Quantity and type of peer-reviewed evidence for popular free medical apps: cross-sectional review

David C Wong[1] (corresponding author), Khine Nwe[2], Ruth Evans[2], Natalie Nelissen[2], Mark E Larsen[3]

1 Centre for Health Informatics, University of Manchester UK

2 Leeds Institute of Health Sciences, University of Leeds, UK

3 Leeds Institute of Data Analytics, University of Leeds, UK

4 Black Dog Institute, UNSW Sydney, Randwick, New South Wales, Australia

Abstract

Introduction – Mobile apps are being increasingly used as a tool to deliver clinical care. Evidence of efficacy for such apps varies, and appropriate levels of evidence may depend on the app's intended use. The UK's National Institute for Health and Care Excellence (NICE) recently developed an evidence standards framework, aiming to explicitly set out the required standards of evidence for different categories of digital health technologies. To determine current compliance with the evidence standards framework, the current study quantified the amount and type of peer-reviewed evidence associated with a cross-section of popular medical apps.

Methods – Apps were identified by selecting the top 100 free medical apps in the Apple App Store and all free apps in the NHS Apps Library. Each app was assigned to one of the four tiers (1, 2, 3a, 3b) in the NICE evidence standards framework. For each app, we conducted searches in Ovid-MEDLINE, Web of Science, Google Scholar, and via manufacturer websites to identify any published articles that assessed the app. This allowed us to determine our primary outcome, whether apps in tiers 3a/3b were more likely than apps in tier 1/2 to be associated with academic peer-reviewed evidence.

Results – We reviewed 125 apps in total (Apple App Store (n=72), NHS Apps Library (n=45), both (n=8), of which 54 were categorized into the higher evidence standards framework tiers, 3a/3b. After screening, we extracted 105 relevant articles which were associated with 25 of the apps. Only 6 articles, pertaining to 3 apps, were reports of randomized controlled trials. Apps in tiers 3a/3b were more likely to be associated with articles than apps in lower tiers ($\chi^2 = 5.54$, $p = .01$). The percentage of tier 3a/3b apps with associated articles was similar for both the NHS Apps Library (10/28) and Apple App store (7/24), ($\chi^2 = 0.042$, $p = .84$).

Discussion – Apps that were in higher tiers 3a and 3b, indicating higher clinical risk, were more likely to have an associated article than those in lower categories. However, even in these tiers, supporting peer-reviewed evidence was missing in the majority of instances. In our sample, Apps from the NHS Apps Library were more no more likely to have supporting evidence than popular Apple App Store apps. This is of concern, given that NHS approval may influence uptake of app usage.

Key Words

Digital Health, Health Apps, Evaluation, mHealth, NHS Apps Library

1. Introduction

The versatility of mobile healthcare (mHealth) apps has meant that, historically, it has been challenging to establish confidence in their clinical quality [1]. More widely, the popularity of healthcare technology is not always associated with strong evidence of the technology's effectiveness. For instance, Cristea et al. recently showed how the highest-valued healthtech start-ups had little supporting academic evidence [2].

Whilst there is robust evidence of effectiveness for some specific mHealth interventions, evidence for the efficacy of mHealth more generally remains limited [3]. One complementary problem is that there is no standard acceptable level of evidence for an mhealth app and indeed, different apps may require different evidence [4]. Some have previously suggested that evidence required to demonstrate safety and effectiveness ought to be proportional to an app's clinical risk and technical complexity [5].

The UK's National Institute for Health and Care Excellence (NICE) recently developed an evidence standards framework, aiming to explicitly set out the required standards of evidence for different categories of digital health technologies (DHTs) [6]. Whilst the framework is designed primarily for DHTs commissioned in the UK health and care system (where professional duty of care is clear) [7], the recommended standards may also be appropriate for direct-to-user applications. The framework has previously been used in this context (e.g by Clarke et al [8]).

Given the new framework, we sought to assess whether a wide selection of popular mHealth apps met the corresponding clinical evidence criteria. Specifically, technologies assigned to the highest tier within the framework (tier 3a or 3b, apps to prevent, detect, manage or treat one or more specific conditions) should, at a minimum, demonstrate effectiveness via 'high quality observational or quasi-experimental studies demonstrating relevant outcomes' [6]. Technologies in this tier are defined as those that are used to 'prevent and manage disease' (3a), or those with 'measurable user benefits, including tools used for treatment and diagnosis, as well as influencing clinical management through active monitoring or calculation' (3b). We reviewed this by assessing the quantity and type of peer-reviewed evidence associated with the identified apps.

2. Methods

This study was a cross-sectional review of a sample of mHealth apps across all health domains.

2.1 Data Collection – Mobile Apps

We selected apps available from the Apple App Store and the NHS Apps Library. The NHS Apps Library [9] provides a list of digital health tools for patients and healthcare professionals so that they can make informed choices for digital health care. The apps registered in the NHS Apps Library have been curated by expert reviewers using criteria outlined in the Digital Assessment Questionnaire [10]. The criteria include 'evidence of outcomes' (clinical, economic or behavioral benefits), clinical safety, data privacy and security, usability and accessibility, interoperability and technical stability. We screened all 70 apps registered in the NHS Apps Library on 27 February 2019.

