BMJ Open

Evidence assessing the diagnostic performance of medical smartphone apps: a systematic review and exploratory meta-analysis

Rahel Buechi,[1] Livia Faes,[2] Lucas M Bachmann,[2] Michael A Thiel,[1] Nicolas S Bodmer,[2] Martin K Schmid,[1] Oliver Job,[1] Kenny R Lienhard[3]

To cite: Buechi R, Faes L, Bachmann LM, *et al*. Evidence assessing the diagnostic performance of medical smartphone apps: a systematic review and exploratory meta-analysis. *BMJ Open* 2017;**7**:e018280. doi:10.1136/bmjopen-2017-018280

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2017-018280).

RB and LF contributed equally.

Received 19 June 2017
Revised 22 October 2017
Accepted 27 October 2017

[1]Eye Clinic, Cantonal Hospital of Lucerne, Lucerne, Switzerland
[2]Medignition Inc., Research Consultants, Zurich, Switzerland
[3]Department of Information Systems, Faculty of Business and Economics, University of Lausanne, Lausanne, Switzerland

Correspondence to
Professor Lucas M Bachmann; bachmann@medignition.ch

ABSTRACT

Objective The number of mobile applications addressing health topics is increasing. Whether these apps underwent scientific evaluation is unclear. We comprehensively assessed papers investigating the diagnostic value of available diagnostic health applications using inbuilt smartphone sensors.

Methods Systematic Review—MEDLINE, Scopus, Web of Science inclusive Medical Informatics and Business Source Premier (by citation of reference) were searched from inception until 15 December 2016. Checking of reference lists of review articles and of included articles complemented electronic searches. We included all studies investigating a health application that used inbuilt sensors of a smartphone for diagnosis of disease. The methodological quality of 11 studies used in an exploratory meta-analysis was assessed with the Quality Assessment of Diagnostic Accuracy Studies 2 tool and the reporting quality with the 'STAndards for the Reporting of Diagnostic accuracy studies' (STARD) statement. Sensitivity and specificity of studies reporting two-by-two tables were calculated and summarised.

Results We screened 3296 references for eligibility. Eleven studies, most of them assessing melanoma screening apps, reported 17 two-by-two tables. Quality assessment revealed high risk of bias in all studies. Included papers studied 1048 subjects (758 with the target conditions and 290 healthy volunteers). Overall, the summary estimate for sensitivity was 0.82 (95 % CI 0.56 to 0.94) and 0.89 (95 %CI 0.70 to 0.97) for specificity.

Conclusions The diagnostic evidence of available health apps on Apple's and Google's app stores is scarce. Consumers and healthcare professionals should be aware of this when using or recommending them.

PROSPERO registration number 42016033049.

Strengths and limitations of this study

- A comprehensive literature search was used to retrieve the published evidence, applying stringent inclusion criteria and assessed the methodological quality of the studies systematically.
- The primary studies found had low methodological quality and level of reporting. All but one of included studies used diagnostic case–control designs.
- The summary estimates from the exploratory meta-analysis need to be interpreted very cautiously.
- We were unable to test all but one of the apps that had been assessed in this review because they were unavailable in the stores, and thus lack first-hand experience.

INTRODUCTION

Within recent years, the number, awareness and popularity of mobile health applications (apps) have increased substantially.[1 2] Currently, over 165 000 apps covering a medical topic are available on the two largest mobile platforms Android and iOS, 9% of them addressing topics of screening, diagnosis and monitoring of various illnesses.[3] Also, the Medical Subject Heading (MeSH) term 'Mobile Applications' that was introduced in MEDLINE in 2014 is currently indexing approximately 1000 records.[4] However, while some authors predicted that mobile health apps will be the game changer of the 21st century, others pointed out that the scientific basis of mobile health apps remains thin.[5 6]

While information used for personal healthcare is traditionally captured via self-report surveys and doctor consultations, mobile devices with embedded sensors offer opportunities to entertain a continued exchange of information between patients and physicians. This dialogue is of particular importance for patients with chronic illnesses.

