BMJ Open

Evidence assessing the diagnostic performance of medical smartphone apps - A Systematic Review and exploratory meta-analysis

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-018280
Article Type:	Research
Date Submitted by the Author:	19-Jun-2017
Complete List of Authors:	Buechi, Rahel; Cantonal Hospital Lucerne, Eye Clinic Faes, Livia; medignition Research Consultants, Bachmann, Lucas; medignition Research Consultants Thiel, Michael; Eye Clinic, Cantonal Hospital of Lucerne, Bodmer, Nicolas; medignition Research Consultants Schmid, Martin; Eye Clinic, Cantonal Hospital of Lucerne, Job, Oliver; Eye Clinic, Cantonal Hospital of Lucerne, Lienhard, Kenny; Faculty of Business and Economics, University of Lausanne, Information Systems
Primary Subject Heading:	Evidence based practice
Secondary Subject Heading:	Health informatics, Health services research
Keywords:	mobile health apps, evidence-based medicine, systematic review, diagnostic research

Note: The following files were submitted by the author for peer review, but cannot be converted to PDF. You must view these files (e.g. movies) online.

Appendix_studies_description.xls

Evidence assessing the diagnostic performance of medical smartphone apps - A Systematic Review and exploratory meta-analysis

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

medignition Inc, Research Consultants, 8004 Zurich, Switzerland, Rahel Büchi research fellow, Livia Faes, research fellow, Lucas M Bachmann, professor of epidemiology, director, Nicolas S Bodmer Research Fellow, Eye Clinic, Cantonal Hospital of Lucerne Michael A Thiel, professor of ophthalmology, director, Martin K Schmid co-director, Oliver Job, head neuro-ophthalmology, Department of Information Systems, Faculty of Business and Economics, University of Lausanne Kenny R Lienhard, research fellow

Correspondence to: Lucas M. Bachmann: bachmann@medignition.ch

Rahel Buechi and Livia Faes contributed equally.

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 in-built 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 December 15th, 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 in-built sensors of a smartphone for diagnosis of disease. The methodological quality of 11 studies used in an exploratory meta-analysis was assessed with the QUADAS-2 tool and the STARD statement. Sensitivity and specificity of studies reporting two-by-two tables were calculated and summarized.

Results We screened 3'296 references for eligibility and included 30 papers investigating 35 diagnostic health applications into the narrative analysis. Quality assessment revealed high risk of bias in all studies. Included papers studied 1'048 subjects (758 with the target conditions and 290 healthy volunteers). Eleven studies, most of them assessing melanoma screening apps, reported 17 two-by-two tables. Overall, the summary estimate for sensitivity was 0.82 (95 % confidence interval (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.

This systematic review was prospectively registered at PROSPERO under the number 42016033049.

Strength and Limitations of this Study

Strength

- A comprehensive literature search to retrieve the published evidence, applying stringent inclusion criteria and assessed the methodological quality of the studies systematically.

Limitations

- 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 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, nine percent 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 health care 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 dialog is of particular importance for patients with chronical 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 fact that most apps target at a diagnostic problem, it would be helpful to gauge the scientific basis of them. In this comprehensive systematic review we thus summarized the currently available papers assessing diagnostic properties of mobile health apps.

Methods

This review was conducted according to the PRISMA(10) statement recommendations.

Data Sources

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

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 in-built 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, i.e. the possibility to calculate sensitivity and specificity (O). We accepted all reference tests (C) used 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) summarized in the meta-analysis was made using the QUADAS-2 tool. Reporting quality was assessed using the STARD statement (22, 23). Quality assessment involved scrutinizing 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 standardized 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. Two-by-two tables consisted of true-positive (TP), false-positive (FP), false-negative (FN), and true-negative (TN) results. For the analysis, we called a result a true positive if the index test finding was in agreement with the reference standard findings. We calculated sensitivity as TP/(TP+FN) and specificity as TN/(FP+TN). Sensitivity and specificity were pooled with the unified method implemented into Stata 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(24) pooling, at least 4 studies on the same target condition had to be available. Therefore, no separate analysis for health apps on Parkinson's disease, falling in chronic stroke patients and atrial fibrillation was possible.

All analyses were done using Stata 14.1 statistics software package (StataCorp LP,College Station, TX, USA).

Role of the Funding Source

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 sector, via an unrestricted research grant. LMB holds shares of medignition. LMB was responsible for the design and the statistical analysis of the study.

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, thirty studies fulfilled the inclusion criteria for the systematic review (2, 11-21, 25-42). Details on these studies are available in the **Appendix**. A subset of eleven studies reported seventeen two-by-two tables and allowed calculating sensitivity and specificity (11-21). Details of these studies are available in **Table 1**. The study selection process is outlined in **Figure 1**.

Figure 1. Flow Chart According to the PRISMA Statement

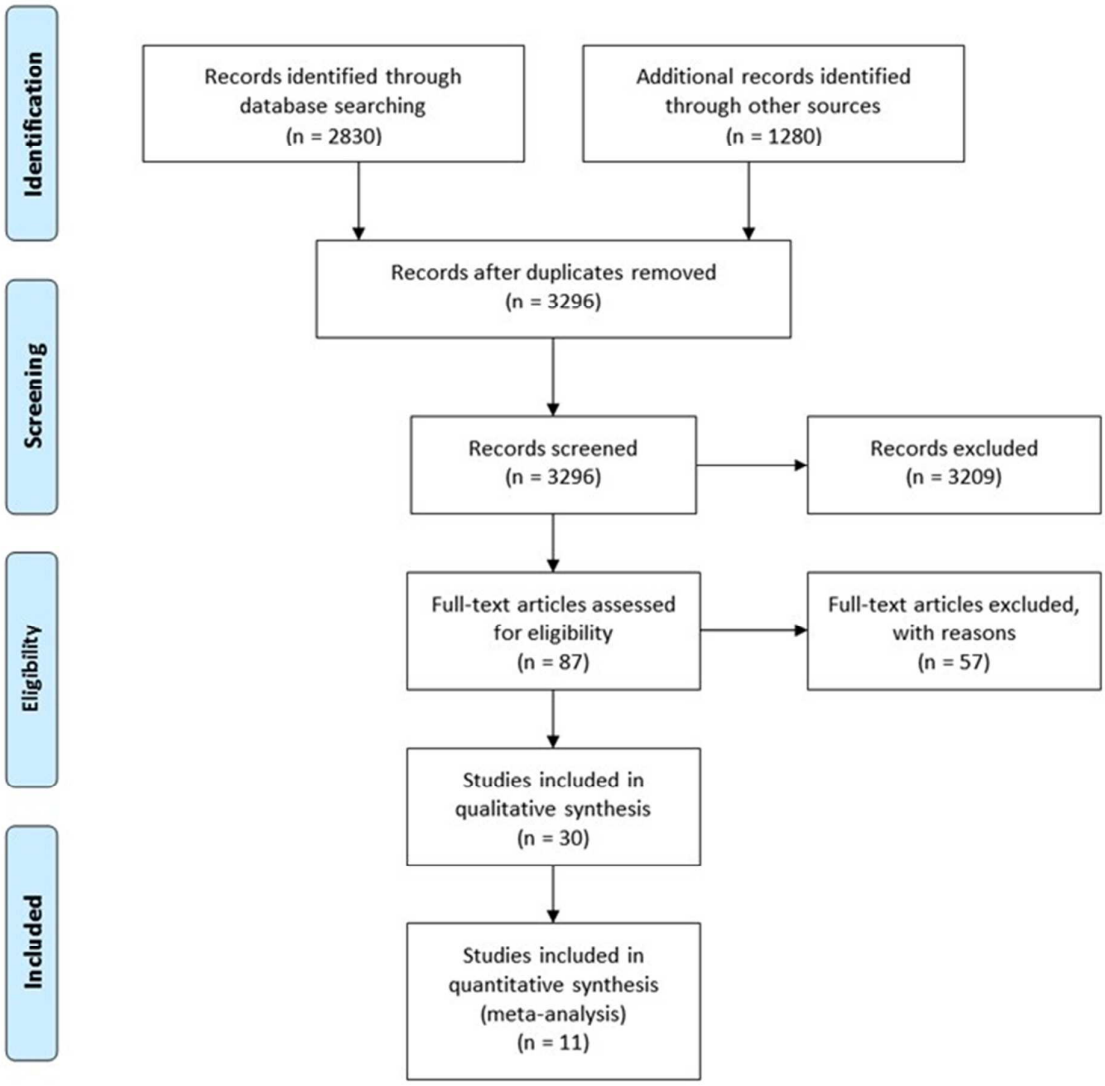


Table 1: Characteristics of included studies. The elven studies allowed constructing 17 2x2 tables.

First Author's Name and Year of Publication	Target Disease	Design	Consecutive Enrolment	n	Average Age (SD)	Female
Arora et al 2014	Parkinson's Disease	Diagnostic Case Control Study	No	10	65.1 (9.8)	Not reported
Arora et al 2015	Parkinson's Disease	Diagnostic Case Control Study	No	10	65.1 (9.8)	30%
Chadwick et al 2014	Melanoma	Diagnostic Case Control Study	No	15	Not applicable	Not applicable
Chadwick et al 2014	Melanoma	Diagnostic Case Control Study	No	15	Not applicable	Not applicable
Chadwick et al.2014	Melanoma	Diagnostic Case Control Study	No	15	Not applicable.	Not applicable
Chadwick et al 2014	Melanoma	Diagnostic Case Control Study	No	15	Not applicable	Not applicable
Chadwick et al 2014	Melanoma	Diagnostic Case Control Study	No	15	Not applicable	Not applicable
Kostikis et al 2015	Parkinson's Disease	Diagnostic Case Control Study	No	23	78	52%
Lagido et al 2014	Atrial Fibrillation	Prospective Cohort Study	No	43	Not reported	Not reported
Maier et al 2014	Melanoma	Diagnostic Case Control Study	Yes	195	Not applicable	Not applicable
Ramlakhan et al 2011	Melanoma	Diagnostic Case Control Study	No	46	Not applicable	Not applicable
Takuya et al 2015	Falling in Chronic Stroke Patients	Diagnostic Case Control Study	No	11	70.5 (12.5)	Not reported
Wadhawan et al 2011	Melanoma	Diagnostic Case Control Study	No	1300	Not applicable	Not applicable
Wadhawan et al 2011	Melanoma	Diagnostic Case Control Study	No	347	Not applicable	Not applicable
Wolf et al 2013	Melanoma	Diagnostic Case Control Study	No	188	Not applicable	Not applicable
Wolf et al 2013	Melanoma	Diagnostic Case Control Study	No	188	Not applicable	Not applicable
Wolf et al 2013	Melanoma	Diagnostic Case Control Study	No	188	Not applicable	Not applicable

Study characteristics

The 30 included papers investigated 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 stroke patients (n=1)(20). The studies altogether involved 1'048 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 **appendix**.

Eleven studies (11-21) that investigated 13 diagnostic health apps allowing the construction of 17 two-by-two tables qualified for the meta-analysis. 12 tables reported on diagnosis of melanoma, three on Parkinson's disease, one assessed falling in chronic stroke patients, and another atrial fibrillation. Ten studies had a diagnostic case-control design and one studies was a prospective cohort study(14). Only in one paper, patients were sampled in a consecutive manner (15). A summary of inclusion and exclusion criteria within individual studies is shown in **Table 2**

.