The App Store [11] is an international distribution platform providing apps for use on Apple devices. We screened the 100 top ranked free apps in the 'medical' category in the UK store from 20-24th February 2019.

Apps were excluded if they were solely intended for educational purposes for healthcare professionals, such as digital versions of medical textbooks, or if they were not free. Finally, interactive websites ('web-apps') listed in the NHS Apps Library were excluded as we did not consider them to be a mobile app. Apps included in both the NHS Apps Library and the Apple App Store were de-duplicated.

For each included app, we manually extracted the app title, screenshots, and developer's description. For apps in the NHS Apps Library, we also recorded their App Store-assigned category (e.g. "Health and Fitness" or "Education"). Apps were not downloaded.

2.2 Data Extraction – Functional Tiers

Using the descriptions and screenshots, a reviewer (KN) classified each app into a functional tier using the NICE Evidence Standards Framework [6]. Abridged descriptions of each tier are shown in Table 1. In the event of uncertainty, two further reviewers (DW,NN) reviewed the app descriptions to reach consensus agreement. The robustness of classification using the framework has been discussed previously [12].

Evidence Tier	Functional Classification	Description
Tier 1	System service	Improves system efficiency. Unlikely to have direct and measurable individual patient outcomes.
Tier 2	Information	Provides information and resources to patients or the public.
	Simple monitoring	Allows users to record health parameters to create health diaries.

	Communicate	Allows 2-way communication between users and professionals, carers, third-party organisations or peers.
Tier 3a	Preventative behaviour change	Designed to change user behaviour related to health issues, for example, smoking, eating, alcohol, sexual health, sleeping and exercise.
	Self-manage	Aims to help people with a diagnosed condition to (better) manage their health.
Tier 3b	Treat	Provides treatment for a diagnosed condition (such as Cognitive Behavioural Therapy for anxiety), or guides treatment decisions.
	Active monitoring	Automatically records information and transmits the data to a professional, carer or third-party organisation, without any input from the user, to inform clinical management decisions.
	Calculate	Performs clinical calculations that are likely to affect clinical care decisions.
	Diagnose	Diagnoses a condition in a patient, or guides a diagnostic decision made by a healthcare professional.

Table 1: Abridged definitions of digital health technology functional tiers from the NICE evidence standards evaluation framework [6].

2.3 Data Collection – Peer-reviewed Evidence search

We conducted a series of searches to determine the quantity of peer-reviewed evidence for each app. Initial scoping searches indicated that relevant literature was often absent in standard medical databases such as MEDLINE. To maximise the probability of retrieving all relevant literature, we searched app developer websites, Google Scholar (limited to the first 50 search results), Ovid-MEDLINE, and Web of Science. In each case, the search term was the full title of the app, as stated on the Apple App Store or NHS Apps Library. All fields were included in the search. Developer websites were extracted from the Apple App Store 'App Support' section and the NHS Apps Library 'Visit Website' link.

Records were de-duplicated, and the primary reviewer (KN), initially screened each abstract. Uncertainties and disagreements were resolved through consultation with a second reviewer (DW or NN). Following this, the corresponding full text articles were examined. Articles were excluded if the full text was unavailable, an app was not evaluated in the article or if the article was not peer-reviewed. Research study protocols, literature reviews, articles solely describing the function of an app, conference abstracts and posters, and journal editorials or commentaries were also excluded.

These exclusion criteria were intended to limit our results to articles that were peer-reviewed assessments of an app. We extracted article identifiers (title of the paper, URL and accessed date), authors’ names, year of publication, and journal name.

Three reviewers (DW, NN, RE) assessed all papers and recorded the study type and study outcome category. In this phase, DW and NN conducted an initial pilot assessment of 11 papers together to confirm the suitability of the categories. The remaining papers were reviewed individually, and any uncertainty over categorisation was resolved by discussion between reviewers.

Study type was coded into five categories. Category A studies focused on evaluating the specific app in a representative group. Category B studies described the use of a specific app in one or possibly a few individual cases, but do not attempt a formal evaluation. In category C, the app was one of many included in a systematic or cross-sectional review, but where the effect of the app was not directly evaluated. In category D, the app was part of a larger, complex intervention and any effects cannot be specifically ascribed to the app. In category E studies, the app was only used as a data collection tool.

For category A or D studies, we recorded whether the study was a randomised controlled trial (RCT), and the category of primary and secondary outcomes. Category (i) outcomes were patient-related, direct clinical outcomes, such as mortality or HbA1c level. Category (ii) comprised indirect clinical outcomes, such as step count or the predictive value of an algorithm. Category (iii) included any non-clinical outcomes such as usability or acceptability.