Three recent reviews focused on the efficacy, effectiveness and usability of mobile health apps in different clinical areas.[7–9] They did not find reasonably sized randomised trials and called for a staged process in the scientific evaluation of mobile health apps. To date, rigorous evidence syntheses of diagnostic studies are missing. In view of the fact that most apps target at a diagnostic

Table 1 Characteristics of included studies

First author's name and year of publication	Target disease	Design	Consecutive enrolment	n	Average age (SD)	% Female	Inclusion criteria	Exclusion criteria
Arora *et al*[21] 2014	Parkinson's disease	Diagnostic case–control study	No	10	65.1 (9.8)	Not reported	Not reported	Other Parkinsonian or tremor disorders
Arora *et al*[11] 2015	Parkinson's disease	Diagnostic case–control study	No	10	65.1 (9.8)	30%	Not reported	Not reported
Chadwick *et al*[12] 2014	Melanoma	Diagnostic case–control study	No	15	Not applicable	Not applicable	Not reported	Not reported
Kostikis *et al*[13] 2015	Parkinson's disease	Diagnostic case–control study	No	23	78	52%	Not reported	Not reported
Lagido *et al*[14] 2014	Atrial fibrillation	Prospective cohort study	No	43	Not reported	Not reported	Not reported	Not reported
Maier *et al*[15] 2015	Melanoma	Diagnostic case–control study	Yes	195	Not applicable	Not applicable	Not reported	Quality images, other elements in the image not belonging to the lesion, for example, hair, images containing more than one lesion, incomplete imaged lesions, non-melanocytic lesions and two-point differences cases.
Ramlakhan *et al*[16] 2011	Melanoma	Diagnostic case–control study	No	46	Not applicable	Not applicable	Not reported	Not reported
Takuya *et al*[20] 2015	Falling in patients with chronic stroke	Diagnostic case–control study	No	11	70.5 (12.5)	Not reported	More than 12 months since stroke onset and ability to walk 16 metres independently with or without a single-point cane and/or an orthosis	Severe cardiovascular, respiratory, musculoskeletal or neurological disorder other than stroke that affected gait performance; unable to understand the instructions because of communication problem or moderate to severe cognitive dysfunction (ie, five or more errors on the SPMSQ; household ambulators walked only indoors or only mobilised during rehabilitation sessions
Wadhawan *et al*[17] 2011	Melanoma	Diagnostic case–control study	No	1300	Not applicable	Not applicable	Not reported	Image artefacts
Wadhawan *et al*[17] 2011	Melanoma	Diagnostic case–control study	No	347	Not applicable	Not applicable	Not reported	Image artefacts
Wolf *et al*[19] 2013	Melanoma	Diagnostic case–control study	No	188	Not applicable	Not applicable	Images for which there was a clear histological diagnosis rendered by a board-certified pathologist	Images containing identifiable features such as facial features, tattoos or labels with patient information. Lesions with equivocal diagnoses such as 'melanoma cannot be ruled out' or 'atypical melanocytic proliferation', Spitz nevi, pigmented spindle cell nevus of Reed and other uncommon or equivocal lesions, lesions with moderate or high-grade atypia poor quality or resolution of images

SPMSQ, Short Portable Mental Status Questionnaire.

