between whether apps were downloaded and the time to publication, but no such relationship for the number of apps reviewed in the studies, we conducted a *post-hoc* analysis to examine whether there was a relationship between these two review features. Figure 5 shows the variation in the number of apps reviewed, based on whether apps were downloaded as part of the review process, and a significant difference was found (p<0.001). Follow-up tests identified significantly more apps were reviewed when no or some apps were downloaded, compared with all apps being downloaded (median number of apps: 130, 175, and 43 respectively).
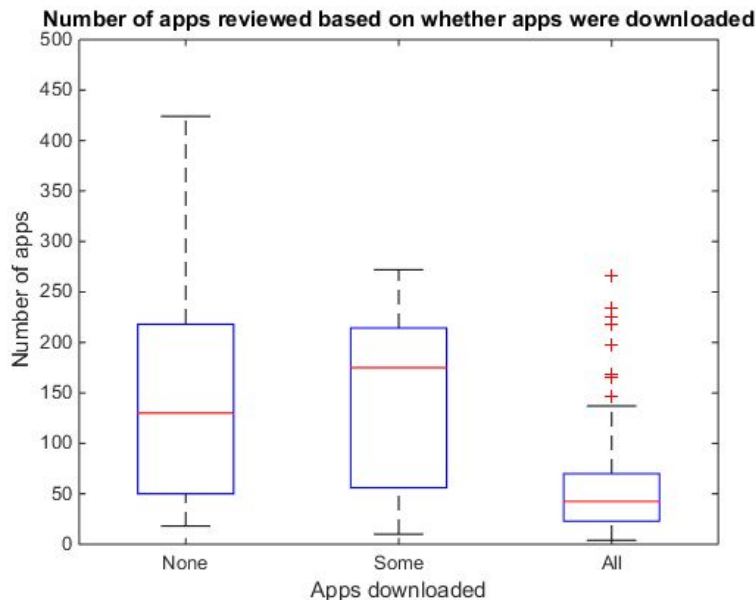
Figure 5. Variation in the number of apps reviewed, depending on whether the apps were downloaded as part of the review process. The y-axis has been truncated, and not all outlier values are shown.

When the time to publication was normalised by the number of apps included in the review, the review time per app was 8.2 days. There were significant differences between studies based on whether apps were downloaded ($p<0.001$) – with those studies where all apps were downloaded taking significantly longer (median 12.1 days/app) compared with the no downloaded (2.9 days/app) and some downloaded (4.2 days/app) apps.

Discussion

This study aimed to quantify the extent to which data presented in cross-sectional app reviews is up-to-date by examining the delay between the selection of apps for review and the time of publication. By the time that most app reviews become available for use, a considerable period of time has elapsed (median: 431 days, 14.2 months). This measure still exceeds one year (387 days, 12.7 months) when a more lenient measure of the recency of findings at the time of publication is used. This delay is not wholly attributable to factors outside the control of researchers, such as the peer review process: when the estimated time to submission was calculated to provide a crude metric of the time to conduct the study, excluding journal peer-review and editing processes, there was still substantial time between app search and manuscript submission (269 days, 8.8 months). The time taken to publish app reviews was influenced by the nature of the analysis, with findings indicating that reviews that downloaded apps for analysis took significantly longer to publish than those that did not. Surprisingly, the number of apps reviewed did not influence publication time, however a post-hoc analysis indicated papers that downloaded apps reviewed significantly fewer apps than those that did not.

Given previous research indicating a high rate of turnover in the app marketplace, with 50% of mental health search results changing within approximately three months,[27] the observed delay in publication raises questions about the validity of study findings at the time they become available to the research, clinical, and broader community, particularly where such reviews focus on recommendations concerning specific products. Reviews may recommend the use of apps which are no longer supported by the developers, have been withdrawn from the app stores or, conversely, have been updated substantially since the review. Contrary to our hypothesis, we did not find that studies making specific recommendations had a shorter time to publication than other reviews. Recommendations for specific apps in published reviews cannot, therefore, automatically be considered reliable. The delay in publication may also mean that more recent, potentially high quality, apps are not made known to the research or clinical communities.