Table 2. Inclusion and Exclusion Criteria

First Author's Name and Year of Publication	Inclusion Criteria	Exclusion Criteria
Arora et al 2014	Not reported.	Other parkinsonian or tremor disorders.
Arora et al 2015	Not reported.	Not reported.
Chadwick et al 2014	Not reported.	Not reported.
Kostikis et al 2015	Not reported.	Not reported.
Lagido et al 2014	Not reported.	Not reported.
Maier et al 2014	Not reported.	Quality images, other elements in the image not belonging to the lesion e.g. hair, images containing more than one lesion, incomplete imaged lesions, non-melanocytic lesions, two-point differences cases.
Ramlakhan et al 2011	Not reported.	Not reported.
Takuya et al 2015	More than 12 months since stroke onset and ability to walk 16 meters independently with or without a single-point cane and/or an orthosis.	Severe cardiovascular, respiratory, musculoskeletal, or neurologic disorder other than stroke that affected gait performance; unable to understand the instructions because of communication problem or moderate to severe cognitive dysfunction (i.e., 5 or more errors on the Short Portable Mental Status Questionnaire [SPMSQ]); household ambulators walked only indoors or only mobilized during rehabilitation sessions.
Wadhawan et al 2011	Not reported.	Image artifacts.
Wolf et al 2013	Images for which there was a clear histologic diagnosis rendered by a board-certified dermato-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.

Methodological Quality

A summary of the methodological quality is shown in **Table 3**

Table 3 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 2014	*Yes*	*Yes*	*Yes*	*Yes*
Arora et al 2015	*Yes*	*Yes*	*No*	*Yes*
Chadwick et al 2014	*Yes*	*Yes*	*No*	*Yes*
Kostikis et al 2015	*Yes*	*Yes*	*No*	*Yes*
Lagido et al 2014	*Yes*	*Yes*	*Yes*	*Yes*
Maier et al 2014	*Yes*	*Yes*	*No*	*Yes*
Ramlakhan et al 2011	*Yes*	*Yes*	*Yes*	*Yes*
Takuya et al 2015	*Yes*	*Yes*	*Yes*	*Yes*
Wadhawan et al 2011	*Yes*	*Yes*	*No*	*Yes*
Wadhawan et al 2011	*Yes*	*Yes*	*Yes*	*Yes*
Wolf et al 2013	*Yes*	*Yes*	*Yes*	*Yes*

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 i.e. reasons for non-adherence, simplicity of use and difficulties and comprehensibility.

Exploratory analyses of diagnostic accuracy

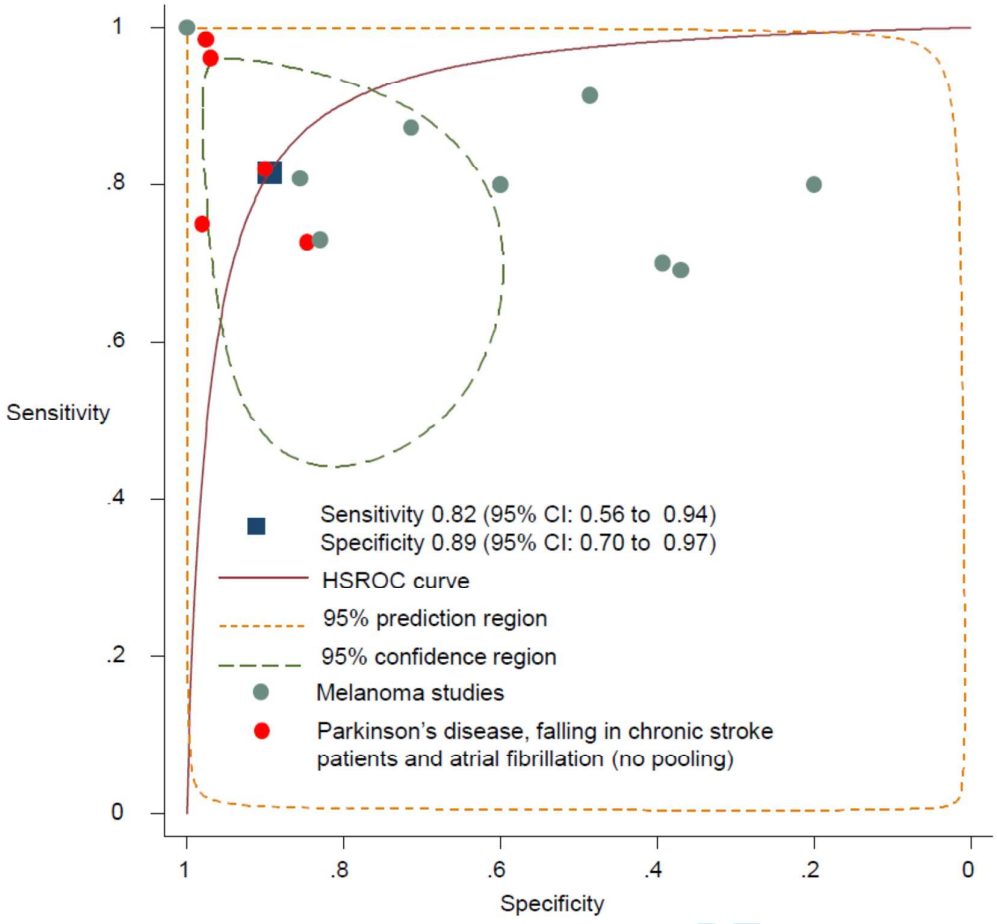
The summary estimate for sensitivity was 82 percent (95 % confidence interval (CI); 0.56 to 0.94) and pooled specificity was 89 percent (95%CI; 0.70 to 0.97). In a sub-group analysis of 12 reports, pooled sensitivity of studies assessing melanoma was 0.73 (95%CI; 0.36 to 0.93) and pooled specificity was 0.84 (95%CI; 0.54 to 0.96). No pooling was possible for Parkinson's disease, falling in chronic stroke patients 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 4** and the hierarchical summary receiver operating characteristic curve (HSROC) is seen in **Figure 2**. .

Table 4 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 2014	96%	97%	10	0	0	10	Not reported	Not reported	Parkinson's Disease
Arora et al 2015	99%	98%	10	0	0	10	Not reported	Not reported	Parkinson's Disease
Chadwick et al 2014	0%	100%	0	0	5	10	Not reported	Skin Scan	Melanoma
Chadwick et al 2014	0%	100%	0	0	4	5	Not reported	Mel App	Melanoma
Chadwick et al.2014	80%	20%	4	8	1	2	Not reported	Mole Detective	Melanoma
Chadwick et al 2014	80%	60%	4	4	1	6	Not reported	SpotMole Plus	Melanoma
Chadwick et al 2014	80%	60%	4	4	1	6	Not reported	Dr.Mole Premium	Melanoma
Kostikis et al 2015	75%	98%	19	2	4	18	0.94	Not reported	Parkinson's Disease
Lagido et al 2014	75%	98%	6	1	2	34	Not reported	Not reported	Atrial Fibrillation
Maier et al 2014	73%	83%	19	20	7	98	Not reported	Not reported	Melanoma
Ramlakhan et al 2011	91%	49%	42	19	4	18	Not reported	Not reported	Melanoma
Takuya et al 2015	73%	85%	8	2	3	11	0.75	Not reported	Falling in Chronic Stroke Patients
Wadhawan et al 2011	81%	86%	30	12	7	75	0.91	Skin Scan	Melanoma
Wadhawan et al 2011	87%	71%	96	68	14	169	Not reported	7-Point Checklist	Melanoma
Wolf et al 2013	70%	39	42	74	18	48	Not reported	Not reported	Melanoma
Wolf et al 2013	69%	37%	41	79	19	46	Not reported	Not reported	Melanoma
Wolf et al 2013	6%	94%	4	7	56	103	Not reported	Not reported	Melanoma

Figure 2 Hierarchical Summary Receiver Operating Characteristic Curve (HSROC)



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, 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 co-workers, who systematically summarized 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 randomized controlled trials to validate mental mobile health apps in clinical care (8). Likewise, Majeed-Ariss and co-authors, 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. Our study has several limitations. First, the primary studies found, had 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, they usually lead 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 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 US, 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.

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.

Competing interests

This study was supported via an unrestricted research grant of medignition Inc. LMB holds shares of medignition. LMB was responsible for the design and the statistical analysis of the study. All other authors declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years, no other relationships or activities that could appear to have influenced the submitted work.

Copyright

Contributions

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 co-authors provided intellectual input and approved the final manuscript. LMB is the study guarantor.

Transparency declaration

LMB affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

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 https://www.tractica.com/resources/white-papers/enterprise-wearable-technology-case-studies/: Tractica; 2015 [
2. Kaiser PK. Emerging therapies for neovascular age-related macular degeneration: drugs in the pipeline. Ophthalmology. 2013;120(5 Suppl):S11-5.
3. Institute I. Patient Adoption of mHealth. http://www.imshealth.com/en/thought-leadership/ims-institute/reports/patient-adoption-of-mhealth; 2013.
4. https://www.ncbi.nlm.nih.gov/mesh/?term=mobile+application aa. last visited: 2/2/2017.
5. Misra S. New report finds more than 165,000 mobile health apps now available, takes close look at characteristics & use 2015 [Available from: http://www.imedicalapps.com/2015/09/ims-health-apps-report/#.
6. F. BSaN. Healthcare's digital future http://www.mckinsey.com/industries/healthcare-systems-and-services/our-insights/healthcares-digital-future: McKiinsey & Company; 2014 [
7. Majeed-Ariss R, Hall AG, McDonagh J, Fallon D, Swallow V. 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(2):e40.
8. Bakker D, Kazantzis N, Rickwood D, Rickard N. Mental Health Smartphone Apps: Review and Evidence-Based Recommendations for Future Developments. JMIR Mental Health. 2016;3(1):e7.
9. Anderson K. BO, Emmerton L. Mobile Health Apps to Facilitate Self-Care: A Qualitative Study of User Experiences. PLOS ONE. 2016.
10. Moher D LA, Tetzlaff J, Altman DG, The PRISMA Group. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. . BMJ 2009.
11. Arora S, Venkataraman V, Zhan A, Donohue S, Biglan K, Dorsey E, et al. Detecting and monitoring the symptoms of Parkinson's disease using smartphones: a pilot study. Parkinsonism & related disorders. 2015;21(6):650-3.
12. Chadwick X, Loescher, L. J., Janda, M., & Soyer, H. P. 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, Kotsavasiloglou C. A Smartphone-Based Tool for Assessing Parkinsonian Hand Tremor. IEEE journal of biomedical and health informatics. 2015;19(6):1835-42.
14. Lagido RB, Lobo J, Leite S, Sousa C, Ferreira L, Silva-Cardoso J, editors. 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: IEEE.
15. Maier T, Kulichova D, Schotten K, Astrid R, Ruzicka T, Berking C, et al. Accuracy of a smartphone application using fractal image analysis of pigmented moles compared to clinical diagnosis and histological result. Journal of the European Academy of Dermatology and Venereology. 2015;29(4):663-7.
16. Ramlakhan K, Shang Y, editors. A mobile automated skin lesion classification system. 2011 IEEE 23rd International Conference on Tools with Artificial Intelligence; 2011: IEEE.
17. Wadhawan T, Situ N, Lancaster K, Yuan X, Zouridakis G. SkinScan©: A PORTABLE LIBRARY FOR MELANOMA DETECTION ON HANDHELD DEVICES. Proceedings / IEEE International Symposium on Biomedical Imaging: from nano to macro IEEE International Symposium on Biomedical Imaging. 2011;2011:133-6.
18. Wadhawan T, Situ N, Rui H, Lancaster K, Yuan X, Zouridakis G. Implementation of the 7-Point Checklist for Melanoma Detection on Smart Handheld Devices. Conference Proceedings. 2011;2011:3180-3.