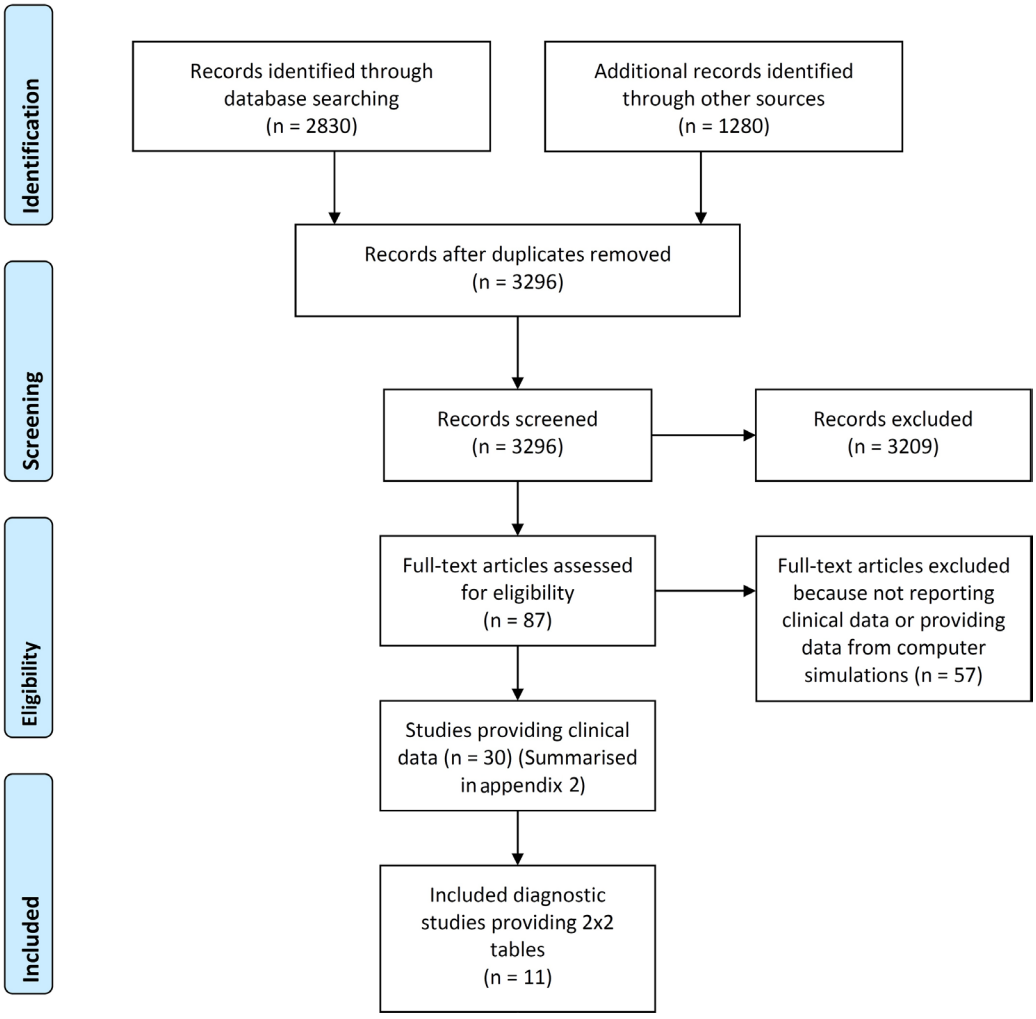

Figure 1 Flow chart according to the PRISMA statement. PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-analyses.

Table 2 Summary of methodological quality assessed with the QUADAS-2[22]

	QUADAS-2: patient selection	QUADAS-2: index test	QUADAS-2: reference standard	QUADAS-2: flow and timing
First author's name and year of publication	Could the selection of patients have introduced bias?	Could the conduct or interpretation of the index test have introduced bias?	Could the reference standard, its conduct, or its interpretation have introduced bias?	Could the patient flow have introduced bias?
Arora *et al*[21] 2014	Yes	Yes	Yes	Yes
Arora *et al*[11] 2015	Yes	Yes	No	Yes
Chadwick *et al*[12] 2014	Yes	Yes	No	Yes
Kostikis *et al*[13] 2015	Yes	Yes	No	Yes
Lagido *et al*[14] 2014	Yes	Yes	Yes	Yes
Maier *et al*[15] 2015	Yes	Yes	No	Yes
Ramlakhan *et al*[16] 2011	Yes	Yes	Yes	Yes
Takuya *et al*[20] 2015	Yes	Yes	Yes	Yes
Wadhawan *et al*[18] 2011	Yes	Yes	No	Yes
Wadhawan *et al*[18] 2011	Yes	Yes	Yes	Yes
Wolf *et al*[19] 2013	Yes	Yes	Yes	Yes

QUADAS, Quality Assessment of Diagnostic Accuracy Studies.

Table 3 Test performance characteristics

First author's name and year of publication	Sensitivity (%)	Specificity (%)	TP	FP	FN	TN	AUC	Application's name	Target disease
Arora *et al*[21] 2014	100.0	100.0	10	0	0	10	Not reported	Not reported	Parkinson's disease
Arora *et al*[11] 2015	100.0	100.0	10	0	0	10	Not reported	Not reported	Parkinson's disease
Chadwick *et al*[12] 2014	0.0	100.0	0	0	5	10	Not reported	Skin Scan	Melanoma
Chadwick *et al*[12] 2014	0.0	100.0	0	0	4	5	Not reported	Mel App	Melanoma
Chadwick *et al*[12] 2014	80.0	20.0	4	8	1	2	Not reported	Mole Detective	Melanoma
Chadwick *et al*[12] 2014	80.0	60.0	4	4	1	6	Not reported	SpotMole Plus	Melanoma
Chadwick *et al*[12] 2014	80.0	60.0	4	4	1	6	Not reported	Dr Mole Premium	Melanoma
Kostikis *et al*[13] 2015	82.6	90.0	19	2	4	18	0.94	Not reported	Parkinson's disease
Lagido *et al*[14] 2014	75.0	97.1	6	1	2	34	Not reported	Not reported	Atrial fibrillation
Maier *et al*[15] 2014	73.1	83.1	19	20	7	98	Not reported	Not reported	Melanoma
Ramlakhan *et al*[16] 2011	91.3	48.6	42	19	4	18	Not reported	Not reported	Melanoma
Takuya *et al*[20] 2015	72.7	84.6	8	2	3	11	0.75	Not reported	Falling in patients with chronic stroke
Wadhawan *et al*[17] 2011	81.1	86.2	30	12	7	75	0.91	Skin scan	Melanoma
Wadhawan *et al*[18] 2011	87.3	71.3	96	68	14	169	Not reported	7-point checklist	Melanoma
Wolf *et al*[19] 2013	70.0	39.3	42	74	18	48	Not reported	Not reported	Melanoma
Wolf *et al*[19] 2013	68.3	36.8	41	79	19	46	Not reported	Not reported	Melanoma
Wolf *et al*[19] 2013	6.7	93.6	4	7	56	103	Not reported	Not reported	Melanoma

AUC, area under the curve; FN, False Negative; FP, False Positive; TN, True Negative; TP, True Positive.

problem, it would be helpful to gauge the scientific basis of them. In this comprehensive systematic review, we thus summarised the currently available papers assessing diagnostic properties of mobile health apps.