There are a number of potential strategies that could mitigate this issue. The first is for review authors to refresh their results prior to publication. Pre-publication update is a standard practice in systematic reviews. For example, the Cochrane Collaboration will not publish reviews unless the most recent search date is less than twelve months (and ideally less than six months) old.[28] As part of efforts to improve the quality of cross-sectional app reviews,[29] editors and peer reviewers should consider at least asking for justification where there is a long period between search and publication. The practicality of this solution must nevertheless consider review-specific factors that may affect the time to publication, the feasibility of update, and whether the review *intends* to guide clinical and public uses.

The second potential strategy is for authors to adjust the numbers of apps incorporated in their reviews. The relationship between downloading of apps and the number of apps reviewed may indicate that study authors hold some shared perceptions about what represents a 'publishable unit' of work. This may be achieved by either downloading a smaller number of apps, or by reviewing the app store descriptions for a larger number of apps. Both appear to result in publication in approximately one year (± three months). While both have merit, it seems likely that studies that scrutinise app content directly are likely to yield richer insights than those relying on summary information presented in app stores for the purposes of marketing. However, with the longer time to publication associated with downloading, it may be appropriate to focus on a smaller sample of the most popular, most used, or top-ranked apps, which can be published more quickly.

A third possible strategy is to remove app assessment and review from the academic sphere, to organisations whose resources are not subject to the constraints of the academic publication process and are, at least in principle, resourced to be able to respond to app dynamics such as update and withdrawal. Indeed, continuous app scanning and review approaches have now been adopted by a number of health organisations, including the NHS Apps Library[30] and the American Psychiatric Association.[31] However, despite the intention to provide continuous review, it is unclear how often reviews and recommendations that appear within these portals are actually updated. Further, even within app portals, the large number of available apps often necessitates that thoroughness be balanced with expediency,[32] potentially still limiting broad utility of such resources.

Importantly, we do not suggest that our findings imply that cross-sectional assessments of health apps have no utility. App reviews may not be the optimal source of timely information about the function and quality of specific apps, but research-based methods are well suited to identifying and providing unbiased estimates of the nature and extent of thematic issues affecting specific populations of apps. Such insights can and have guided systemic responses to app quality problems. Research-led studies have arguably been important both in identifying *new* issues affecting apps, particularly where identification of issues involves complex exploratory and technical analysis,[33] and in devising systematic strategies for their identification, such as MARS. Unless specifically resourced to do so, it seems unlikely that continuous scanning programmes and app portals will be able to fulfill this discovery function.

Limitations

While this study examined app reviews across all health domains, to characterise the publication delay generally for the mobile health field of research, it did not examine differences across specific sub-domains. It is possible that different outcomes would be observed for different health conditions, for example due to resourcing availability/constraints across clinical domains, or due to differences in self-management approaches for different conditions. These differences were not considered in the current study due to substantial observed heterogeneity in the scope of reviews: some considered only technical domains of app quality (for example, data privacy), some considered broad categories (for example, mental health or physical health), and some considered specific conditions. Furthermore, the databases selected for the literature searches may not have provided complete coverage across all reviews focused on health and wellbeing domains (for example, those reported in allied health publications) or technical domains (for example, those focusing on data privacy and security).

The time to submission was calculated in addition to the primary outcome, time to publication, to provide an approximate measure of the time to conduct the review process. However, this is only a crude estimate as it cannot account for manuscripts submitted to multiple journals prior to acceptance.

Finally, given the research question addressed by this study, we acknowledge the time taken to conduct and publish this review. However, given the relatively slower pace of research compared to app development, these findings from 2019 are likely to still hold relevance.

Conclusions

The majority of health app reviews present data that is at least a year out of data at the time of publication. Given the high rate of observed turnover of health apps in public marketplaces, it may not be appropriate, therefore, for these reviews to be presented as a resource concerning specific products for commissioners, clinicians and the public. Authors of such reviews should, where possible, take steps to minimise the delay to publication, update their results prior to publication and consider whether making specific product recommendations is appropriate. App reviews may nevertheless fulfill important functions to identify novel and thematic issues and guide policy and systemic responses to health app quality and safety. App users should consider alternative sources of information about apps

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

that are better calibrated to the dynamics of the app marketplace, such as continuous scanning services offered by dedicated health app portals.