19. Wolf JA, Moreau JF, Akilov O, Patton T, English JC, Ho J, et al. Diagnostic inaccuracy of smartphone applications for melanoma detection. JAMA dermatology. 2013;149(4):422-6.
20. Takuya I. HT, Shigeru U. Accelerometry-based gait characteristics evaluated using a smartphone and their association with fall risk in people with chronic stroke. Journal of Stroke and Cerebrovascular Diseases. 2015;24(6):1305-11.
21. Arora S. VV, Donohue S., Biglan K.M., Dorsey E.R., Little M.A., editor High Accuracy Discrimination of Parkinson's Disease Participants from Healthy Controls Using Smartphones. 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2014 2014.
22. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011;155(8):529-36.
23. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, 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, Egger M, Deeks JJ, Shang A, et al. An empirical comparison of methods for meta-analysis of diagnostic accuracy showed hierarchical models are necessary. J Clin Epidemiol. 2008;61(11):1095-103.
25. Bove R, White CC, Giovannoni G, Glanz B, Golubchikov V, Hujol J, et al. Evaluating more naturalistic outcome measures A 1-year smartphone study in multiple sclerosis. Neurology-Neuroimmunology Neuroinflammation. 2015;2(6):e162.
26. Daneault J-F, Carignan B, Codère CÉ, Sadikot AF, Duval C. Using a smart phone as a standalone platform for detection and monitoring of pathological tremors. Frontiers in human neuroscience. 2013;6:357.
27. Do T-T, Zhou Y, Zheng H, Cheung N-M, Koh D, editors. Early melanoma diagnosis with mobile imaging. 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society; 2014: IEEE.
28. Doukas C. SP, Kiranoudis C.T., Maglogiannis I. Automated skin lesion assessment using mobile technologies and cloud platforms. Conf Proc IEEE Eng Med Biol Soc. 2012:2444-7.
29. Ellis RJ, Ng YS, Zhu S, Tan DM, Anderson B, Schlaug G, et al. A Validated Smartphone-Based Assessment of Gait and Gait Variability in Parkinson's Disease. PLoS one. 2015;10(10):e0141694.
30. Kostikis N, Hristu-Varsakelis D, Arnaoutoglou M, Kotsavasiloglou C, editors. Smartphone-based evaluation of parkinsonian hand tremor: Quantitative measurements vs clinical assessment scores. 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society; 2014: IEEE.
31. Lee J, Reyes BA, McManus DD, Maitas O, Chon KH. Atrial fibrillation detection using an iPhone 4S. IEEE Transactions on Biomedical Engineering. 2013;60(1):203-6.
32. McManus DD, Lee J, Maitas O, Esa N, Pidikiti R, Carlucci A, 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(3):315-9.
33. Nishiguchi S, Ito H, Yamada M, Yoshitomi H, Furu M, Ito T, et al. Self-assessment tool of disease activity of rheumatoid arthritis by using a smartphone application. TELEMEDICINE and e-HEALTH. 2014;20(3):235-40.
34. Printy BP, Renken LM, Herrmann JP, Lee I, Johnson B, Knight E, et al., editors. Smartphone application for classification of motor impairment severity in parkinson's disease. 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society; 2014: IEEE.

35. Raknim P, Lan K-c. Gait Monitoring for Early Neurological Disorder Detection Using Sensors in a Smartphone: Validation and a Case Study of Parkinsonism. Telemedicine and e-Health. 2016;22(1):75-81.
36. Shinohara A, Ito T, Ura T, Nishiguchi S, Ito H, Yamada M, et al., editors. 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 Y-Z, He Y-G, Mitzel G, Zhang S, Bartlett MB. Handheld Shape Discrimination Hyperacuity (hSDH) Test on a Mobile Device for Remote Monitoring of Visual Function in Maculopathy. Investigative Ophthalmology & Visual Science. 2012;53(14):2914-.
38. Winther C, Frisen L. Self-Testing of Vision in Age-Related Macula Degeneration: A Longitudinal Pilot Study Using a Smartphone-Based Rarebit Test. J Ophthalmol. 2015;2015:285463.
39. Woods AM, Nowostawski M, Franz EA, Purvis M. Parkinson's disease and essential tremor classification on mobile device. Pervasive and Mobile Computing. 2014;13:1-12.
40. Fuadah YN, Setiawan AW, Mengko TL, editors. Mobile cataract detection using optimal combination of statistical texture analysis. 2015 4th International Conference on Instrumentation, Communications, Information Technology, and Biomedical Engineering (ICICI-BME); 2015: IEEE.
41. Yamada M, Aoyama T, Mori S, Nishiguchi S, Okamoto K, Ito T, et al. Objective assessment of abnormal gait in patients with rheumatoid arthritis using a smartphone. Rheumatology international. 2012;32(12):3869-74.
42. Zhu S, Ellis RJ, Schlaug G, Ng YS, Wang Y, editors. Validating an iOS-based Rhythmic Auditory Cueing Evaluation (iRACE) for Parkinson's Disease. Proceedings of the 22nd ACM international conference on Multimedia; 2014: ACM.
43. 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.
44. Bachmann LM, Coray R, Estermann P, Ter Riet G. Identifying diagnostic studies in MEDLINE: reducing the number needed to read. J Am Med Inform Assoc. 2002;9(6):653-8.

Appendix: Search strategies

SCOPUS

TITLE-ABS-KEY ((((**patient*** OR **outpatient*** OR **ambulant** OR **ambulatory**) W/3 **monitor***) OR **self-monitor***) AND (((**cell** OR **cellular** OR **mobile** OR **smart**) W/3 **phone**) OR (**smartphone*** OR **iphone***) OR ((**telemedicine** OR **ehealth** OR **mhealth** OR **telematic**) AND **mobile**)))

Medline (Ovid interface)

1 monitoring, ambulatory/
2 ((patient* or outpatient* or ambulant or ambulatory) adj3 monitor*).ti,ab.
3 self-monitor*.ti,ab.
4 1 or 2 or 3
5 exp cell phones/
6 ((cell or cellular or mobile or smart) adj3 phone).ti,ab.
7 (smartphone* or iPhone*).ti,ab.
8 or/5-7
9 exp telemedicine/ or exp telemetry/
10 (telemedicine or ehealth or mhealth or telematic).ti,ab.
11 9 or 10
12 mobile.ti,ab.
13 11 and 12
14 8 or 13
15 4 and 14

Business Source Premier

S1 ((patient OR outpatient OR ambulant OR ambulatory) N3 monitoring) OR selfmonitoring
S2 (((cell or cellular or mobile or smart) N3 phone)) OR ((smartphone* or iPhone*)) OR ((telemedicine or ehealth or mhealth or telematic) AND mobile))
S3 S1 AND S2

Science Citation Index

1 TS=(((patient* or outpatient* or ambulant or ambulatory) NEAR/3 monitor*) OR self-monitor*)
2 TS=(((cell or cellular or mobile or smart) NEAR/3 phone) OR (smartphone* or iPhone*) OR ((telemedicine or ehealth or mhealth or telematic) AND mobile))
3 #2 AND #1
4 PUBLICATION NAME: (JMIR MHEALTH "AND" UHEALTH OR JMIR MEDICAL INFORMATICS)
5 #4 AND #3
6 #2 AND #1 Refined by: RESEARCH AREAS: (MEDICAL INFORMATICS)
7 #6 AND #5
8 #6 OR #5

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.	3, 4
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	4
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.	5
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.	5
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.	appendix
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).	5
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.	5, 6
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	6
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.	5,6
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	6
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.	6

PRISMA 2009 Checklist

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).	-
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	n.a.
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.	7
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.	11,12, 14
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.	13, 14
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	13
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]).	n.a.
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).	16
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).	16
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	17
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.	6

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed1000097

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

BMJ Open

Evidence assessing the diagnostic performance of medical smartphone apps - A Systematic Review and exploratory meta-analysis

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-018280.R1
Article Type:	Research
Date Submitted by the Author:	29-Sep-2017
Complete List of Authors:	Buechi, Rahel; Cantonal Hospital Lucerne, Eye Clinic Faes, Livia; medignition Research Consultants, Bachmann, Lucas; medignition Research Consultants Thiel, Michael; Eye Clinic, Cantonal Hospital of Lucerne, Bodmer, Nicolas; medignition Research Consultants Schmid, Martin; Eye Clinic, Cantonal Hospital of Lucerne, Job, Oliver; Eye Clinic, Cantonal Hospital of Lucerne, Lienhard, Kenny; Faculty of Business and Economics, University of Lausanne, Information Systems
Primary Subject Heading:	Evidence based practice
Secondary Subject Heading:	Health informatics, Health services research
Keywords:	mobile health apps, evidence-based medicine, systematic review, diagnostic research

Evidence assessing the diagnostic performance of medical smartphone apps - A Systematic Review and exploratory meta-analysis

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

medignition Inc, Research Consultants, 8004 Zurich, Switzerland, Rahel Büchi research fellow, Livia Faes, research fellow, Lucas M Bachmann, professor of epidemiology, director, Nicolas S Bodmer Research Fellow, Eye Clinic, Cantonal Hospital of Lucerne Michael A Thiel, professor of ophthalmology, director, Martin K Schmid co-director, Oliver Job, head neuro-ophthalmology, Department of Information Systems, Faculty of Business and Economics, University of Lausanne Kenny R Lienhard, research fellow

Correspondence to: Lucas M. Bachmann: bachmann@medignition.ch

Rahel Buechi and Livia Faes contributed equally.

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 in-built 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 December 15th, 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 in-built sensors of a smartphone for diagnosis of disease. The methodological quality of 11 studies used in an exploratory meta-analysis was assessed with the QUADAS-2 tool and the reporting quality with the STARD statement. Sensitivity and specificity of studies reporting two-by-two tables were calculated and summarized.

Results We screened 3'296 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 1'048 subjects (758 with the target conditions and 290 healthy volunteers). Overall, the summary estimate for sensitivity was 0.82 (95 % confidence interval (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.

This systematic review was prospectively registered at PROSPERO under the number 42016033049.

Strength and Limitations of this Study

Strength

- A comprehensive literature search to retrieve the published evidence, applying stringent inclusion criteria and assessed the methodological quality of the studies systematically.

Limitations

- 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, nine percent 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 health care 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 dialog is of particular importance for patients with chronical 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 fact that most apps target at a diagnostic problem, it would be helpful to gauge the scientific basis of them. In this comprehensive systematic review we thus summarized the currently available papers assessing diagnostic properties of mobile health apps.

Methods

This review was conducted according to the PRISMA(10) statement recommendations.

Data Sources

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

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 in-built 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, i.e. the possibility to calculate sensitivity and specificity (O). We accepted all reference tests (C) used 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 2x2 table data that were summarized in the meta-analysis was made using the QUADAS-2 tool. Reporting quality was assessed using the STARD statement (22, 23). Quality assessment involved scrutinizing 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 standardized 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 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(24) pooling, at least 4 studies on the same target condition had to be available. Therefore, no separate analysis for health apps on Parkinson's disease, falling in chronic stroke patients and atrial fibrillation was possible.

All analyses were done using Stata 14.1 statistics software package (StataCorp LP,College Station, TX, USA).

Role of the Funding Source

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 sector, via an unrestricted research grant. LMB holds shares of medignition. LMB was responsible for the design and the statistical analysis of the study.