METHODS

This review was conducted according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses[10] statement recommendations.

Data sources

Electronic searches were performed without any language restriction on MEDLINE (PubMed interface), Scopus (both databases from inception until 15 December 2016) and Web of Science inclusive Medical Informatics and Business Source Premier (by citation of reference). The full search algorithm is provided in the online supplementary appendix 1.

Study selection

We applied the PICOS format as follows: We included all studies examining subjects in a clinical setting (P) and investigating a health app that used inbuilt sensors of a smartphone (I) for diagnosis of an illness. Minimum requirement to be included in an exploratory meta-analysis was the availability of original data and the possibility to construct a two-by-two table, that is, the possibility to calculate sensitivity and specificity (O). We accepted all reference tests (C) used

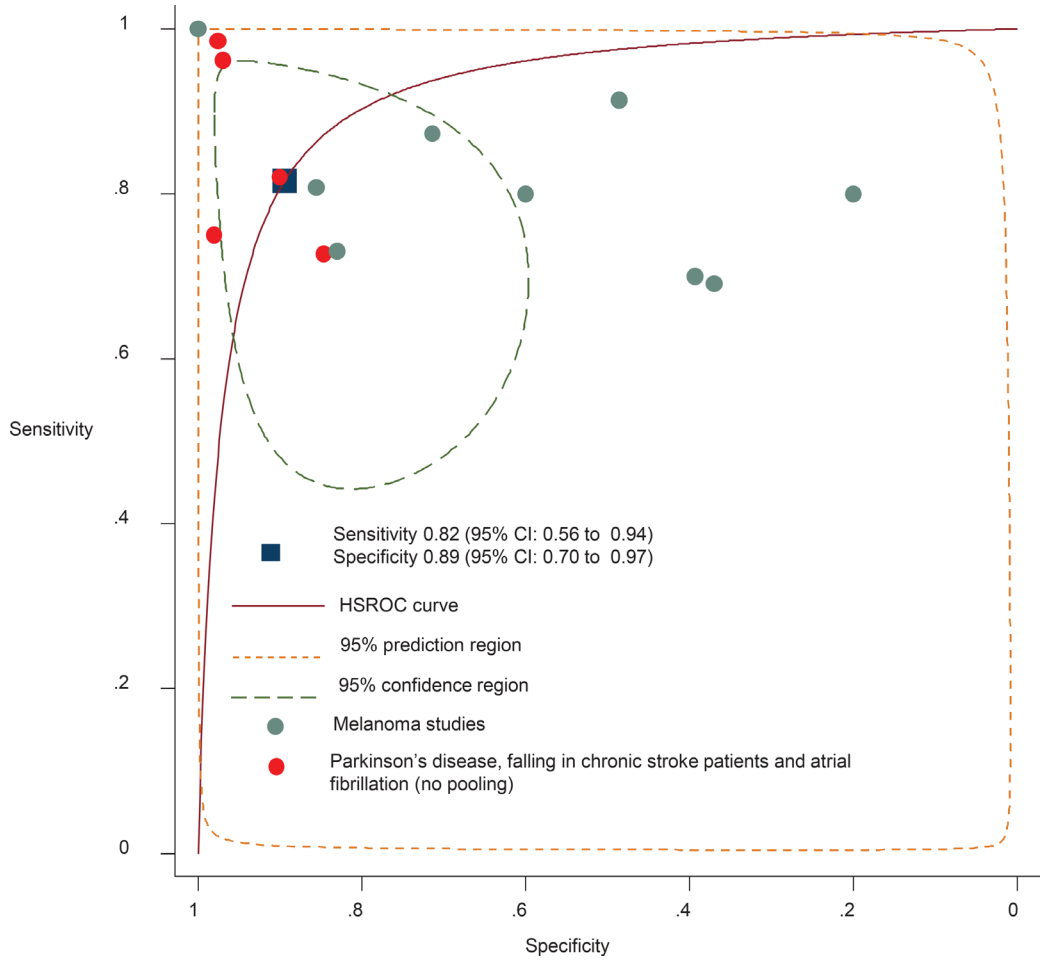

Figure 2 Hierarchical summary receiver operating characteristic curve (HSROC).

in these studies to classify presence or absence of disease. No selection on study design was made (S).