Statement of contributorship

ML and KH conceived the study and are guarantors. JN and KH devised the search strategies and performed the searches. ML, JN and KH performed initial screening and ML collated the results. KH developed and executed the automatic data extraction process. KH prepared the manual data extraction tools, allocated reviewer tasks and collated results. All authors performed data extraction. ML performed principal data analysis. ML, JN and KH wrote the first draft of the manuscript. All authors provided comments to and/or contributed revisions to the draft manuscript. All authors reviewed and approved the final manuscript version.

Declaration of interests

KH, ML, JN are authors of studies that were included in this review. They declare no other competing financial or non-financial interests. No specific funding was received to undertake this study. All other authors declare no competing financial or non-financial interests.

References

1.	Kitsiou S, Pare G, Jaana M, Gerber B. Effectiveness of mhealth interventions for patients with diabetes: An overview of systematic reviews. *PLoS One* 2017; **12**(3): e0173160.

2.	Coorey GM, Neubeck L, Mulley J, Redfern J. Effectiveness, acceptability and usefulness of mobile applications for cardiovascular disease self-management: Systematic review with meta-synthesis of quantitative and qualitative data. *Eur J Prev Cardiol* 2018; **25**(5): 505-21.

3.	Whitehead L, Seaton P. The effectiveness of self-management mobile phone and tablet apps in long-term condition management: A systematic review. *J Med Internet Res* 2016; **18**(5): e97.

4.	Han M, Lee E. Effectiveness of mobile health application use to improve health behavior changes: A systematic review of randomized controlled trials. *Healthc Inform Res* 2018; **24**(3): 207-26.

5.	Zhao J, Freeman B, Li M. Can mobile phone apps influence people's health behavior change? An evidence review. *J Med Internet Res* 2016; **18**(11): e287.

6.	Byambasuren O, Beller E, Glasziou P. Current knowledge and adoption of mobile health apps among australian general practitioners: Survey study. *JMIR Mhealth Uhealth* 2019; **7**(6): e13199.

7.	Boyle L, Grainger R, Hall RM, Krebs JD. Use of and beliefs about mobile phone apps for diabetes self-management: Surveys of people in a hospital diabetes clinic and diabetes health professionals in new zealand. *JMIR Mhealth Uhealth* 2017; **5**(6): e85.

8.	Rubanovich CK, Mohr DC, Schueller SM. Health app use among individuals with symptoms of depression and anxiety: A survey study with thematic coding. *JMIR Ment Health* 2017; **4**(2): e22.

9.	Australian Government Department of Health and Ageing. E-Mental health strategy for Australia. Canberra, Australia: Commonwealth of Australia, 2012.

10.	California Mental Health Services Authority. "The Technology Suite" Driving access to behavioral health care thru innovation. 2018.

11.	NHS England. The NHS Long Term Plan. London, United Kingdom: HM Government, 2019.

12.	Huckvale K, Adomaviciute S, Prieto JT, Leow MK, Car J. Smartphone apps for calculating insulin dose: a systematic assessment. *BMC Med* 2015; **13**: 106.

13.	Huckvale K, Car M, Morrison C, Car J. Apps for asthma self-management: a systematic assessment of content and tools. *BMC Med* 2012; **10**: 144.

14.	Larsen M, Nicholas J, Christensen H. A Systematic Assessment of Smartphone Tools for Suicide Prevention. *PLoS One* 2016; **11**(4): e0152285.

15.	Nicholas J, Larsen ME, Proudfoot J, Christensen H. Mobile Apps for Bipolar Disorder: A Systematic Review of Features and Content Quality. *J Med Internet Res* 2015; **17**(8): e198.

16.   Donker T, Petrie K, Proudfoot J, Clarke J, Birch MR, Christensen H. Smartphones for smarter delivery of mental health programs: a systematic review. *J Med Internet Res* 2013; **15**(11): e247.

17.   Thornton L, Quinn C, Birrell L, et al. Free smoking cessation mobile apps available in australia: A quality review and content analysis. *Australian and New Zealand Journal of Public Health* 2017; **41**(6): 625-30.

18.   Goodwin GM, Haddad PM, Ferrier IN, et al. Evidence-based guidelines for treating bipolar disorder: Revised third edition recommendations from the British Association for Psychopharmacology. *J Psychopharmacol* 2016; **30**(6): 495-553.