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, thirty studies provided some clinical data (2, 11-21, 25-42). Details on these studies are available in the **Appendix**. Eleven studies reporting seventeen 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**.

Table 1: Characteristics of included studies. The elven 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 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 2015	Parkinson's Disease	Diagnostic Case Control Study	No	10	65.1 (9.8)	30%	Not reported.	Not reported.
Chadwick et al 2014	Melanoma	Diagnostic Case Control Study	No	15	Not applicable	Not applicable	Not reported.	Not reported.
Kostikis et al 2015	Parkinson's Disease	Diagnostic Case Control Study	No	23	78	52%	Not reported.	Not reported.
Lagido et al 2014	Atrial Fibrillation	Prospective Cohort Study	No	43	Not reported	Not reported	Not reported.	Not reported.
Maier et al 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 e.g. hair, images containing more than one lesion, incomplete imaged lesions, non- melanocytic lesions, two-point differences cases.
Ramlakhan et al 2011	Melanoma	Diagnostic Case Control Study	No	46	Not applicable	Not applicable	Not reported.	Not reported.
Takuya et al 2015	Falling in Chronic Stroke Patients	Diagnostic Case Control Study	No	11	70.5 (12.5)	Not reported	More than 12 months since stroke onset and ability to walk 16 meters independently with or without a single-point cane and/or an orthosis.	Severe cardiovascular, respiratory, musculoskeletal, or neurologic disorder other than stroke that affected gait performance; unable to understand the instructions because of communication problem or moderate to severe cognitive dysfunction (i.e., 5 or more errors on the Short Portable Mental Status Questionnaire [SPMSQ]); household ambulators walked only indoors or only mobilized during rehabilitation sessions.

Wadhawan et al 2011	Melanoma	Diagnostic Case Control Study	No	1300	Not applicable	Not applicable	Not reported.	Image artefacts.
Wadhawan et al 2011	Melanoma	Diagnostic Case Control Study	No	347	Not applicable	Not applicable	Not reported.	Image artefacts.
Wolf et al 2013	Melanoma	Diagnostic Case Control Study	No	188	Not applicable	Not applicable	Images for which there was a clear histologic 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.

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 stroke patients (n=1)(20). The studies altogether involved 1'048 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 **appendix**.

Eleven studies (11-21) that investigated 13 diagnostic health apps allowing the construction of 17 two-by-two tables qualified for the meta-analysis. 12 tables reported on diagnosis of melanoma, three on Parkinson's disease, one assessed falling in chronic stroke patients, and another atrial fibrillation.

Methodological Quality

A summary of the methodological quality is shown in **Table 2**

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 2014	*Yes*	*Yes*	*Yes*	*Yes*
Arora et al 2015	*Yes*	*Yes*	*No*	*Yes*
Chadwick et al 2014	*Yes*	*Yes*	*No*	*Yes*
Kostikis et al 2015	*Yes*	*Yes*	*No*	*Yes*
Lagido et al 2014	*Yes*	*Yes*	*Yes*	*Yes*
Maier et al 2015	*Yes*	*Yes*	*No*	*Yes*
Ramlakhan et al 2011	*Yes*	*Yes*	*Yes*	*Yes*
Takuya et al 2015	*Yes*	*Yes*	*Yes*	*Yes*
Wadhawan et al 2011	*Yes*	*Yes*	*No*	*Yes*
Wadhawan et al 2011	*Yes*	*Yes*	*Yes*	*Yes*
Wolf et al 2013	*Yes*	*Yes*	*Yes*	*Yes*

Ten studies had a diagnostic case-control design and one studies 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 i.e. reasons for non-adherence, simplicity of use and difficulties and comprehensibility.

Exploratory analyses of diagnostic accuracy

The summary estimate for sensitivity was 82 percent (95 % confidence interval (CI); 0.56 to 0.94) and pooled specificity was 89 percent (95%CI; 0.70 to 0.97). In a sub-group 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 chronic stroke patients 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 (HSROC) is seen in **Figure 2**. .

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 2014	96%	97%	10	0	0	10	Not reported	Not reported	Parkinson's Disease
Arora et al 2015	99%	98%	10	0	0	10	Not reported	Not reported	Parkinson's Disease
Chadwick et al 2014	0%	100%	0	0	5	10	Not reported	Skin Scan	Melanoma
Chadwick et al 2014	0%	100%	0	0	4	5	Not reported	Mel App	Melanoma
Chadwick et al.2014	80%	20%	4	8	1	2	Not reported	Mole Detective	Melanoma
Chadwick et al 2014	80%	60%	4	4	1	6	Not reported	SpotMole Plus	Melanoma
Chadwick et al 2014	80%	60%	4	4	1	6	Not reported	Dr.Mole Premium	Melanoma
Kostikis et al 2015	75%	98%	19	2	4	18	0.94	Not reported	Parkinson's Disease
Lagido et al 2014	75%	98%	6	1	2	34	Not reported	Not reported	Atrial Fibrillation
Maier et al 2014	73%	83%	19	20	7	98	Not reported	Not reported	Melanoma
Ramlakhan et al 2011	91%	49%	42	19	4	18	Not reported	Not reported	Melanoma
Takuya et al 2015	73%	85%	8	2	3	11	0.75	Not reported	Falling in Chronic Stroke Patients
Wadhawan et al 2011	81%	86%	30	12	7	75	0.91	Skin Scan	Melanoma
Wadhawan et al 2011	87%	71%	96	68	14	169	Not reported	7-Point Checklist	Melanoma
Wolf et al 2013	70%	39	42	74	18	48	Not reported	Not reported	Melanoma
Wolf et al 2013	69%	37%	41	79	19	46	Not reported	Not reported	Melanoma
Wolf et al 2013	6%	94%	4	7	56	103	Not reported	Not reported	Melanoma

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, 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 co-workers, who systematically summarized 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 randomized controlled trials to validate mental mobile health apps in clinical care (8). Likewise, Majeed-Ariss and co-authors, 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 over-inclusive 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 percent 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 found, had 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, they usually lead 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 US, 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.

Competing interests

This study was supported via an unrestricted research grant of medignition Inc. LMB holds shares of medignition. LMB was responsible for the design and the statistical analysis of the study. All other authors declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years, no other relationships or activities that could appear to have influenced the submitted work.

other languages, create adaptations, reprints, include within collections and create summaries, extracts and/or, abstracts of the Contribution, iii) create any other derivative work(s) based on the Contribution, iv) to exploit all subsidiary rights in the Contribution, v) the inclusion of electronic links from the Contribution to third party material where-ever it may be located; and, vi) licence any third party to do any or all of the above.

Contributions

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 co-authors provided intellectual input and approved the final manuscript. LMB is the study guarantor.

Transparency declaration

LMB affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

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 https://www.tractica.com/resources/white-papers/enterprise-wearable-technology-case-studies/: Tractica; 2015 [
2. Kaiser PK. Emerging therapies for neovascular age-related macular degeneration: drugs in the pipeline. Ophthalmology. 2013;120(5 Suppl):S11-5.
3. Institute I. Patient Adoption of mHealth. http://www.imshealth.com/en/thought-leadership/ims-institute/reports/patient-adoption-of-mhealth; 2013.
4. https://www.ncbi.nlm.nih.gov/mesh/?term=mobile+application aa. last visited: 2/2/2017.
5. Misra S. New report finds more than 165,000 mobile health apps now available, takes close look at characteristics & use 2015 [Available from: http://www.imedicalapps.com/2015/09/ims-health-apps-report/#.
6. F. BSaN. Healthcare's digital future http://www.mckinsey.com/industries/healthcare-systems-and-services/our-insights/healthcares-digital-future: McKiinsey & Company; 2014 [
7. Majeed-Ariss R, Hall AG, McDonagh J, Fallon D, Swallow V. 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(2):e40.
8. Bakker D, Kazantzis N, Rickwood D, Rickard N. Mental Health Smartphone Apps: Review and Evidence-Based Recommendations for Future Developments. JMIR Mental Health. 2016;3(1):e7.
9. Anderson K. BO, Emmerton L. Mobile Health Apps to Facilitate Self-Care: A Qualitative Study of User Experiences. PLOS ONE. 2016.
10. Moher D LA, Tetzlaff J, Altman DG, The PRISMA Group. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. . BMJ 2009.
11. Arora S, Venkataraman V, Zhan A, Donohue S, Biglan K, Dorsey E, et al. Detecting and monitoring the symptoms of Parkinson's disease using smartphones: a pilot study. Parkinsonism & related disorders. 2015;21(6):650-3.
12. Chadwick X, Loescher, L. J., Janda, M., & Soyer, H. P. 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, Kotsavasiloglou C. A Smartphone-Based Tool for Assessing Parkinsonian Hand Tremor. IEEE journal of biomedical and health informatics. 2015;19(6):1835-42.
14. Lagido RB, Lobo J, Leite S, Sousa C, Ferreira L, Silva-Cardoso J, editors. 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: IEEE.
15. Maier T, Kulichova D, Schotten K, Astrid R, Ruzicka T, Berking C, et al. Accuracy of a smartphone application using fractal image analysis of pigmented moles compared to clinical diagnosis and histological result. Journal of the European Academy of Dermatology and Venereology. 2015;29(4):663-7.
16. Ramlakhan K, Shang Y, editors. A mobile automated skin lesion classification system. 2011 IEEE 23rd International Conference on Tools with Artificial Intelligence; 2011: IEEE.
17. Wadhawan T, Situ N, Lancaster K, Yuan X, Zouridakis G. SkinScan©: A PORTABLE LIBRARY FOR MELANOMA DETECTION ON HANDHELD DEVICES. Proceedings / IEEE International Symposium on Biomedical Imaging: from nano to macro IEEE International Symposium on Biomedical Imaging. 2011;2011:133-6.
18. Wadhawan T, Situ N, Rui H, Lancaster K, Yuan X, Zouridakis G. Implementation of the 7-Point Checklist for Melanoma Detection on Smart Handheld Devices. Conference Proceedings. 2011;2011:3180-3.