We excluded all studies examining apps providing psychological assessments, questionnaires or mobile alternatives of paper-based tests. We further excluded apps using external sensors, such as clip-on lenses, for the diagnostic assessment or studies, where the app was only used as the transmitter of data.

Data extraction and quality assessment

The methodological quality of all 11 studies[11–21] providing 2×2table data that were summarised in the meta-analysis was made using the Quality Assessment of Diagnostic Accuracy Studies 2 tool. Reporting quality was assessed using the 'STAndards for the Reporting of Diagnostic accuracy studies' (STARD) statement.[22 23] Quality assessment involved scrutinising the methods of data collection (prospective, retrospective) and patient selection (consecutive enrolment, convenience sample) and descriptions of the test (the type of test and analysis performed by the app) and the reference standard (method to rule-in or rule-out the illness).

Two reviewers independently assessed papers and extracted data using a standardised form. Discrepancies were resolved by discussion between the two reviewers by correspondence with study authors or arbitration by a third reviewer. This was necessary in five cases.

Apps of included studies were searched in Apple's App Store and on Google Play.

Data synthesis and analysis

Data to fill the two-by-two table were extracted of each study, and sensitivity and specificity were calculated. Sensitivity and specificity were pooled with the unified method implemented into Stata V.14.2 under the routine 'metandi'. Metandi fits a two-level mixed logistic regression model, with independent binomial distributions for the true positives and true negatives within each study and a bivariate normal model for the logit transforms of sensitivity and specificity between studies. For pooling, at least four studies on the same target condition had to be available.[24] Therefore, no separate analysis for health apps on Parkinson's disease, falling in patients with chronic stroke and atrial fibrillation was possible.

All analyses were done using Stata V.14.2 statistics software package.

RESULTS

Study selection

Electronic searches retrieved 4010 records. After excluding duplicates, 3296 records remained and were screened based on title and abstract. Subsequently, 3209

studies were excluded because they did not fulfil the eligibility criteria. The large majority of records were excluded because they did not contain original data but expressed personal opinion about the possible role of medical smartphone apps. Eighty-seven articles were finally retrieved and read in full text to be considered for inclusion. Out of these, 30 studies provided some clinical data.[2 11–21 25–42] Details on these studies are available in the online supplementary appendix 2. Eleven studies reporting 17 two-by-two tables were considered in this review.[11–21] Details of these studies are available in table 1. The study selection process is outlined in figure 1.

Study characteristics

The 30 papers providing some clinical data 35 diagnostic health apps for various clinical conditions: They included: screening for melanoma (n=8),[12 15–19 27 28] Parkinson's disease monitoring (n=6)[11 21 29 34 35 42] tremor in Parkinson' disease, in multiples sclerosis or of essential tremor (n=4),[13 26 30 39] atrial fibrillation (n=3),[14 31 32] rheumatoid arthritis (n=3),[33 36 41] wet age-related macular degeneration and diabetic retinopathy (n=3),[2 37 38] multiples sclerosis (n=1),[25] cataract (n=1)[40] and falling in patients with stroke (n=1).[20] The studies altogether involved 1048 subjects, 758 subjects with the target condition and 290 healthy volunteers or controls. One paper reported on approximately 3000 skin lesions of an unknown number of patients.[28] The complete data abstraction of these studies is available in the online supplementary appendix 2.

Eleven studies[11–21] that investigated 13 diagnostic health apps allowing the construction of 17 two-by-two tables qualified for the meta-analysis. Twelve tables reported on diagnosis of melanoma, three on Parkinson's disease, one assessed falling in patients with chronic stroke and another atrial fibrillation.

Methodological quality

A summary of the methodological quality is shown in table 2.

Ten studies had a diagnostic case–control design and one was a prospective cohort study.[14] Only in one paper, patients were sampled in a consecutive manner.[15]

A high risk of bias was assessed in all cases. Most high-risk ratings were assigned in domains of 'Patient Selection', 'Index Test' and 'Flow and Timing', whereas fewest high-risk ratings were found within the domain of the 'Reference Standard'. Hence, several sources of bias were identified that may have affected study estimates. Methodological criteria that were frequently inadequately addressed were 'interpretation of reference standard without knowledge of the index test' and vice versa.

USABILITY

Only four studies assessed usability of the investigated diagnostic health app.[2 28 36 37] None used a validated instrument. Questions on usability involved that is, reasons for non-adherence, simplicity of use and difficulties and comprehensibility.