19.   Wyatt JC, Thimbleby H, Rastall P, Hoogewerf J, Wooldridge D, Williams J. What makes a good clinical app? Introducing the RCP Health Informatics Unit checklist. *Clinical Medicine Journal (London)* 2015; **15**(6): 519-21.

20.   Torous JB, Chan SR, Gipson SY-MT, et al. A hierarchical framework for evaluation and informed decision making regarding smartphone apps for clinical care. *Psychiatric Services* 2018; **69**(5): 498-500.

21.   Stephens H, Uccellini M, McKay F. Guidelines for creating healthy living apps. Melbourne, Australia: Dialog Consultking, 2015.

22.   Centers for Disease Control and Prevention. Best practices user guide: Health communications in tobacco prevention and control. Atlanta: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health; 2018.

23.   Rose KJ, Petrut C, L'Heveder R, de Sabata S. IDF Europe's position on mobile applications in diabetes. *Diabetes Res Clin Pract* 2019; **149**: 39-46.

24.   Ronchi E, Woskie LR, Milstein JA. Mobile technology-based services for global health and wellness: opportunities and challenges. Boston, USA: Organisation for Economic Cooperation and Development (OECD), 2017.

25.   World Bank Group. Empowering households and individuals to co-produce positive health outcomes for dignified, person-centered care amidst demographic change Washington, DC, USA: World Bank, 2018.

26.   Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: A new tool for assessing the quality of health mobile apps. *JMIR mHealth and uHealth* 2015; **3**(1): e27-e.

27.   Larsen ME, Nicholas J, Christensen H. Quantifying App Store Dynamics: Longitudinal Tracking of Mental Health Apps. *JMIR Mhealth Uhealth* 2016; **4**(3): e96.

28.   Tariq S, Akhtar N, Afzal H, et al. A novel co-training-based approach for the classification of mental illnesses using social media posts. *IEEE Access* 2019; **7**: 166165-72.

29.   BinDhim NF, Hawkey A, Trevena L. A systematic review of quality assessment methods for smartphone health apps. *Telemed J E Health* 2015; **21**(2): 97-104.

30.   NHS. NHS Apps Library. https://www.nhs.uk/apps-library/ (accessed 25 February 2020).

31.   American Psychiatric Association. App evaluation model.
https://www.psychiatry.org/psychiatrists/practice/mental-health-apps/app-evaluation-model (accessed 25 February 2020).

32.   Meek T. Tom's digital disruptors: testing times for apps. Digital Health; 2015.

33.   Huckvale K, Torous J, Larsen ME. Assessment of the Data Sharing and Privacy Practices of Smartphone Apps for Depression and Smoking Cessation. *JAMA Netw Open* 2019; **2**(4): e192542.