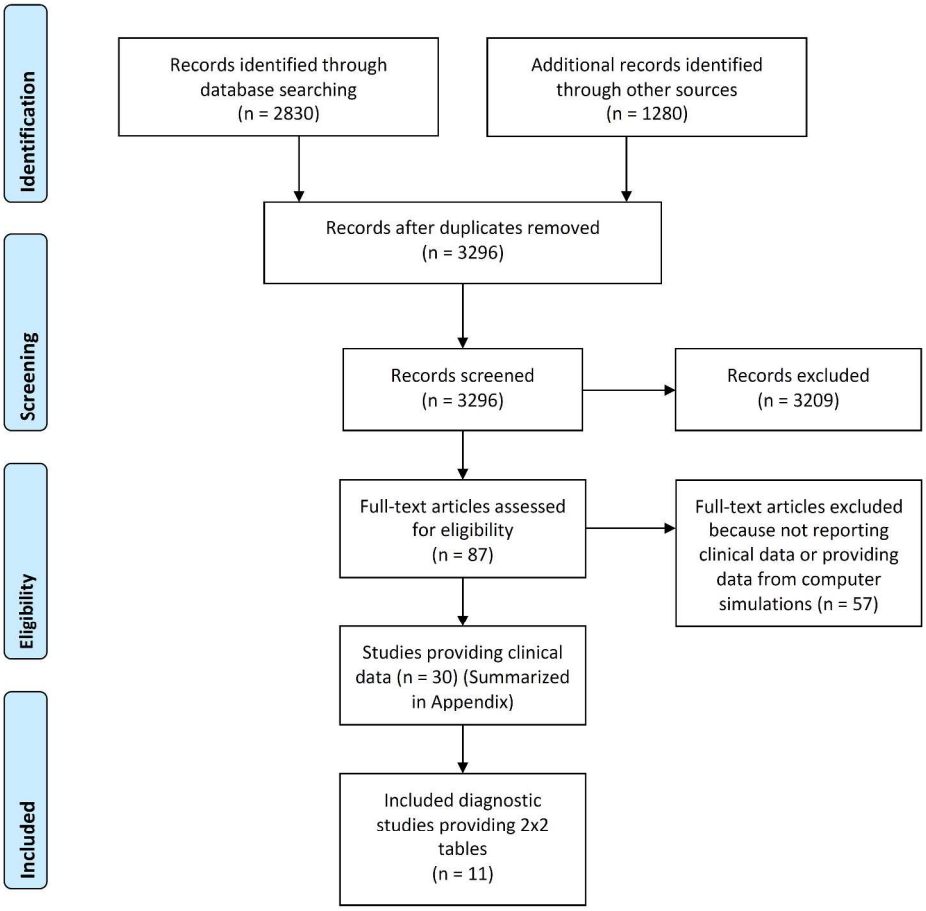
19. Wolf JA, Moreau JF, Akilov O, Patton T, English JC, Ho J, et al. Diagnostic inaccuracy of smartphone applications for melanoma detection. JAMA dermatology. 2013;149(4):422-6.
20. Takuya I. HT, Shigeru U. Accelerometry-based gait characteristics evaluated using a smartphone and their association with fall risk in people with chronic stroke. Journal of Stroke and Cerebrovascular Diseases. 2015;24(6):1305-11.
21. Arora S. VV, Donohue S., Biglan K.M., Dorsey E.R., Little M.A., editor High Accuracy Discrimination of Parkinson's Disease Participants from Healthy Controls Using Smartphones. 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2014 2014.
22. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011;155(8):529-36.
23. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, 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, Egger M, Deeks JJ, Shang A, et al. An empirical comparison of methods for meta-analysis of diagnostic accuracy showed hierarchical models are necessary. J Clin Epidemiol. 2008;61(11):1095-103.
25. Bove R, White CC, Giovannoni G, Glanz B, Golubchikov V, Hujol J, et al. Evaluating more naturalistic outcome measures A 1-year smartphone study in multiple sclerosis. Neurology-Neuroimmunology Neuroinflammation. 2015;2(6):e162.
26. Daneault J-F, Carignan B, Codère CÉ, Sadikot AF, Duval C. Using a smart phone as a standalone platform for detection and monitoring of pathological tremors. Frontiers in human neuroscience. 2013;6:357.
27. Do T-T, Zhou Y, Zheng H, Cheung N-M, Koh D, editors. Early melanoma diagnosis with mobile imaging. 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society; 2014: IEEE.
28. Doukas C. SP, Kiranoudis C.T., Maglogiannis I. Automated skin lesion assessment using mobile technologies and cloud platforms. Conf Proc IEEE Eng Med Biol Soc. 2012:2444-7.
29. Ellis RJ, Ng YS, Zhu S, Tan DM, Anderson B, Schlaug G, et al. A Validated Smartphone-Based Assessment of Gait and Gait Variability in Parkinson's Disease. PLoS one. 2015;10(10):e0141694.
30. Kostikis N, Hristu-Varsakelis D, Arnaoutoglou M, Kotsavasiloglou C, editors. Smartphone-based evaluation of parkinsonian hand tremor: Quantitative measurements vs clinical assessment scores. 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society; 2014: IEEE.
31. Lee J, Reyes BA, McManus DD, Maitas O, Chon KH. Atrial fibrillation detection using an iPhone 4S. IEEE Transactions on Biomedical Engineering. 2013;60(1):203-6.
32. McManus DD, Lee J, Maitas O, Esa N, Pidikiti R, Carlucci A, 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(3):315-9.
33. Nishiguchi S, Ito H, Yamada M, Yoshitomi H, Furu M, Ito T, et al. Self-assessment tool of disease activity of rheumatoid arthritis by using a smartphone application. TELEMEDICINE and e-HEALTH. 2014;20(3):235-40.
34. Printy BP, Renken LM, Herrmann JP, Lee I, Johnson B, Knight E, et al., editors. Smartphone application for classification of motor impairment severity in parkinson's disease. 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society; 2014: IEEE.

35. Raknim P, Lan K-c. Gait Monitoring for Early Neurological Disorder Detection Using Sensors in a Smartphone: Validation and a Case Study of Parkinsonism. Telemedicine and e-Health. 2016;22(1):75-81.
36. Shinohara A, Ito T, Ura T, Nishiguchi S, Ito H, Yamada M, et al., editors. 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 Y-Z, He Y-G, Mitzel G, Zhang S, Bartlett MB. Handheld Shape Discrimination Hyperacuity (hSDH) Test on a Mobile Device for Remote Monitoring of Visual Function in Maculopathy. Investigative Ophthalmology & Visual Science. 2012;53(14):2914-.
38. Winther C, Frisen L. Self-Testing of Vision in Age-Related Macula Degeneration: A Longitudinal Pilot Study Using a Smartphone-Based Rarebit Test. J Ophthalmol. 2015;2015:285463.
39. Woods AM, Nowostawski M, Franz EA, Purvis M. Parkinson's disease and essential tremor classification on mobile device. Pervasive and Mobile Computing. 2014;13:1-12.
40. Fuadah YN, Setiawan AW, Mengko TL, editors. Mobile cataract detection using optimal combination of statistical texture analysis. 2015 4th International Conference on Instrumentation, Communications, Information Technology, and Biomedical Engineering (ICICI-BME); 2015: IEEE.
41. Yamada M, Aoyama T, Mori S, Nishiguchi S, Okamoto K, Ito T, et al. Objective assessment of abnormal gait in patients with rheumatoid arthritis using a smartphone. Rheumatology international. 2012;32(12):3869-74.
42. Zhu S, Ellis RJ, Schlaug G, Ng YS, Wang Y, editors. Validating an iOS-based Rhythmic Auditory Cueing Evaluation (iRACE) for Parkinson's Disease. Proceedings of the 22nd ACM international conference on Multimedia; 2014: ACM.
43. 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.
44. Bachmann LM, Coray R, Estermann P, Ter Riet G. Identifying diagnostic studies in MEDLINE: reducing the number needed to read. J Am Med Inform Assoc. 2002;9(6):653-8.
45. Fryback DG, Thornbury JR. The efficacy of diagnostic imaging. Med Decis Making. 1991;11(2):88-94.

Figure Legends

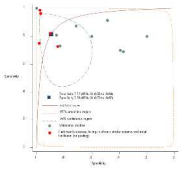
Figure 1. Flow Chart According to the PRISMA Statement

Figure 2 Hierarchical Summary Receiver Operating Characteristic Curve (HSROC)



Flow Chart According to the PRISMA Statement

391x366mm (300 x 300 DPI)



Hierarchical Summary Receiver Operating Characteristic Curve (HSROC)

355x266mm (300 x 300 DPI)

Appendix: Search strategies

SCOPUS

TITLE-ABS-KEY ((((**patient*** OR **outpatient*** OR **ambulant** OR **ambulatory**) W/3 **monitor***) OR **self-monitor***) AND (((**cell** OR **cellular** OR **mobile** OR **smart**) W/3 **phone**) OR (**smartphone*** OR **iphone***) OR ((**telemedicine** OR **ehealth** OR **mhealth** OR **telematic**) AND **mobile**)))

Medline (Ovid interface)

1 monitoring, ambulatory/
2 ((patient* or outpatient* or ambulant or ambulatory) adj3 monitor*).ti,ab.
3 self-monitor*.ti,ab.
4 1 or 2 or 3
5 exp cell phones/
6 ((cell or cellular or mobile or smart) adj3 phone).ti,ab.
7 (smartphone* or iPhone*).ti,ab.
8 or/5-7
9 exp telemedicine/ or exp telemetry/
10 (telemedicine or ehealth or mhealth or telematic).ti,ab.
11 9 or 10
12 mobile.ti,ab.
13 11 and 12
14 8 or 13
15 4 and 14

Business Source Premier

S1 ((patient OR outpatient OR ambulant OR ambulatory) N3 monitoring) OR selfmonitoring
S2 (((cell or cellular or mobile or smart) N3 phone)) OR ((smartphone* or iPhone*)) OR ((telemedicine or ehealth or mhealth or telematic) AND mobile))
S3 S1 AND S2

Science Citation Index

1 TS=(((patient* or outpatient* or ambulant or ambulatory) NEAR/3 monitor*) OR self-monitor*)
2 TS=(((cell or cellular or mobile or smart) NEAR/3 phone) OR (smartphone* or iPhone*) OR ((telemedicine or ehealth or mhealth or telematic) AND mobile))
3 #2 AND #1
4 PUBLICATION NAME: (JMIR MHEALTH "AND" UHEALTH OR JMIR MEDICAL INFORMATICS)
5 #4 AND #3
6 #2 AND #1 Refined by: RESEARCH AREAS: (MEDICAL INFORMATICS)
7 #6 AND #5
8 #6 OR #5

author	year	design	sampling	Exclusion criteria	Inclusion criteria	Recruitment	consecutive	duration	clinical condition	Type of measurement
Bove	2015	Cohort study, paired control-group, feasibility study, human observational trial	Convenient	not reported	Informed consent, Multiple sclerosis	Pairs consisting of 1 patient with demyelinating disease and 1 healthy cohabitant, were recruited at the Partners MS Center, a large referral clinical center in the northeastern United States.	no	1 year monitoring	Severity of MS	Questionnaires and visual tests should classify severity of MS
Bove	2015	Cohort study, paired control-group, feasibility study, human observational trial	Convenient	not reported	Informed consent	Pairs consisting of 1 patient with demyelinating disease and 1 healthy cohabitant, all aged 18–55 years, were recruited at the Partners MS Center, a large referral clinical center in the northeastern United States. Cohabitant pairs were recruited to control for common environment.	no	1 year monitoring	Healthy cohabitants of participating patients	Questionnaires and visual tests should provide a control for environmental factors
Kaiser	2013	Open-label, single-arm, multicentre study, Pilot study	Prospective	Patients were excluded if they had concomitant ocular disease in the study eye; neurologic impairment that would interfere with study assessments; use of systemic medications known to be toxic to the lens, retina, or optic nerve; or use of any other investigational agents within 60 days of screening.	Active CNV secondary to AMD (either newly diagnosed and treatment-naive or successfully treated with anti–vascular endothelial growth factor therapy for 1 year) in at least 1 eye, eligible for ranibizumab therapy, with best corrected visual acuity (BCVA) letter score 24 or higher (20/320 Snellen equivalent) by "Early Treatment Diabetic Retinopathy Study chart" at 4 m. Only one eye was required to fulfil entry criteria for a patient to enrol in the study; if both of the patient's eyes had CNV-AMD, then both were included in the analyses.	This study was conducted at 24 centers in the United States (NCT01542866).	no	16 weeks	Wet age-related macular degeneration	Distortion algorithm
Printy	2014	Cohort study, Pilot study	unclear	not reported	not reported	In a movement disorders clinic on the day of an outpatient appointment.	no	not reported	Parkinson's disease and severity	Quantification of the severity of Parkinson's motors symptoms using an application that collects kinematic data and extracts quantitative features using signal processing techniques, support vector machine classifier.
Lagido	2014	Diagnostic case-control	unclear	not reported	not reported	Samples were collected from heart failure patients at rest in Hospital S. Joao in Porto.	no	not reported	Atrial fibrillation	Detect heart rate and heart rate variability using a photoplethysmogram signal with the user's fingertip placed over the smartphone camera.