Study Type	**Description**	**Example**
A - study of app	The app is the main focus of the evaluation.	BlueIce [13]
B - case study/series	Use of the app is described for an individual or small number of individuals.	Oviva [14]
C - review of apps	A qualitative or quantitative assessment of more than one app.	MedScape [15]
D - app as part of complex intervention	The app is one component in a complex intervention.	Microguide [16]
E - app as data collection	The app is used as part of a clinical study as a means of data collection.	SnoreLab [17]

Table 2: categories and definitions of study types used to classify articles relating to apps

2.4 Data Analysis – peer reviewed evidence

Primary Outcome

We provided descriptive statistics of the number of apps categorized to each functional tier, and the proportion of apps within each tier associated with at least one peer-reviewed article. We determined whether apps in tiers 3a and 3b (i.e. tiers that require experimental evidence, according to the framework) were more likely than apps in tiers 1 and 2 to be associated with evidence by comparing proportions.

Secondary Outcomes

In secondary analysis, we provided a descriptive summary of supporting evidence and type of study (categories A-E) and outcome (categories i-iii) for each app. We then determined whether apps in the NHS Apps Library were more likely to have associated evidence than apps within the Apple App Store. We compared the proportion of tier 3a and 3b apps, from both sources (excluding apps duplicated in both), that had at least one peer-reviewed evaluation.

All proportions were compared using a two-sided chi-squared test at α=.05 with Yates's continuity correction in R 4.0.0 [18].

3. Results

App Search

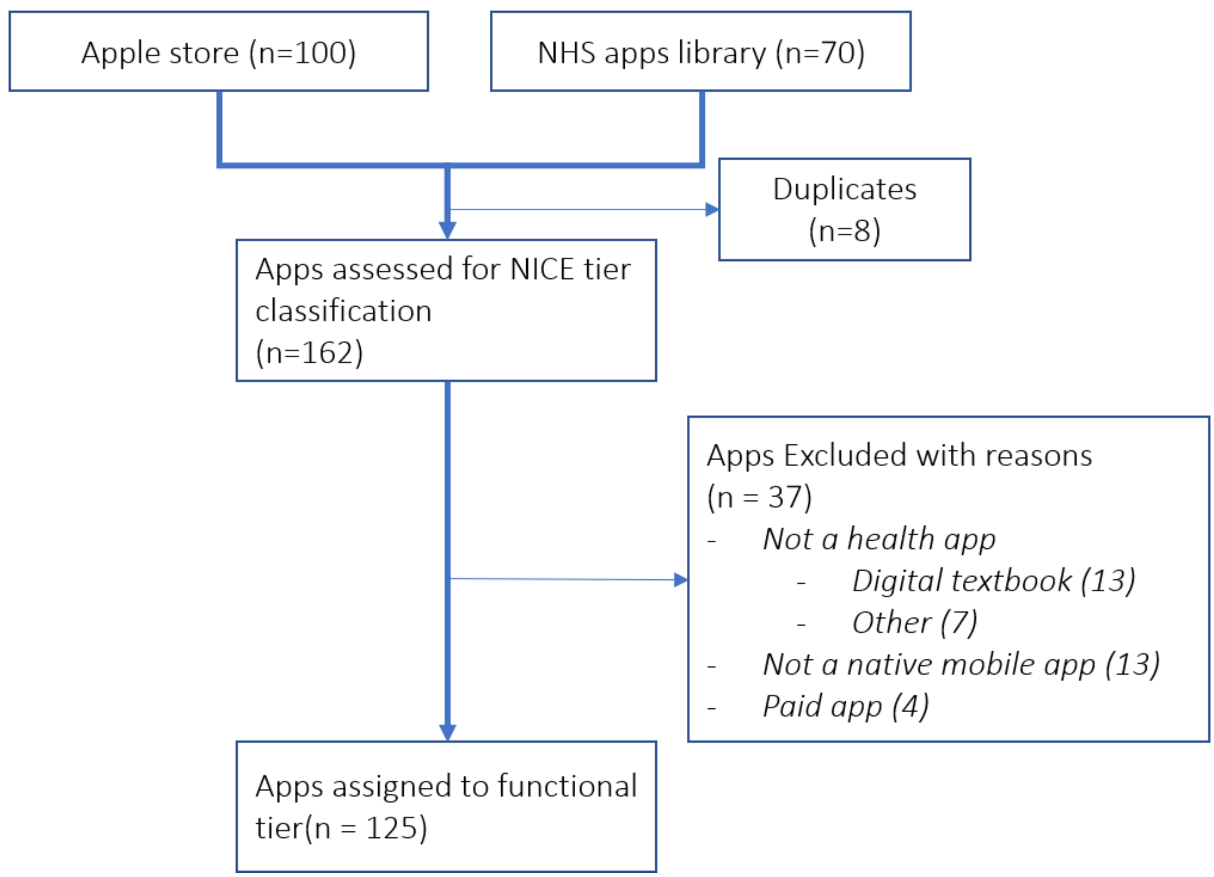

Figure 1: Flow diagram of app search, showing the total number of apps included from the Apple App store and NHS Apps Library

3.1 App Search and Classification

We identified 170 apps in the Apple App Store (n=100) and NHS Apps Library (n=70). Following screening, 125 apps (n=72 from Apple App Store n=45 from NHS Apps Library, n=8 in both) were retained (Figure 1).