20 / 20

Authors	Title	Journal
O'Loughlin	Reviewing	Internet Interventions
Quevedo R	Mobile pho	Endocrinologia, Diabetes y Nutricion
Mousavi Ja	Top mobile	Malaysian Journal of Medical Sciences
Middelweer	Apps to pro	The International Journal of Behavioral Nutrition and Physical Activity
Singer, Mat	Quality of s	Frontiers in Psychiatry
Bennett, Mc	A review of	Journal of Smoking Cessation
Bardus, Ma	A review an	The International Journal of Behavioral Nutrition and Physical Activity
Jeon E, Pai	Analysis of	Healthcare Informatics Research
Mendiola M	Valuable fe	JMIR MHealth and UHealth
Payne HE,	Health Beh	JMIR Serious Games
Ramo DE,	Cannabis M	JMIR MHealth and UHealth
Mani M, Ka	Review and	JMIR MHealth and UHealth
Hale K, Cap	A Framewo	JMIR MHealth and UHealth
Taki S, Car	Infant Feed	Interactive Journal of Medical Research
Chen J, Ca	The Most F	JMIR MHealth and UHealth
Mangone E	Mobile Pho	JMIR MHealth and UHealth
Con D, De	Mobile Pho	JMIR MHealth and UHealth
Iribarren S,	Smartphon	JMIR MHealth and UHealth
Masterson	Review and	JMIR MHealth and UHealth
Zaidan S, F	Popular Mc	JMIR MHealth and UHealth
Pereira-Aze	Expert Invo	JMIR MHealth and UHealth
Rivera J, M	Mobile App	JMIR MHealth and UHealth
Franco RZ,	Popular Nu	JMIR MHealth and UHealth
Wilson H, S	The Quality	JMIR MHealth and UHealth
Geuens J,	A Review o	JMIR MHealth and UHealth
Huang ET,	Safe Sex M	JMIR MHealth and UHealth
Chen E, Ma	A Systemat	JMIR MHealth and UHealth
Santo K, Ri	Mobile Pho	JMIR MHealth and UHealth
Sullivan RK	Smartphon	JMIR MHealth and UHealth
Grainger R	Apps for Pe	JMIR MHealth and UHealth
Christmann	Stress Man	JMIR MHealth and UHealth
Ginossar T	Content, Us	JMIR MHealth and UHealth
Formagini T	A review of	Cadernos de Saude Publica
Bright T, Pa	Validated S	JMIR Rehabilitation And Assistive Technologies
Cheng F, X	Content An	JMIR MHealth and UHealth
Tinschert P	The Potenti	JMIR MHealth and UHealth
Matava C, I	Mobile App	JMIR medical education
Muthing J,	Client-Focu	JMIR MHealth and UHealth
Lalloo C, S	Commercia	JMIR MHealth and UHealth
Sudol NT, A	In Search o	Female Pelvic Medicine & Reconstructive Surgery
Frie K, Har	Insights Fro	JMIR MHealth and UHealth
Wallace SL	In Search o	Female Pelvic Medicine & Reconstructive Surgery
Metelmann	Medical Co	JMIR MHealth and UHealth
Park JYE, I	Mobile Pho	JMIR MHealth and UHealth
Devan H, F	Evaluation	JMIR MHealth and UHealth
Yang G, Lo	The Charac	JMIR MHealth and UHealth
Abroms L.C	iPhone app	American Journal of Preventive Medicine
Cowan L.T.	Apps of Ste	Health Education and Behavior
Muessig K.	Mobile pho	Journal of medical Internet research
Azar K.M.J	Mobile appl	American Journal of Preventive Medicine
Abroms L.C	A content a	American Journal of Preventive Medicine
Cantudo Cu	Availability	European Journal of Hospital Pharmacy
Haffey F., E	Smartphon	British Journal of Clinical Pharmacology
Stevens D.	Obesity sur	Obesity Surgery
Wallace L.S	A systemat	Journal of Opioid Management

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Cantudo-C A better reg Farmacia Hospitalaria
Choi J., No Smoking ce Journal of medical Internet research
Arnhold M., Mobile app Journal of medical Internet research
Huckvale K Apps for as BMC Medicine
Connor K., Contempor Hernia
Bailey S.C. The availab Journal of the American Medical Informatics Association
Mobasheri Smartphon Breast
Lalloo C., J there's a pa Clinical Journal of Pain
Stevens D. Smartphon Urolithiasis
Wurzer P., Smartphon Burns
van Meche Sports injur British journal of sports medicine
Ouhbi S., F Free Blood Journal of Medical Systems
Kumar N., A content a Journal of the American Society of Hypertension
Schnall R., Review and American Journal of Infection Control
Shelmerdin Smartphon Pediatric Radiology
Huckvale K The evoluti BMC Medicine
Wong M.C. Mobile app Otolaryngology - Head and Neck Surgery (United States)
Lupton D., 'It's like hav Social Science and Medicine
Jutel A., Lu Digitizing d Diagnosis
Al Badi M.F Smartphon International Journal of Pharmaceutical Sciences and Research
Kassianos Smartphon British Journal of Dermatology
Kalz M., Le Smartphon Journal of medical Internet research
Huckvale K Smartphon BMC Medicine
Pereira-Az Health in PLoS ONE
Cheng N.Mi Phone app Telemedicine journal and e-health : the official journal of the American Telemedi
Fijacko N., Mobile App Journal of Medical Systems
Zhang M.W Analysis of Obesity Surgery
Morrissey E Behavior ch American Journal of Preventive Medicine
Thomas G. Threats and Health, Risk and Society
Kim M.S., E Exploring th Journal of the American College of Radiology
Lee H., Sul Smartphon British journal of sports medicine
Perry R., Lu An evaluati Contraception
Coulon S.M A systemat American Journal of Preventive Medicine
Darby A., S A Review o Diabetes Technology and Therapeutics
Crane D., C Behavior ch Journal of medical Internet research
Hoeppner E How Smart Nicotine and Tobacco Research
Ong A.A., C Overview o World Journal of Otorhinolaryngology - Head and Neck Surgery
Robustillo C High quanti Telemedicine journal and e-health : the official journal of the American Telemedi
Nguyen A.D Mobile app International Journal of Medical Informatics
Portelli P., A quality re British Journal of Pain
Brzan P.P., Mobile App Journal of Medical Systems
Penzensta Quality of S European Addiction Research
Gal N., Zite Evaluation Research in social & administrative pharmacy : RSAP
Gibbs J., G 'Can you re Sexually Transmitted Infections
Lunde B., F An Evaluati Women's Health Issues
Calero J.J., Apps for ra Translational Oncology
Collado-Bo Smartphon Farmacia hospitalaria : organo oficial de expresion cientifica de la Sociedad Esp
Radovic A., Smartphon Cyberpsychology, behavior and social networking
Nieto-Gutie Features of Journal of Emergency Medicine, Trauma and Acute Care
Nicholas J. Mobile App Journal of medical Internet research
Sucala M., Anxiety: Th Depression and Anxiety
Schoeppe S Apps to imp International Journal of Behavioral Nutrition and Physical Activity
Tellez W.A. Sunscreen European Research in Telemedicine
Hoeppner E There is an Journal of Substance Abuse Treatment
Meghani S. Clinician-Ta Journal of Palliative Medicine
Edwards E. Gamificatio BMJ open