Maier	2014	Diagnostic case-control	Prospective	Images were excluded from evaluation in case of poor quality images, other elements in the image not belonging to the lesion e.g. hair, images containing more than one lesion, incomplete imaged lesions, non-melanocytic lesions, two-point differences cases (results in non-consecutive risk classes mainly due to inappropriate imaging angle or distance). The cases with an equal number of results in two consecutive risk classes, so-called tie cases (e.g. 1 high risk, 1 medium risk and 1 low risk results), were also excluded.	not reported	We included 195 melanocytic lesions in consecutive patients seen routinely for skin cancer screening at the Department of Dermatology, University Hospital of Munich, Germany after obtaining written informed consent.	yes	not reported	Melanoma	Risk assessment algorithm
Ellis	2015	Diagnostic case-control with age-matched healthy controls	unclear	not reported	A telephone questionnaire was first administered to screen out potential subjects who (1) are not within the age range of 40 to 85; (2) have any problems with their hearing; (3) are not able to walk independently without an aid; (4) have joint problems or other neurological, musculoskeletal or medical problems that can affect walking; (5) have sustained a fall within the past year that continues to affect their walking pattern; (6) have had surgery to implant a device (e.g., deep brain stimulation or pacemaker). Subjects who satisfied all six criteria were invited to participate in the study. Upon arrival at the testing location, four clinical assessments were administered.	All subjects were recruited through the Singapore General Hospital clinics. For safety reasons, the inclusion/exclusion criteria for the present study precluded patients with severe gait dysfunction; most patients in the present sample would be considered to have “moderately advanced” disease. Whether SmartMOVE would perform as well in the case of severe gait dysfunction (e.g., shuffling steps or frequent gait freezing episodes) is thus unknown.	no	not reported	Parkinson's disease	Smartphone’s inertial measurement unit to record gait movements during walking. / The accuracy of smartphone-based gait analysis (utilizing the smartphone’s built-in tri-axial accelerometer and gyroscope to calculate successive step times and step lengths) was validated against two heel contact–based measurement devices: heel-mounted footswitch sensors (to capture step times) and an instrumented pressure sensor mat (to capture step lengths).

Ellis	2015	Diagnostic case-control with age-matched healthy controls	unclear	not reported	A telephone questionnaire was first administered to screen out potential subjects who (1) are not within the age range of 40 to 85; (2) have any problems with their hearing; (3) are not able to walk independently without an aid; (4) have joint problems or other neurological, musculoskeletal or medical problems that can affect walking; (5) have sustained a fall within the past year that continues to affect their walking pattern; (6) have had surgery to implant a device (e.g., deep brain stimulation or pacemaker). Subjects who satisfied all six criteria were invited to participate in the study. Upon arrival at the testing location, four clinical assessments were administered.	All subjects were recruited through the Singapore General Hospital clinics. For safety reasons, the inclusion/exclusion criteria for the present study precluded patients with severe gait dysfunction; most patients in the present sample would be considered to have "moderately advanced" disease. Whether SmartMOVE would perform as well in the case of severe gait dysfunction (e.g., shuffling steps or frequent gait freezing episodes) is thus unknown.	no	not reported	Healthy subjects	Smartphone's inertial measurement unit to record gait movements during walking. / The accuracy of smartphone-based gait analysis (utilizing the smartphone's built-in tri-axial accelerometer and gyroscope to calculate successive step times and step lengths) was validated against two heel contact–based measurement devices: heel-mounted footswitch sensors (to capture step times) and an instrumented pressure sensor mat (to capture step lengths).
Nishiguchi	2014	Open-label, follow-up study	unclear	Other musculoskeletal disorders, cognitive disorders, Parkinson's disease, stroke, or unable to walk unassisted over 15m using walking aids. Patients with previous surgery in the lower extremities were also excluded.	Patients with rheumatoid arthritis defined by the 1987 or 2010 American College of Rheumatology criteria were included.	not reported	no	unclear	Rheumatoid arthritis, disease activity	The modified Health Assessment Questionnaire (mHAQ), self- assessed TJC (self-assessed tender joint count, out of 49 joints), and self-assessed SJC (self-assessed swollen joint count, out of 46 joints) were recorded on the smartphone application that we developed. The mHAQ, a self-reported measure of physical function to quantify functional disability. The mHAQ is expressed on a scale ranging from 0 to 3, where 0 = no disability and 3 = severe functional disability. Gait Analysis: The participants were instructed to walk along a 15-m walkway at their preferred speed. Trunk linear accelerations were measured by participants themselves with the smartphone as they walked on the walkway. The smartphone was kept adjacent to the L3 spinous process, which is close to where the body's center of mass is believed to be located during quiet standing using a semi-elastic belt.

Shinohara	2013	Feasibility study	unclear	Other musculoskeletal disorders, cognitive disorders, Parkinson's disease, stroke, or unable to walk over 10 m unassisted. Patients with previous surgery in the lower extremities were also excluded.	Patients with rheumatoid arthritis as defined by the American College of Rheumatology 1987 or 2010 criteria were included.	The participants were patients who attended the rheumatology outpatient clinic of Kyoto University Hospital.	no	Until the next hospital visit: mean duration, 35.0 ± 11.3 days	Rheumatoid arthritis, disease activity	Linear trunk accelerations are gathered by the participants' smartphones (kept in a waist pouch) as they walked for 10 seconds at their preferred speed. Peak frequency (PF), autocorrelation peak (AC), and coefficient of variance (CV) of the acceleration peak intervals. The PF value indicates the gait cycle, which is the time taken for 1 step. The AC value indicates the degree of gait balance, so a higher AC value indicates a greater degree of balance. The CV value indicates the degree of gait variability, i.e., the variability in the elapsed time between the first contacts of 2 consecutive footfalls. The modified Health Assessment Questionnaire (mHAQ), self- assessed TJC (self-assessed tender joint count, out of 49 joints), and self-assessed SJC (self-assessed swollen joint count, out of 46 joints) were recorded on the smartphone application that we developed. The mHAQ, a self-reported measure of physical function to quantify functional disability in RA. The mHAQ is expressed on a scale ranging from 0 to 3, where 0 = no disability and 3 = severe functional disability. General health condition and pain condition were recorded on the smartphone using a visual analogue scale (VAS). Gait Analysis: The participants were instructed to walk along a 15-m walkway at their preferred speed. Trunk linear accelerations were measured by participants themselves with the smartphone as they walked on the walkway. The smartphone was kept adjacent to the L3 spinous process, which is close to where the body's center of mass is believed to be located during quiet standing using a semi-elastic belt.

Yamada	2011	Cross-sectional study	unclear	We excluded participants based on the following exclusion criteria: other musculoskeletal disorders, cognitive disorders, Parkinson's disease, stroke, or unable to walk unassisted over 15 m using current walking aids.	Patients with rheumatoid arthritis defined by the American College of Rheumatology 1987 criteria were included.	This was a cross-sectional study performed between April 2011 and May 2011 in the rheumatology outpatient clinics of Kyoto University Hospital. A total of 39 RA patients (mean age, 65.9 ± 10.0 years) participated.	not reported	not reported	Rheumatoid arthritis, disease activity	The smartphone used in this study includes an acceleration sensor, a recording device, and a computer program for processing the acceleration signals. Trunk linear accelerations were measured using the smartphone while the subject walked on the walkway. The smartphone was attached to the L3 spinous process using a semi-elastic belt. Before measurements, the accelerometer of the smartphone was calibrated statically against gravity. The accelerometer of the smartphone sampled at 33 Hz. The recorded signals were analysed by the application developed in the android environment. Gait analysis: The participants were instructed to walk on a 20-m walk- way at their preferred speed. All participants wore their usual walking shoes, avoiding high heels and hard-soled shoes. The mid (10-m) walking time was measured using an electronic stopwatch.
Yamada	2011	Cross-sectional study	unclear	We excluded participants based on the following exclusion criteria: other musculoskeletal disorders, cognitive disorders, Parkinson's disease, stroke, or unable to walk unassisted over 15 m using current walking aids.	not reported	Twenty older individuals also took part in this experiment as control participants.	not reported	not reported	Healthy Control	The smartphone used in this study includes an acceleration sensor, a recording device, and a computer program for processing the acceleration signals. Trunk linear accelerations were measured using the smartphone while the subject walked on the walkway. The smartphone was attached to the L3 spinous process using a semi-elastic belt. Before measurements, the accelerometer of the smartphone was calibrated statically against gravity. The accelerometer of the smartphone sampled at 33 Hz. The recorded signals were analysed by the application developed in the android environment. Gait analysis: The participants were instructed to walk on a 20-m walk- way at their preferred speed. All participants wore their usual walking shoes, avoiding high heels and hard-soled shoes. The mid (10-m) walking time was measured using an electronic stopwatch.

Lee	2013	Case-control study, with-in	Prospective	not reported	not reported	Patients who presented for electrical cardioversion to the University of Massachusetts Medical Center (UMMC) cardiac electrophysiology laboratory were recruited by trained study personnel (McManus, Mathias)	no	not reported	Atrial fibrillation	Detect Atrial fibrillation (AF) and non-sinus rhythm (NSR) using a photoplethysmogram signal with the user's fingertip placed over the smartphone camera. AF and NSR detection is based on threshold values derived from the MIT-BIH AF and MIT-BIH NSR databases using statistical method RMSSD (Root mean square of successive differences).
Lee	2013	Case-control study, with-in	Prospective	not reported	not reported	Patients who presented for electrical cardioversion to the University of Massachusetts Medical Center (UMMC) cardiac electrophysiology laboratory were recruited by trained study personnel (McManus, Mathias)	no	not reported	Atrial fibrillation	Detect Atrial fibrillation (AF) and non-sinus rhythm (NSR) using a photoplethysmogram signal with the user's fingertip placed over the smartphone camera. AF and NSR detection is based on threshold values derived from the MIT-BIH AF and MIT-BIH NSR databases using statistical method RMSSD (Root mean square of successive differences).
Lee	2013	Case-control study, with-in	Prospective	not reported	not reported	Patients who presented for electrical cardioversion to the University of Massachusetts Medical Center (UMMC) cardiac electrophysiology laboratory were recruited by trained study personnel (McManus, Mathias)	no	not reported	Atrial fibrillation	Detect Atrial fibrillation (AF) and non-sinus rhythm (NSR) using a photoplethysmogram signal with the user's fingertip placed over the smartphone camera. AF and NSR detection is based on threshold values derived from the MIT-BIH AF and MIT-BIH NSR databases using statistical method RMSSD (Root mean square of successive differences).
Zhu	2014	unclear	unclear	None of the patients had a coexisting dementia (Mini-Mental State Examination score >24 points) or other diagnosed neurological impairments.	Patients diagnosed with idiopathic Parkinson's disease from Singapore General Hospital; All patients were under stable medication regiments for the preceding four weeks, and were tested at least 30 minutes after taking morning medications.	Patients diagnosed with idiopathic Parkinson's disease were recruited from Singapore General Hospital.	no	not reported	Parkinson's disease	To help “scale up” rhythmic auditory cueing (RAC) for wider distribution, we have developed an iOS-based Rhythmic Auditory Cueing Evaluation (iRACE) mobile application to deliver RAC and assess motor performance in PD patients. The touchscreen of the mobile device is used to assess motor timing during index finger tapping, and the device's built-in tri-axial accelerometer and gyro- scope to assess step time and step length during walking. Novel machine learning-based gait analysis algorithms have been developed for iRACE, including heel strike detection, step length quantification, and left-versus-right foot identification.