Of the 20 apps excluded as being irrelevant, 15 were digital medical textbooks or educational resources. Of the other 5 apps, 2 were Bluetooth controllers for peripheral medical devices, 1 reported reasons for a student's absence from school, 1 located cannabis retailers, and 1 was a photo editor. Four paid apps and 13 web-apps were also excluded.

All included apps from the NHS Apps Library were available in the Apple App Store, but only eight were ranked within the top 100 medical apps. Of the 53 apps available from the NHS Apps Library, 34/53 were not listed in the 'Medical' category and instead classed as 'Health and

Fitness' (n=28), 'Lifestyle' (n=3), 'Education' (n=1), 'Social Networking' (n=1) and 'Games: role-play' (n=1).

Table 3 shows the tier classification of the 125 apps. Four apps were categorized as tier 1 (4/125, 3.2%), and the majority were categorized as tier 2 (67/125, 53.6%). A quarter of the apps were categorized as tier 3a (32/125, 25.6%), and the remaining 22/125 (17.6%) were in tier 3b.

Origin of App	NICE Evidence Standards Framework function tier			
	1	**2**	**3a**	**3b**
Apple App Store	4	40	10	18
NHS Apps Library	0	21	22	2
Both	0	6	0	2
Total	4	67	32	22

Table 3: functional tier classification of all selected apps using the NICE evidence framework

3.2 Evidence Search

We performed a literature search for each of the 125 apps (Figure 2). In total, we screened 7442 records based on title and abstract. After assessment for eligibility, we retrieved 105 peer-reviewed articles. Details about these articles are available in the supplementary material.

Evidence search

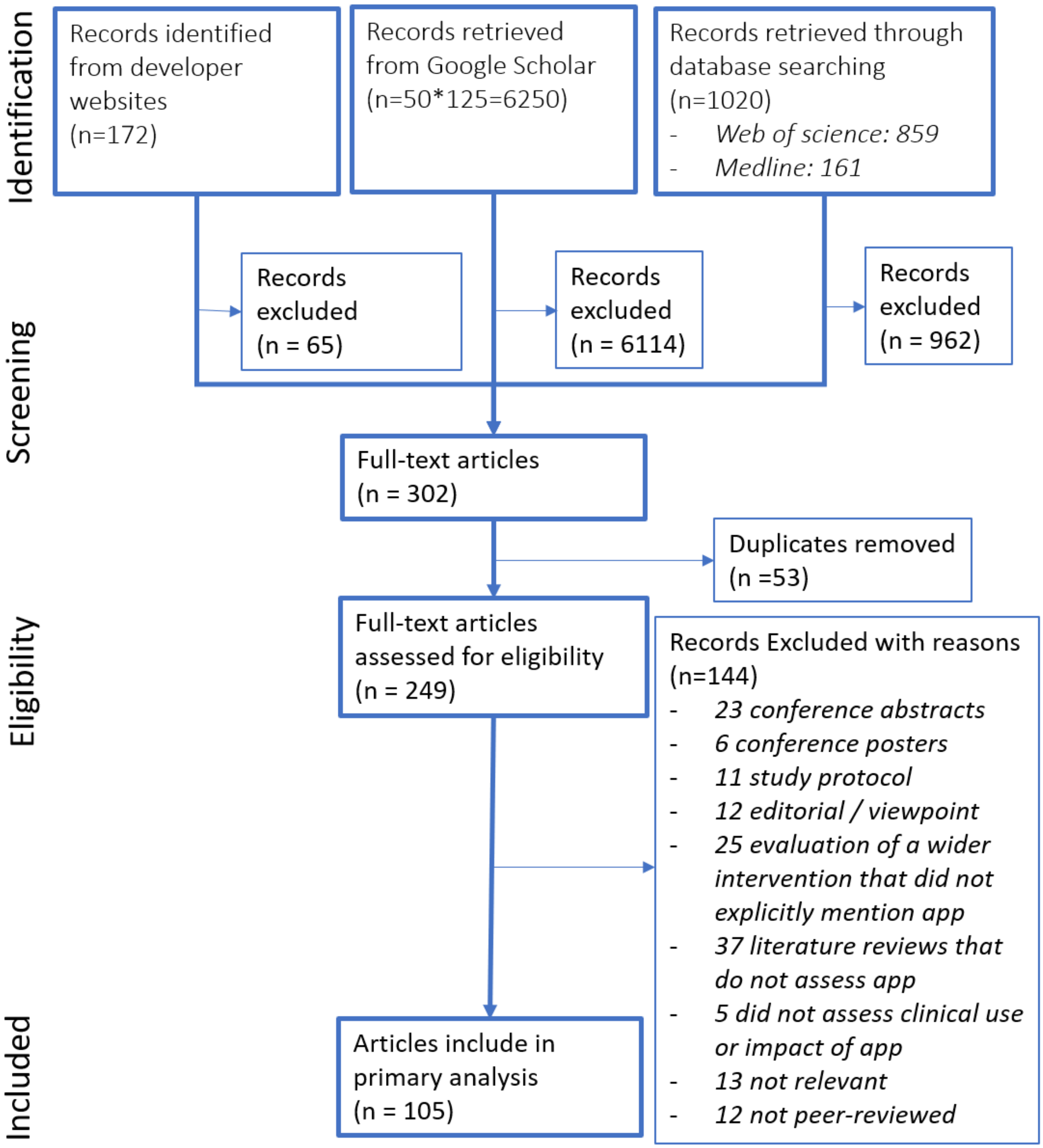

Figure 2: Flowchart of the identification of app evaluation studies.