Exploratory analyses of diagnostic accuracy

The summary estimate for sensitivity was 82% (95% CI; 0.56 to 0.94) and pooled specificity was 89% (95% CI 0.70 to 0.97). In a subgroup analysis of 12 reports, pooled sensitivity of studies assessing melanoma was 0.73 (95%CI 0.36 to 0.93) and pooled specificity was 0.84 (95% CI 0.54 to 0.96). No pooling was possible for Parkinson's disease, falling in patients with chronic stroke and atrial fibrillation due to the limited number of studies.

Only one of the apps assessed in this review was available on Apple's or Google's app stores.[12] A summary of test performance characteristics is shown in table 3 and the hierarchical summary receiver operating characteristic curve is seen in figure 2.

DISCUSSION

Main findings

This systematic review of studies assessing the performance of diagnostic health apps using smartphone sensors showed that scientific evidence is scarce. Available studies were small and had low methodological quality. Only one-third of available reports assessed parameters of diagnostic accuracy. Only one app included in the meta-analysis is currently available on app stores. The large majority of health apps available in the stores have not undergone a solid scientific enquiry prior to dissemination.

Results in light of existing literature

To the best of our knowledge, this is the first systematic review assembling the evidence of diagnostic mobile health apps in a broader context. We are aware of one recent paper by Donker and coworkers, who systematically summarised the efficacy of mental health apps for mobile devices.[43] In line with our findings, Donker and colleagues call for further research into evidence-based mental health apps and for a discussion about the regulation of this industry. Other reviews, examining efficacy and effectiveness of mobile health apps support our findings.[7–9] For example, Bakker and colleagues called for randomised controlled trials to validate mental mobile health apps in clinical care.[8] Likewise, Majeed-Ariss and coauthors, who systematically investigated mobile health apps in chronically ill adolescents, pointed at the need of scientific evaluation involving healthcare providers' input at all developmental stages.[7]

Strength and limitations

We conducted a comprehensive literature search to retrieve the published evidence, applied stringent inclusion criteria and assessed the methodological quality of the studies systematically. We applied an overinclusive definition of diagnosis, because for example, symptom monitoring might contribute in the diagnostic work-up of a patient. Out of the papers qualifying for inclusion into

this review, only about 25% investigated the diagnostic accuracy of the app. We believe that a broader concept of diagnosis in this particular context was useful to capture the relevant literature. Our study has several limitations. First, the primary studies were found to have low methodological quality and level of reporting. All but one of included studies used diagnostic case–control designs. While this design might be helpful in early evaluation of diagnostic tests, it usually leads to higher test performance characteristics than could be expected in clinical practice. From that viewpoint, the summary estimates from the exploratory meta-analysis need to be interpreted very cautiously. The searches performed in the electronic databases had low specificity leading to a large number of irrelevant records. Correspondingly, the 'number needed to read' was very high.[44] Although we assessed the records in duplicate by two experienced systematic reviewers, we cannot fully rule out that we missed potentially relevant articles. Finally, we were unable to test all but one of the apps[12] that had been assessed in this review, because they were not available anymore, and thus lack first-hand experience.

Implications for research

Led by the consumer electronics industry, the production of mobile health apps has gained in importance and popularity within recent years. Unfortunately, the scientific work-up of the clinical usefulness of these apps is leaping behind. While many studies have highlighted the potential and possible clinical usefulness of health apps, research conducted according to the well-established standards of design, sampling and analysis are missing. The regulation applied in the USA, the EU and other countries does not go far enough. Ensuring that medical health apps meet criteria on technical concerns is only one important element of regulation. From the consumers or patients' perspective, a trustworthy source showing the amount and level of scientific data underpinning the claims made in the app descriptions would be very useful. In our view, it is very important that technical, clinical and methodological experts jointly form an interdisciplinary development team. While the IT experts take care of the technical developments, data safety and compliance with regulatory requirement, clinical expert certify that the app addresses the right medical context, and researchers finally impose appropriate scientific methods to validly quantify the clinical yield. We believe that developers of a (diagnostic) mobile health app should adopt the same hierarchical framework that has been proposed for imaging testing in the seminal paper of Fryback and Thornbury.[45]

CONCLUSION

In this comprehensive systematic review, we found a lack of scientific evidence quantifying the diagnostic value of health apps in the medical literature. The information about the diagnostic accuracy of currently available health apps on Apple's and Google's app stores is almost absent. Consumers and healthcare professionals should be aware of this when using or recommending them.

Contributors RB, LF, LMB, KRL, NSB and MAT obtained and appraised data. LMB and MKS wrote the paper with considerable input from OJ, MAT, RB and KRL. All coauthors provided intellectual input and approved the final manuscript. LMB was responsible for the design and the statistical analysis of the study and is the study guarantor.