1
2
3   Shaia K.L., An Evaluati Telemedicine journal and e-health : the official journal of the American Telemedi
4   Ye Q., Kha An Analysis Journal of Diabetes Science and Technology
5   Hoppe C.D An evaluati Journal of human nutrition and dietetics : the official journal of the British Dietetic
6   Househ M., A cross-sec Health informatics journal
7   Thornton L.Free smoki Australian and New Zealand journal of public health
8   Lambert K. Should We Journal of renal nutrition : the official journal of the Council on Renal Nutrition of
9   Leigh S., O Effective? E Evidence-based mental health
10 Zhang M.W Current sta BMJ Open
11 Zhou A.H., Mobile App Annals of Otology, Rhinology and Laryngology
12 Sharpe J.D A systemat Sexual Health
13 Casale M., Mobile appl Laryngoscope Investigative Otolaryngology
14 Jovicic S., Quality eva Clinical Chemistry and Laboratory Medicine
15 Fougerouse A Cross-Se AMIA ... Annual Symposium proceedings. AMIA Symposium
16 Yu J.S., Ku Smartphon Translational behavioral medicine
17 Khander A. Identificatio Telemedicine journal and e-health : the official journal of the American Telemedi
18 Ben-Mussa Popular app Health informatics journal
19 Wisniewski Understand Evidence-based mental health
20 Brown H.M A review of Maternal and Child Nutrition
21 Talwar D., `Characteris European Journal of Human Genetics
22 Owens O.L Systematic American journal of men's health
23 Kwan V., B A Review o The Journal of head trauma rehabilitation
24 Carmody J. Content an Journal of pediatric psychology
25 Grundy Q., Data sharin BMJ (Online)
26 Singh K., D Patients' ar Clinical journal of the American Society of Nephrology : CJASN
27 Luo D., Wa Mobile App Journal of Clinical Rheumatology
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28 icine Association
29
30
31
32
33
34
35
36
37
38
39
40
41 icine Association
42
43
44
45
46
47
48
49
50
51 panola de Farmacia Hospitalaria
52
53
54
55
56
57
58
59
60

1
2
3 icine Association
4
5 Association
6
7
8 the National Kidney Foundation
9
10
11
12
13
14
15
16
17 icine Association
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	2
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	4
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	5
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	N/A
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	6
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	5
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	5
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	6
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	7
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	7
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	N/A
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	8
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	8

PRISMA 2009 Checklist

Page 1 of 2

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	N/A
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	9
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	9
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	S1
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	N/A
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	N/A
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	10
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	N/A
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	12
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	13
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	15
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	15
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	17

For more information, visit: **www.prisma-statement.org**.