3.3 Evidence associated with apps

The 105 articles were associated with 25 apps. All articles were published between 2011 and 2019. 57 articles pertained to one app, Kardia. The number of sampled apps with and without associated evidence for each tier are described in Table 4.

	Apps without evidence	Apps with evidence	A - study of app	B – case study	C - review	D – part of intervention	E – data collection tool
Tier 1	4	0	0	0	0	0	0
Tier 2	59	8	6	0	0	0	2
Tier 3a	20	12	11	1	1	0	1
Tier 3b	17	5	4	1	2	2	1
Total	100	25	21	2	3	2	4

Table 4: Evidence and study types associated with apps at each tier. The left part of the table shows the number of apps without and with at least one peer-reviewed publication. The right part shows the number of apps with at least one study of designated category (one app can be associated with multiple study types, and with multiple studies of the same type). Definitions of each study type are described in Table 2.

	App Name	A - study of app	B – case study	C - review	D – part of intervention	E – data collection	Total
Tier 2	Baby+	**1**	0	0	0	0	**1**
	Brush DJ	**1**	0	0	0	0	**1**
	Integrated Family Delivered Neonatal Care	**1**	0	0	0	0	**1**
	iResus	**1**	0	0	0	0	**1**
	MyTherapy Pill Reminder	**2**	0	0	0	0	**2**
	Ovia Pregnancy Tracker	0	0	0	0	**1**	**1**
	Snorelab	0	0	0	0	**1**	**1**
	Squeezy	**1**	0	0	0	0	**1**
Tier 3a	Back Pain Relief – Kaia	**3**	0	0	0	0	**3**
	Blue Ice	**2**	0	0	0	0	**2**
	Catch It	**1**	0	0	0	0	**1**
	Fitbit Plus – Health Coaching	**1**	0	0	0	0	**1**
	GDm-Health	**3**	0	0	0	0	**3**
	Migraine Buddy	0	0	0	0	**2**	**2**
	Moodpath: Depression & Anxiety	**1**	0	0	0	0	**1**
	OurPath	**1**	0	0	0	0	**1**
	Oviva	**1**	**1**	0	0	0	**2**
	Owise breast cancer	**1**	0	0	0	0	**1**
	Thrive	**1**	0	0	0	0	**1**

	RR: Eating Disorder Management	**5**	0	**1**	0	0	**6**
Tier 3b	ECG Check	**1**	0	0	0	0	**1**
	Kardia	**33**	**8**	0	**2**	**14**	**57**
	Medscape	**1**	0	**6**	0	0	**7**
	Microguide	**2**	0	0	**2**	0	**4**
	WebMD	0	0	**3**	0	0	**3**
	Total	**64**	**9**	**10**	**4**	**18**	**105**

Table 5: Study types associated with each app.

Of the 32 tier 3a apps, 12 (38%) had at least one associated peer-reviewed article. The types of study conducted for each app are shown in Table 5. In 11 cases, this included at least one direct study of the app (study type A, Table 4). For the remaining app (Migraine Buddy), the app had been used to collect data (study type E), but the app itself was not directly evaluated.

Of the 22 tier 3b apps, 5 (23%) were associated with peer-reviewed articles. In four cases, this included at least one direct study of the app (study type A). The remaining app (WebMD) had been included in three app reviews (study type C).

In comparison, no tier 1 apps (0%) and eight tier 2 apps (8/67, 12%) had at least one associated peer-reviewed publication. App tiers that required experimental evidence according to the NICE framework (tiers 3a/b) were more likely to have at least one associated study than tiers that did not require such evidence (tiers 1/2) ($\chi^2 = 5.54$, $p = .01$, two-sided).

Six articles (associated with apps: iResus, Thrive, Kaia and Kardia (3 articles)) were reports of RCTs in which the app was solely evaluated (study type A), or part of a complex intervention (study type D). 64 articles (21 apps) reported study results that directly assessed an app (study type A, 61% of articles, 17% of apps). Only 7 of these articles (6 apps) assessed direct clinical outcomes (category i, one as secondary outcome only). A further 39 articles did not measure direct clinical outcomes, but did assess an indirect clinical outcome (category ii, 36 as primary outcome, 3 as secondary outcome). The remaining 18 studies considered only non-clinical outcomes (category iii). Of the 46 (7+39) articles that measured direct or indirect clinical outcomes (category i or ii), 21 also reported non-clinical outcomes.