Funding The work presented in this paper was funded by Medignition Inc, a privately owned company in Switzerland providing health technology assessments for the public and private sectors, via an unrestricted research grant.

Competing interests LMB holds shares of Medignition.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement The dataset containing all abstracted data of included studies is available from the Dryad repository: doi:10.5061/dryad.900f8.

REFERENCES

1. Bakker. Enterprise wearable technology case studies: tractica. 2015 https://www.tractica.com/resources/white-papers/enterprise-wearable-technology-case-studies/
2. Kaiser PK. Emerging therapies for neovascular age-related macular degeneration: drugs in the pipeline. *Ophthalmology* 2013;120:S11–15.
3. Institute I. Patient adoption of mhealth. 2013 http://www.imshealth.com/en/thought-leadership/ims-institute/reports/patient-adoption-of-mhealth
4. NCBI. Mobile applications. 2017https://www.ncbi.nlm.nih.gov/mesh/?term=mobile+application
5. Misra S. New report finds more than 165,000 mobile health apps now available, takes close look at characteristics & use. 2015 http://www.imedicalapps.com/2015/09/ims-health-apps-report/#
6. BSaN F. *Healthcare's digital future*: McKiinsey & Company, 2014. http://www.mckinsey.com/industries/healthcare-systems-and-services/our-insights/healthcares-digital-future
7. Majeed-Ariss R, Hall AG, McDonagh J, *et al*. Mobile phone and tablet apps to support young people's management of their physical long-term conditions: a Systematic Review Protocol. *JMIR Res Protoc* 2015;4:e40.
8. Bakker D, Kazantzis N, Rickwood D, *et al*. Mental health smartphone apps: review and evidence-based recommendations for future developments. *JMIR Ment Health* 2016;3:e7.
9. Anderson K, Burford O, Emmerton L. Mobile health apps to facilitate self-care: a qualitative study of user experiences. *PLoS One* 2016;11:e0156164.
10. Moher D, Liberati A, Tetzlaff J, *et al*. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ* 2009;339:b2535.
11. Arora S, Venkataraman V, Zhan A, *et al*. Detecting and monitoring the symptoms of Parkinson's disease using smartphones: A pilot study. *Parkinsonism Relat Disord* 2015;21:650–3.
12. Chadwick X, Loescher LJ, Janda M, *et al*: Mobile medical applications for melanoma risk assessment: false assurance or valuable tool? In: *Proceedings of the Annual Hawaii International Conference on System Sciences*, 2014:2675–84.
13. Kostikis N, Hristu-Varsakelis D, Arnaoutoglou M, *et al*. A smartphone-based tool for assessing parkinsonian hand tremor. *IEEE J Biomed Health Inform* 2015;19:1835–42.
14. Lagido RB, Lobo J, Leite S, Sousa C, Ferreira L, Silva-Cardoso J, *et al*. eds. In: *Using the smartphone camera to monitor heart rate and*