Wolf	2013	Diagnostic case-control	Convenient	Images that contained any identifiable features such as facial features, tattoos, or labels with patient information were either excluded or cropped to remove the identifiable features or information. Lesions with equivocal diagnoses such as “melanoma cannot be ruled out” or “atypical melanocytic proliferation” were excluded, as were Spitz nevi, pigmented spindle cell nevus of Reed, and other uncommon or equivocal lesions. We also excluded lesions with moderate or high-grade atypia given the controversy over their management. Two of the investigators then reviewed all images for image quality and omitted those that were of poor quality or resolution.	Images for which there was a clear histologic diagnosis rendered by a board-certified pathologist. The remaining images were stratified into one of the following categories: invasive melanoma, melanoma in situ, lentigo, benign nevus (including compound, junctional, and low-grade dysplastic nevi), dermatofibroma, seborrheic keratosis, and hemangioma. We only used close-up images of lesions.	not reported	no	not reported	Melanoma	Application 1 uses an automated algorithm to detect the border of the lesion, although it also allows manual input to confirm or change the detected border. It is the only application we tested that has this feature of user input for border detection. The application then analyses the image and gives an assessment of “problematic,” which we considered to be a positive test, “ok,” which we considered to be a negative test, or “error” if the image could not be assessed by the application. We categorized the latter group as unevaluable.
Wolf	2013	Diagnostic case-control	Convenient	Images that contained any identifiable features such as facial features, tattoos, or labels with patient information were either excluded or cropped to remove the identifiable features or information. Lesions with equivocal diagnoses such as “melanoma cannot be ruled out” or “atypical melanocytic proliferation” were excluded, as were Spitz nevi, pigmented spindle cell nevus of Reed, and other uncommon or equivocal lesions. We also excluded lesions with moderate or high-grade atypia given the controversy over their management. Two of the investigators then reviewed all images for image quality and omitted those that were of poor quality or resolution.	Images for which there was a clear histologic diagnosis rendered by a board-certified pathologist. The remaining images were stratified into one of the following categories: invasive melanoma, melanoma in situ, lentigo, benign nevus (including compound, junctional, and low-grade dysplastic nevi), dermatofibroma, seborrheic keratosis, and hemangioma. We only used close-up images of lesions.	not reported	no	not reported	Melanoma	Application 2 uses an automated algorithm to evaluate an image that has been uploaded by the user. The output given is either “melanoma,” which we considered to be a positive test, or “looks good” which we considered to be a negative test. If the image could not be analysed a message of “skin condition not found” was given and we considered the image unevaluable.

Wolf	2013	Diagnostic case-control	Convenient	Images that contained any identifiable features such as facial features, tattoos, or labels with patient information were either excluded or cropped to remove the identifiable features or information. Lesions with equivocal diagnoses such as “melanoma cannot be ruled out” or “atypical melanocytic proliferation” were excluded, as were Spitz nevi, pigmented spindle cell nevus of Reed, and other uncommon or equivocal lesions. We also excluded lesions with moderate or high-grade atypia given the controversy over their management. Two of the investigators then reviewed all images for image quality and omitted those that were of poor quality or resolution.	Images for which there was a clear histologic diagnosis rendered by a board-certified pathologist. The remaining images were stratified into one of the following categories: invasive melanoma, melanoma in situ, lentigo, benign nevus (including compound, junctional, and low-grade dysplastic nevi), dermatofibroma, seborrheic keratosis, and hemangioma. We only used close-up images of lesions.	not reported	no	not reported	Melanoma	Application 3 asks the user to upload an image to the application and then to position it within a box to ensure that the correct lesion is analysed. The output given by the application is “high risk,” which we considered to be a positive test, or “medium risk” or “low risk,” both of which we considered to be a negative test. The presence of a medium risk category in this application presented some difficulty in analysis as it was the only application tested that gave an intermediate output. Thus, we did perform sensitivity and specificity analysis with “medium risk” lesions counting as a positive test as well since it is not clear how a user would interpret such a result. Some lesions generated a message of “error” and these were considered unevaluable.
McManus	2013	Diagnostic case-control, within	Prospective	not reported	not reported	In a prospectively recruited cohort of participants undergoing cardioversion for atrial fibrillation	no	not reported	Atrial fibrillation	Mean square of successive differences (RMSSD): Our application acquired pulsatile signals by illuminating the fingertip using the standard iPhone lamp and recording video signal (30 frames/s) for 2 minutes. The signal was processed by averaging 50x50 green band pixels per frame. We interpolated the pulsatile signal to 30 Hz using a cubic spline algorithm followed by peak detection. As described in prior work, we use a peak detection algorithm that uses a filter bank with estimates of heart rate, variable cut-off frequencies, rank-order nonlinear filters, and decision logic as well as motion noise correction.

	McManus	2013	Diagnostic case-control, within	Prospective	not reported	not reported	In a prospectively recruited cohort of participants undergoing cardioversion for atrial fibrillation	no	not reported	Atrial fibrillation	Shannon entropy (ShE): Our application acquired pulsatile signals by illuminating the fingertip using the standard iPhone lamp and recording video signal (30 frames/s) for 2 minutes. The signal was processed by averaging 50x50 green band pixels per frame. We interpolated the pulsatile signal to 30 Hz using a cubic spline algorithm followed by peak detection. As described in prior work, we use a peak detection algorithm that uses a filter bank with estimates of heart rate, variable cut-off frequencies, rank-order nonlinear filters, and decision logic as well as motion noise correction.
	McManus	2013	Diagnostic case-control, within	Prospective	not reported	not reported	In a prospectively recruited cohort of participants undergoing cardioversion for atrial fibrillation	no	not reported	Atrial fibrillation	Combined statistical method: root mean square of successive RR difference (RMSSD/mean) and Shannon entropy (ShE): Our application acquired pulsatile signals by illuminating the fingertip using the standard iPhone lamp and recording video signal (30 frames/s) for 2 minutes. The signal was processed by averaging 50x50 green band pixels per frame. We interpolated the pulsatile signal to 30 Hz using a cubic spline algorithm followed by peak detection. As described in prior work, we use a peak detection algorithm that uses a filter bank with estimates of heart rate, variable cut-off frequencies, rank-order nonlinear filters, and decision logic as well as motion noise correction.

Takuya	2015	Cross-sectional study	Convenient	Exclusion criteria were (1) severe cardiovascular, respiratory, musculoskeletal, or neurologic disorder other than stroke that affected gait performance; (2) unable to understand the instructions because of communication problem or moderate to severe cognitive dysfunction (i.e., 5 or more errors on the Short Portable Mental Status Questionnaire [SPMSQ]); and (3) household ambulators walked only indoors or only mobilized during rehabilitation sessions.	Inclusion criteria were (1) more than 12 months since stroke onset and (2) ability to walk 16 m independently with or without a single point cane and/or an orthosis.	Community-dwelling adults with chronic stroke receiving day care services were recruited and screened for inclusion and exclusion criteria. conducted in 2 day care centers for elderly adults in Saitama, Japan	no		not reported	Chronic stroke, falls	Gait characteristics
Takuya	2015	Cross-sectional study	Convenient	Exclusion criteria were (1) severe cardiovascular, respiratory, musculoskeletal, or neurologic disorder other than stroke that affected gait performance; (2) unable to understand the instructions because of communication problem or moderate to severe cognitive dysfunction (ie, 5 or more errors on the Short Portable Mental Status Questionnaire [SPMSQ]); and (3) household ambulators walked only indoors or only mobilized during rehabilitation sessions.	Inclusion criteria were (1) more than 12 months since stroke onset and (2) ability to walk 16 m independently with or without a single point cane and/or an orthosis.	Community-dwelling adults with chronic stroke receiving day care services were recruited and screened for inclusion and exclusion criteria. conducted in 2 day care centers for elderly adults in Saitama, Japan	no		not reported	chronic stroke, no falls	Gait characteristics
Chadwick	2014	Case-control study	Convenient	not reported	not reported	The first step in the app testing was the selection of 15 high-quality, full resolution, digital images (.jpg) of melanocytic skin lesions of varying risk (5 melanomas, 10 benign nevi) from the image archive of the study dermatology expert (HPS) to assess the accuracy of the app analysis software.	no		not reported	Melanoma	SkinScan - Asymmetry/Border: Analysis by inbuilt pattern recognition software. / Color: Analysis by input comparison algorithms. / Diameter: no input. / Evolution: no input

Chadwick	2014	Case-control study	Convenient	not reported	not reported	The first step in the app testing was the selection of 15 high-quality, full resolution, digital images (.jpg) of melanocytic skin lesions of varying risk (5 melanomas, 10 benign nevi) from the image archive of the study dermatology expert (HPS) to assess the accuracy of the app analysis software.	no	not reported	Melanoma	MelApp - Asymmetry/Border: Analysis by inbuilt pattern recognition software. Area could be limited to focus the analysis. / Color: Analysis by input comparison algorithms. / Diameter: manual sliding scale input. / Evolution: manual sliding scale input
Chadwick	2014	Case-control study	Convenient	not reported	not reported	The first step in the app testing was the selection of 15 high-quality, full resolution, digital images (.jpg) of melanocytic skin lesions of varying risk (5 melanomas, 10 benign nevi) from the image archive of the study dermatology expert (HPS) to assess the accuracy of the app analysis software.	no	not reported	Melanoma	Mole Detective - Asymmetry/Border: Analysis by inbuilt pattern recognition software. / Color: Analysis by input comparison algorithms. / Diameter: Manual input of <6mm, -mm, or >6mm. / Evolution: No input for analysis. Reminder can be set for future use.
Chadwick	2014	Case-control study	Convenient	not reported	not reported	The first step in the app testing was the selection of 15 high-quality, full resolution, digital images (.jpg) of melanocytic skin lesions of varying risk (5 melanomas, 10 benign nevi) from the image archive of the study dermatology expert (HPS) to assess the accuracy of the app analysis software.	no	not reported	Melanoma	Spot Mole Plus - Asymmetry/Border: Analysis by inbuilt pattern recognition software. Manual adjustment of lesion border available/ Color: Analysis by input comparison algorithms. / Diameter: Manual input of numeric value. / Evolution: no input for past history of change. Can perform serial analysis of lesion images.
Chadwick	2014	Case-control study	Convenient	not reported	not reported	The first step in the app testing was the selection of 15 high-quality, full resolution, digital images (.jpg) of melanocytic skin lesions of varying risk (5 melanomas, 10 benign nevi) from the image archive of the study dermatology expert (HPS) to assess the accuracy of the app analysis software.	no	not reported	Melanoma	Dr. Mole Premium - Asymmetry/Border: Analysis by inbuilt pattern recognition software. Comparison of lesion quadrants for asymmetry./ Color: Analysis by input comparison algorithms. / Diameter: Manual sliding scale input. / Evolution: manual input of "none", "slow" & "fast" with time frames for each.