The remaining 41 articles were: case studies (type B: 9 articles, 2 apps), app reviews (type C: 10 articles, 3 apps), studies in which the app was part of a wider intervention (type D: 4 articles, 2 apps), and studies in which the app was used to collect data (type E: 18 articles, 4 apps).

3.4 NHS Apps Library compliance with evidence standards framework

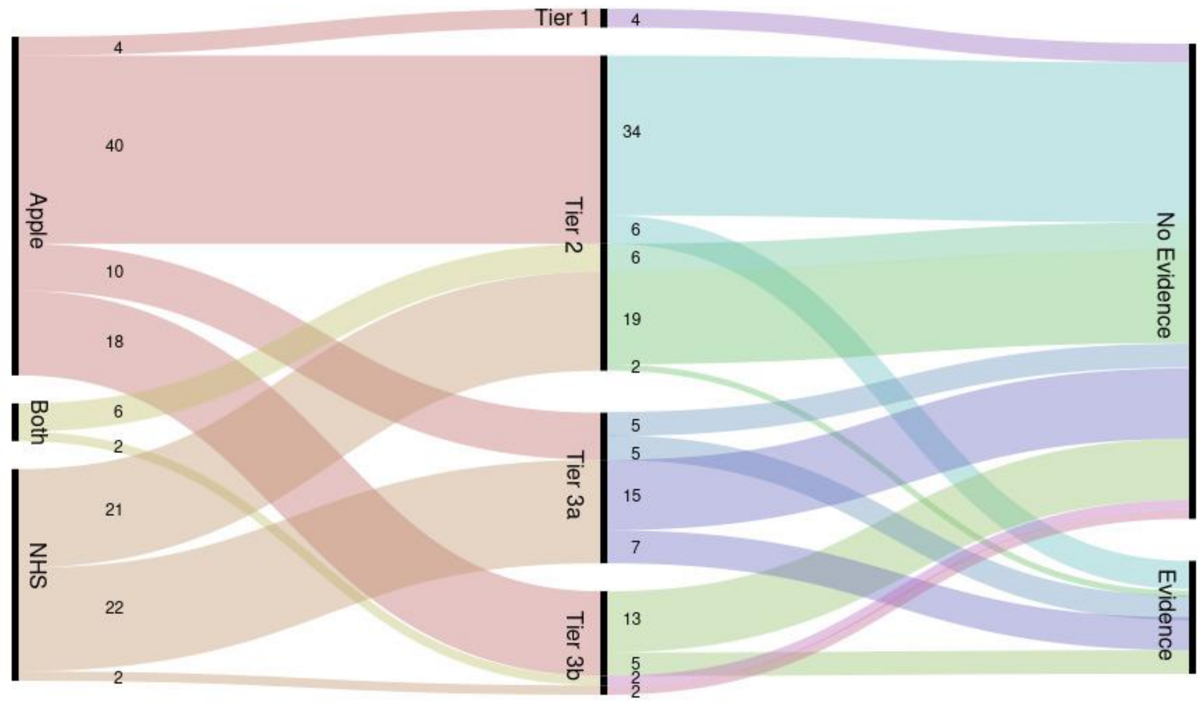

Figure 3: Number of mobile apps, categorized by app store, NICE evidence standards framework functional tier, and presence of peer-reviewed evidence

Figure 3 shows the breakdown of apps by source (Apple App Store/ NHS Apps Library), functional tier, and presence of peer-reviewed evidence. Within the NHS Apps Library, we classified 22 as tier 3a, of which 7 (32%) had any associated peer-reviewed evidence. We classified 2 apps as tier 3b, of which 0 (0%) had any associated evidence. For the Apple App Store, 5/10 (50%) tier 3a apps and 5/18 (28%) tier 3b apps were associated with evidence. Two further tier 3b apps appeared in both the NHS Apps Library and Apple App Store, but neither had associated evidence.

The proportion of tier 3a/3b apps listed on the Apple App Store (10/28) and NHS Apps Library (7/24) that were associated with peer-reviewed literature was similar ($\chi^2 = 0.042$, p = .84, two-sided).

4. Discussion

The NICE evidence standards framework sets the minimum standard of evidence for tier 3a and 3b digital health technologies to demonstrate effectiveness via 'high quality observational or quasi-experimental studies' (interventional for tier 3b) that show improvements in relevant outcomes.

In our cross-sectional review of free popular and NHS-recommended apps, 31% (17/54) of tier 3a/b apps were associated with peer-reviewed evidence. Although we did not assess study quality, this figure is an estimate of the upper-bound on the percentage of apps that meet the minimum standard of evidence in the NICE framework. Only 6% (3/54) of tier 3a/b apps were associated with at least one RCT. This modest proportion of apps indicates the continued absence of evidence to support the effectiveness of medical apps. This is consistent with previous research, which has shown a sparsity of quality evidence to show that mHealth apps improve patient clinical outcomes in numerous domains [19, 20, 21]

Most apps had few (< 5) associated articles (Table 5). However, one app, Kardia, was associated with over half (n=57) of the total number of articles retrieved. We postulate a few reasons for this extreme outlier. First, unlike most other included apps, Kardia requires an external peripheral ECG medical device. Second, the tangible measurements produced, directly relating to clinical outcomes, lend themselves to quantitative study. The peripheral device also makes it a useful tool for data collection (category E), which is how it was used in 14 studies. Finally, the technology is relatively mature – the first published studies from 2013 and 2014 were among the earliest of the papers retrieved.