rhythm in heart failure patients: IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI), 2014.
15. Maier T, Kulichova D, Schotten K, *et al*. Accuracy of a smartphone application using fractal image analysis of pigmented moles compared to clinical diagnosis and histological result. *J Eur Acad Dermatol Venereol* 2015;29:663–7.
16. Ramlakhan K, Shang Y, eds.In: *A mobile automated skin lesion classification system*: IEEE 23rd International Conference on Tools with Artificial Intelligence, 2011.
17. Wadhawan T, Situ N, Lancaster K, *et al*. SkinScan©: a portable library for melanoma detection on handheld devices. *Proc IEEE Int Symp Biomed Imaging* 2011;2011:133–6.
18. Wadhawan T, Situ N, Rui H, *et al*. Implementation of the 7-point checklist for melanoma detection on smart handheld devices. *Conf Proc IEEE Eng Med Biol Soc* 2011;2011:3180–3.
19. Wolf JA, Moreau JF, Akilov O, *et al*. Diagnostic inaccuracy of smartphone applications for melanoma detection. *JAMA Dermatol* 2013;149:422–6.
20. Isho T, Tashiro H, Usuda S. Accelerometry-based gait characteristics evaluated using a smartphone and their association with fall risk in people with chronic stroke. *J Stroke Cerebrovasc Dis* 2015;24:1305–11.
21. Arora S, Donohue S VV, Biglan KM, *et al*. *High accuracy discrimination of parkinson's disease participants from healthy controls using smartphones*. editor: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2014, 2014.
22. Whiting PF, Rutjes AW, Westwood ME, *et al*. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;155:529–36.
23. Bossuyt PM, Reitsma JB, Bruns DE, *et al*. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *Ann Clin Biochem* 2003;40(Pt 4):357–63.
24. Harbord RM, Whiting P, Sterne JA, *et al*. An empirical comparison of methods for meta-analysis of diagnostic accuracy showed hierarchical models are necessary. *J Clin Epidemiol* 2008;61:1095–103.
25. Bove R, White CC, Giovannoni G, *et al*. Evaluating more naturalistic outcome measures: a 1-year smartphone study in multiple sclerosis. *Neurol Neuroimmunol Neuroinflamm* 2015;2:e162.
26. Daneault JF, Carignan B, Codère CÉ, *et al*. Using a smart phone as a standalone platform for detection and monitoring of pathological tremors. *Front Hum Neurosci* 2012;6:357.
27. Do T-T ZY, Zheng H, Cheung N-M, Koh D, *et al*. eds. In: *Early melanoma diagnosis with mobile imaging*: 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2014.
28. Doukas C, Stagkopoulos P, Kiranoudis CT, *et al*. Automated skin lesion assessment using mobile technologies and cloud platforms. *Conf Proc IEEE Eng Med Biol Soc* 2012;2012:2444-7.
29. Ellis RJ, Ng YS, Zhu S, *et al*. A validated smartphone-based assessment of gait and gait variability in parkinson's disease. *PLoS One* 2015;10:e0141694.
30. Kostikis N, Hristu-Varsakelis D, Arnaoutoglou M, Kotsavasiloglou C, *et al*. eds. In: *IEEE Smartphone-based evaluation of parkinsonian hand tremor: Quantitative measurements vs clinical assessment scores*: 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2014.
31. Lee J, Reyes BA, McManus DD, *et al*. Atrial fibrillation detection using an iPhone 4S. *IEEE Trans Biomed Eng* 2013;60:203–6.
32. McManus DD, Lee J, Maitas O, *et al*. A novel application for the detection of an irregular pulse using an iPhone 4S in patients with atrial fibrillation. *Heart Rhythm* 2013;10:315–9.
33. Nishiguchi S, Ito H, Yamada M, *et al*. Self-assessment tool of disease activity of rheumatoid arthritis by using a smartphone application. *Telemed J E Health* 2014;20:235–40.
34. Printy BP, Renken LM, Herrmann JP, *et al*. eds. In: *IEEESmartphone application for classification of motor impairment severity in parkinson's disease*: 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2014.
35. Raknim P, Lan KC. Gait monitoring for early neurological disorder detection using sensors in a smartphone: Validation and a case study of parkinsonism. *Telemed J E Health* 2016;22:75–81.
36. Shinohara A, Ito T, Ura T, *et al*. eds. In: *Development of lifelog sharing system for rheumatoid arthritis patients using smartphone*: Conference proceedings: annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2013.
37. Wang YZ, He YG, Mitzel G, *et al*. Handheld shape discrimination hyperacuity test on a mobile device for remote monitoring of visual function in maculopathy. *Invest Ophthalmol Vis Sci* 2013;54:5497.
38. Winther C, Frisén L. Self-testing of vision in age-related Macula degeneration: a longitudinal pilot study using a smartphone-based rarebit test. *J Ophthalmol* 2015;2015:1–7.
39. Woods AM, Nowostawski M, Franz EA, *et al*. Parkinson's disease and essential tremor classification on mobile device. *Pervasive Mob Comput* 2014;13:1–12.
40. Fuadah YN, Setiawan AW, Mengko TL. *Mobile cataract detection using optimal combination of statistical texture analysis*: 4th International Conference on Instrumentation, Communications, Information Technology, and Biomedical Engineering (ICICI-BME) IEEE, 2015.
41. Yamada M, Aoyama T, Mori S, *et al*. Objective assessment of abnormal gait in patients with rheumatoid arthritis using a smartphone. *Rheumatol Int* 2012;32:3869–74.
42. Zhu S, Ellis RJ, Schlaug G, *et al*. eds. In: *Validating an iOS-based rhythmic auditory cueing evaluation (iRACE) for Parkinson's disease. Proceedings of the 22nd ACM international conference on multimedia*: ACM, 2014.
43. Donker T, Petrie K, Proudfoot J, *et al*. Smartphones for smarter delivery of mental health programs: a systematic review. *J Med Internet Res* 2013;15:e247.
44. Bachmann LM, Coray R, Estermann P, *et al*. Identifying diagnostic studies in MEDLINE: reducing the number needed to read. *J Am Med Inform Assoc* 2002;9:653–8.
45. Fryback DG, Thornbury JR. The efficacy of diagnostic imaging. *Med Decis Making* 1991;11:88–94.