In subgroup analysis, we showed that apps in NHS Apps Library were no more likely to have supporting evidence than popular Apple App Store apps. Whilst weak association between app quality and app popularity has been shown previously [23], the result is somewhat surprising as the curation process for the NHS Apps Library includes 37 questions related to effectiveness. This is concerning, given recent findings showing that NHS approval was the most important factor for clinicians in deciding whether to prescribe an app [22].

The review process for apps to be listed, and remain, on the NHS Apps Library may not be well-understood by clinicians and end-users, and the quantity and quality of the submitted evidence is not explicitly included in the app listings. Currently, the effectiveness evaluation portion of the Digital Assessment Questions for inclusion on the NHS Apps Library allows developers to cite 'Expert opinion without explicit critical appraisal, or based on physiology, bench research or "first principles"'. Furthermore, there is no requirement that any evidence should be peer-reviewed or publicly available [10].

4.1 Limitations

We conducted a pragmatic, two-stage review of supporting evidence for apps. The first step selected and classified apps, and the second stage searched for and categorized available supporting evidence.

The first step was limited in three aspects. First, apps were solely classified according to the app description and screenshots – no apps were downloaded. This was a deliberate decision, as we chose to evaluate evidence based on the app developers' claimed purpose. The evaluation against intended use mirrors how medical devices are currently regulated. Second, we selected a cross-section of apps. Our *a priori*, pragmatic, choice to select free apps and to exclude the

Google Play app store may introduce selection bias. Given that entry criteria for the Google Play app store is commonly considered to be less stringent than the Apple App Store, we do not think it is likely that this would alter the main result, but an extended review is required to answer this definitively.

Third, in previous work, we showed that there was disagreement between independent raters and that these were often caused by ambiguities in the NICE framework or developer's app description that could not easily be resolved [12]. Disagreement mostly occurred when deciding if an app should be classified as tier 3a vs 3b. As apps in both of these tiers ought to be supported by evidence, there is limited impact on the analysis here.

The evidence searches also had limitations. First, the evidence search step involved a bespoke search for each included app. Articles were sourced from developer websites, Google Scholar and medical databases to maximise recall. As most included studies were sourced via developer websites, the search strategy might be biased in favour of commercial apps, versus those that are developed by academic institutions. Second, supporting evidence was limited to peer-reviewed articles only. For the NICE framework, it is unclear whether internal non-published studies would provide acceptable evidence. We justify both potential limitations by suggesting that the value of scientific evidence depends on its availability, and that it is reasonable to exclude evidence that cannot be found via reasonable effort or is not publicly available.

Finally, we acknowledge that the initial app searches were conducted over 1 year ago, a common issue with app reviews [24]. This was partly due to the time required to undertake the subsequent large number of literature searches corresponding to each app. Despite the churn due to increase in popular telehealth and COVID apps in 2020, 61/70 and 52/100 remain listed on the NHS Apps Library and App Store, respectively (as of 03 Dec 2020) suggesting that results are still applicable.

5. Conclusion

This cross-sectional review of supporting evidence for free apps showed that most apps did not have sufficient evidence to meet the minimum criteria within the NICE Evidence Standards Framework. Furthermore, we found no indication that apps curated within the NHS Apps Library had any more evidence than popular apps within the Apple App Store.

The introduction of the evidence standards framework provides a benchmark that can help to improve the quality of digital health technologies by highlighting areas where there is no supporting good-quality evidence. As this review primarily considers apps published prior to the establishment of the NICE framework, future work should consider whether the framework has had an impact on the quality and quantity of evidence generated for more recent apps.

Author's Contributions

All authors contributed to the design, analysis, and final manuscript writing. KN undertook the initial app and literature searches. DW, KN, ML, NN undertook the app data extraction. DW, NN, RE undertook the literature data extraction.

Acknowledgements

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. KN was supported by a Chevening Scholarship

Conflicts of Interest

DW receives royalties from Sensyne Health. Sensyne Health license the GDm-Health app that was included in this review.

Summary Table

What was already known on the topic:

- evidence for the efficacy of medical mobile apps has historically been very limited.

- the UK's National Institute for Health and Care Excellence (NICE) have developed a new framework to assess the evidence required for digital health technologies.

What this study added to our knowledge:

- Free medical apps are associated with published peer-reviewed literature in a minority of instances, even for apps that may have a strong bearing on direct patient care.

- Medical apps that had been manually curated in the NHS Apps library were not associated with more academic literature than popular free apps.

- Most medical apps sampled in our study would not meet the requirements set out in the NICE digital health technologies evidence standards framework

Word count (excluding abstract + refs : 3090)

References

1. Magrabi F, Habli I, Sujan M, Wong D, Thimbleby H, Baker M, Coiera E. Why is it so difficult to govern mobile apps in healthcare?. BMJ Health and Care Informatics. 2019 Nov 18.
2. Cristea IA, Cahan EM, Ioannidis JP. Stealth research: lack of peer-reviewed evidence from healthcare unicorns. Eur J Clin Invest. 2019 Apr 1;49(4):e13072-8.

3. Marcolino MS, Oliveira JA, D'Agostino M, Ribeiro AL, Alkmim MB, Novillo-Ortiz D. The impact of mHealth interventions: systematic review of systematic reviews. JMIR mHealth and uHealth. 2018;6(1):e23.
4. Greaves F, Joshi I, Campbell M, Roberts S, Patel N, Powell J. What is an appropriate level of evidence for a digital health intervention?. The Lancet. 2018 Dec 22;392(10165):2665-7.
5. Lewis TL, Wyatt JC. mHealth and mobile medical Apps: a framework to assess risk and promote safer use. J Med Internet Res 2014 Sep 15;16(9):e210
6. Evidence Standards Framework for Digital Health Technologies. London: National Institute for Health and Care Excellence; 2019 Mar.
7. Burr, C., & Morley, J. (2020). 'Empowerment or Engagement? Digital Health Technologies for Mental Healthcare'. In C. Burr and S. Milano (Eds.) The 2019 Yearbook of the Digital Ethics Lab, pp. 67–88.
8. Clarke S, Hanna D, Mulholland C, Shannon C, Urquhart C. A systematic review and meta-analysis of digital health technologies effects on psychotic symptoms in adults with psychosis. Psychosis. 2019 Oct 2;11(4):362-73.
9. NHS Digital. NHS Apps Library. London: NHS URL: https://www.nhs.uk/apps-library/ [accessed 2019-10-17]
10. NHS Digital. Digital Assessment Questionnaire V2.2. https://digital.nhs.uk/binaries/content/assets/website-assets/services/nhs-apps-library/daq.pdf
11. Apple Inc. 2020. App Store. [online] [Accessed 8th June 2020]. Available from: https://www.apple.com/uk/ios/app-store/
12. Nwe K, Larsen ME, Nelissen N, Wong DCW. Medical Mobile App Classification Using the National Institute for Health and Care Excellence Evidence Standards Framework for Digital Health Technologies: Interrater Reliability Study. J Med Internet Res 2020;22(6):e17457
13. Stallard P, Porter J, Grist R. A smartphone app (BlueIce) for young people who self-harm: open phase 1 pre-post trial. JMIR mHealth and uHealth. 2018;6(1):e32.
14. Jones L (2018) Diabetes structured education: using technology to transform care. Primary Health Care. doi: 10.7748/phc.2018.e1459
15. Quality Assessment of Medical Apps that Target Medication-Related Problems
16. Fox-Lewis S, Pol S, Miliya T, Day NP, Turner P, Turner C. Utilization of a clinical microbiology service at a Cambodian paediatric hospital and its impact on appropriate antimicrobial prescribing. Journal of Antimicrobial Chemotherapy. 2018 Feb 1;73(2):509-16.
17. Piskin B, Senel B, Uyar A, Karakoc O, Tasci C. Utilization of Sleep Application and Pulse Oximetry in the Short-Term Follow-Up of Mandibular Advancement Device Therapy: A Case Report. Clin Case Rep Int. 2018; 2.;1084.
18. R Core Team (2014). *R: a language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.
19. Wisniewski H, Liu G, Henson P, Vaidyam A, Hajratalli NK, Onnela JP, Torous J. Understanding the quality, effectiveness and attributes of top-rated smartphone health apps. Evidence-based mental health. 2019 Feb 1;22(1):4-9.
20. Rowland SP, Fitzgerald JE, Holme T, Powell J, McGregor A. What is the clinical value of mHealth for patients?. NPJ Digital Medicine. 2020 Jan 13;3(1):1-6.
21. Yerrakalva D, Yerrakalva D, Hajna S, Griffin S. Effects of mobile health app interventions on sedentary time, physical activity, and fitness in older adults: systematic review and meta-analysis. Journal of medical Internet research. 2019;21(11):e14343.
22. Leigh S, Ashall-Payne L, Andrews T. Barriers and Facilitators to the Adoption of Mobile Health Among Health Care Professionals From the United Kingdom: Discrete Choice Experiment. *JMIR Mhealth Uhealth*. 2020;8(7):e17704.

23. Abroms LC, Lee Westmaas J, Bontemps-Jones J, Ramani R, Mellerson J. A content analysis of popular smartphone apps for smoking cessation. Am J Prev Med. 2013;45(6):732–6. https://doi.org/10.1016/j.amepre.2013.07.008.
24. Larsen M, Nicholas J, Han J*, et al* Here's something I prepared earlier: a review of the time to publication of cross-sectional reviews of smartphone health apps. *BMJ Open* 2020;**10:**e039817. doi: 10.1136/bmjopen-2020-03981