Winther	2015	Longitudinal Pilot study	Convenient	Exclusion criteria were the bilateral presence of currently inactive lesions judged to run a small risk of recurrence, the presence of additional, non- maculopathic causes of visual loss, and inability to participate in conventional acuity testing. Twenty-eight patients partook in the study. Those who had the same type of lesions in both eyes (active or inactive) provided results from the least involved eye only whereas those who had different types of lesions provided results from both eyes.	not reported	Patients were recruited from the wet age-related macular degeneration treatment programme of the Retina Unit at the Sahlgrenska University Hospital, a tertiary-care center.	no	Average monitoring of average of 30 weeks. Another time they report an average of 39	Wet age-related macular degeneration	For formal analysis, all MBT plots were carefully evaluated by subjective inspection, epoch by epoch. Epochs showing trends of decreasing scores, or increasing variation, or both, were rated worse. Epochs showing the opposite evolution were rated better. All other epochs were rated stable. ETDRS results were rated similarly, using direct numerical comparisons of scores. The outcomes of the clinical exam nations, which included biomicroscopy and scrutiny of OCT parameters and maps, were summarized in the same manner.
Winther	2015	Longitudinal Pilot study	Convenient	Exclusion criteria were the bilateral presence of currently inactive lesions judged to run a small risk of recurrence, the presence of additional, non- maculopathic causes of visual loss, and inability to participate in conventional acuity testing. Twenty-eight patients partook in the study. Those who had the same type of lesions in both eyes (active or inactive) provided results from the least involved eye only whereas those who had different types of lesions provided results from both eyes.	not reported	Twenty control subjects were recruited primarily from patients' relatives or other accompanying persons.	no	Average monitoring of average of 30 weeks. Another time they report an average of 39	Healthy relatives or accompanying	not reported

Wang	2013	Cross-sectional study	Convenient	not reported	1) AMD or DR with corrected Early Treatment of Diabetic Retinopathy Study (ETDRS) VA of 20/100 or better in at least one eye, 2) ophthalmic evaluation by retina specialists with clinical and spectral-domain (SD)-OCT documentation, 3) no retinal pathology other than AMD or DR, 4) no concurrent systemic illness affecting the retina, and 5) no dementia or other limitation that would prevent the patient from performing a self-test of visual function. Patients with AMD and DR were recruited at various disease stages, including those under active anti-vascular endothelial growth factor treatment. Patients with epiretinal membrane or pigment epithelial detachment were not excluded.	Patients with AMD and DR were recruited from the clinic of the Department of Ophthalmology, UT Southwestern Medical Center.	no	not reported	Age-related macular degeneration

Wang	2013	Cross-sectional study	unclear	not reported	1) AMD or DR with corrected Early Treatment of Diabetic Retinopathy Study (ETDRS) VA of 20/100 or better in at least one eye, 2) ophthalmic evaluation by retina specialists with clinical and spectral-domain (SD)-OCT documentation, 3) no retinal pathology other than AMD or DR, 4) no concurrent systemic illness affecting the retina, and 5) no dementia or other limitation that would prevent the patient from performing a self-test of visual function. Patients with AMD and DR were recruited at various disease stages, including those under active anti-vascular endothelial growth factor treatment. Patients with epiretinal membrane or pigment epithelial detachment were not excluded.	Patients with AMD and DR were recruited from the clinic of the Department of Ophthalmology, UT Southwestern Medical Center.	no	not reported	Diabetic retinopathy

Wang	2013	Cross-sectional study	unclear	not reported	1) AMD or DR with corrected Early Treatment of Diabetic Retinopathy Study (ETDRS) VA of 20/100 or better in at least one eye, 2) ophthalmic evaluation by retina specialists with clinical and spectral-domain (SD)-OCT documentation, 3) no retinal pathology other than AMD or DR, 4) no concurrent systemic illness affecting the retina, and 5) no dementia or other limitation that would prevent the patient from performing a self-test of visual function. Patients with AMD and DR were recruited at various disease stages, including those under active anti-vascular endothelial growth factor treatment. Patients with epiretinal membrane or pigment epithelial detachment were not excluded.	Healthy subjects were recruited from the normal subject database of the Retina Foundation of the Southwest.	no	not reported	Healthy senior volunteers	
Woods	2014	unclear	unclear	not reported	not reported	Participants diagnosed with Parkinson's disease	no	unclear	Parkinson tremor	Smartphone application that uses discrete wavelet transforms and support vector machines to discriminate between Parkinson's and Essential postural tremors / Triaxial, digital acceleration sensor. The 6 experimental tasks were: (1) tremor with eyes open (Vis+); (2) tremor with eyes closed (Vis−); (3) tremor while attending to the active tremor hand (Bubble); (4) tremor while attending to a laser target at 2 m (Laser2); (5) tremor while attending to a laser target at 1 m (Laser1); (6) tremor while not attending to the hand but while counting backwards by 3 (Counting).

Woods	2014	unclear	unclear	not reported	not reported	Participants with essential tremor	no	unclear	Essential Tremor	Smartphone application that uses discrete wavelet transforms and support vector machines to discriminate between Parkinson's and Essential postural tremors / Triaxial, digital acceleration sensor. The 6 experimental tasks were: (1) tremor with eyes open (Vis+); (2) tremor with eyes closed (Vis−); (3) tremor while attending to the active tremor hand (Bubble); (4) tremor while attending to a laser target at 2 m (Laser2); (5) tremor while attending to a laser target at 1 m (Laser1); (6) tremor while not attending to the hand but while counting backwards by 3 (Counting).
Kostikis	2014	Pilot study	Convenient	not reported	not reported	Subjects participating in this study were all Parkinson's disease patients recruited from the outpatient clinic of the 1st Department of Neurology at the Aristotle University of Thessaloniki.	no	not reported	Parkinson's Tremor	Using a smartphone-based platform, which processes the phone's accelerometr and gyroscope signals to detect and measure hand tremor. / In this work we are initially interested in resting tremor so we asked the subjects to "wear" an iPhone (fitted on a glove as in [3]) on top of their hand while sitting in a chair comfortably and resting both their hands on their lap, keeping that position for 30 seconds. The device was mounted on both their hands alternately, and each test was repeated twice for each subject.

Kostikis	2015	Cross-sectional study	Convenient	not reported	not reported	We recruited patients from the outpatient clinic of the first Department of Neurology at the Aristotle University of Thessaloniki. They all agreed to participate after they were offered a detailed explanation of the study's procedure and goals. All of them were right-handed, under L-DOPA treatment and suffering from Parkinson for more than two years.	no	not reported	Parkinson's Tremor	We attached an iPhone on our volunteer's hands using the same custom-made mounting glove]. It consists of a perforated case into which the phone "locks," and a wrist-supporting glove, both commercially available. The glove fits tightly on the volunteer's hand and the case is tightly sewn on the glove using nonelastic thread, ensuring the stability of the device on top of the hand. With the device attached, each participant had to maintain each of two prescribed postures for 30 s, while acceleration and gyroscope data were recorded by the phone. The two postures we used were the same ones used during the clinical evaluation, 1) "Extended," i.e., seated with both hands extended in front of the torso (Postural Tremor of the Hands, component 3.15 of the MDS-UPDRS) and 2) "Rest," i.e., seated with both hands placed on the arms of the chair (Rest Hand Tremor, component 3.17 of the MDS-UPDRS). The procedure was then repeated for the subject's other hand, in the same two postures. In the following, we will specify the combination of a patient's hand (Right of Left upper extremity) during each position as rR, rL, eR, and eL for rest-right, rest-left, extended-right, and extended-left, respectively.

Kostikis	2015	Cross-sectional study	Random	They were screened for several health conditions which could exclude them from the study, such as hypertension or any movement disorder. They were also notified of the procedure and the purpose of the study before agreeing to participate.	not reported	The control group for the study, contains healthy volunteers, none of whom suffered from a movement disorder, hypertension, or diabetes.	no	not reported	Age-matched healthy volunteers	We attached an iPhone on our volunteer's hands using the same custom-made mounting glove]. It consists of a perforated case into which the phone "locks," and a wrist-supporting glove, both commercially available. The glove fits tightly on the volunteer's hand and the case is tightly sewn on the glove using nonelastic thread, ensuring the stability of the device on top of the hand. With the device attached, each participant had to maintain each of two prescribed postures for 30 s, while acceleration and gyroscope data were recorded by the phone. The two postures we used were the same ones used during the clinical evaluation, 1) "Extended," i.e., seated with both hands extended in front of the torso (Postural Tremor of the Hands, component 3.15 of the MDS-UPDRS) and 2) "Rest," i.e., seated with both hands placed on the arms of the chair (Rest Hand Tremor, component 3.17 of the MDS-UPDRS). The procedure was then repeated for the subject's other hand, in the same two postures. In the following, we will specify the combination of a patient's hand (Right of Left upper extremity) during each position as rR, rL, eR, and eL for rest-right, rest-left, extended-right, and extended-left, respectively.

Kostikis	2015	Cross-sectional study	Convenient	not reported	not reported	We recruited patients from the outpatient clinic of the first Department of Neurology at the Aristotle University of Thessaloniki. They all agreed to participate after they were offered a detailed explanation of the study's procedure and goals. All of them were right-handed, under L-DOPA treatment and suffering from Parkinson for more than two years.	no	not reported	De novo Parkinson's disease: During the study, two of them were hospitalized overnight so that they could be tested in the morning before they received their medication, to approximate de novo patients.	We attached an iPhone on our volunteer's hands using the same custom-made mounting glove]. It consists of a perforated case into which the phone "locks," and a wrist-supporting glove, both commercially available. The glove fits tightly on the volunteer's hand and the case is tightly sewn on the glove using nonelastic thread, ensuring the stability of the device on top of the hand. With the device attached, each participant had to maintain each of two prescribed postures for 30 s, while acceleration and gyroscope data were recorded by the phone. The two postures we used were the same ones used during the clinical evaluation, 1) "Extended," i.e., seated with both hands extended in front of the torso (Postural Tremor of the Hands, component 3.15 of the MDS-UPDRS) and 2) "Rest," i.e., seated with both hands placed on the arms of the chair (Rest Hand Tremor, component 3.17 of the MDS-UPDRS). The procedure was then repeated for the subject's other hand, in the same two postures. In the following, we will specify the combination of a patient's hand (Right of Left upper extremity) during each position as rR, rL, eR, and eL for rest-right, rest-left, extended-right, and extended-left, respectively.

Arora	2015	Pilot study	Convenient	Importantly, this study did not include individuals with other parkinsonian or tremor disorders that may be more difficult to differentiate from Parkinson's disease.	not reported9	Individuals with Parkinson diagnosed clinically by a movement disorder specialist, were recruited from an academic movement disorder clinic (Johns Hopkins) and all participants provided informed consent.	no	1month / study duration: average 34.4 days	Parkinson's disease	(1) (voice test) say the sustained phonation 'aaah' for as long and as steadily as possible; (2) (posture test) stand upright unaided for thirty seconds; (3) (gait test) walk twenty steps forward, turn around, and return back to the starting position; (4) (finger tapping test) tap the screen alternately keeping a regular rhythm; and (5) (reaction time test) press and hold the on-screen button as soon as it appears and release it as soon as it disappears. The participants were asked to conduct the above specified tests four times daily: just before taking their first (morning) dose of levodopa (or in one case, rasagiline), one hour later, mid- afternoon, and before going to bed.
Arora	2015	Pilot study	Convenient	Known neurological disorder	not reported	Were recruited from an academic movement disorder clinic (Johns Hopkins) and all participants provided informed consent. Control participants were spouses, caregivers, relatives, or colleagues of an individual with PD	no	1month / study duration: average 34.4 days	Healthy volunteers (Control participants)	(1) (voice test) say the sustained phonation 'aaah' for as long and as steadily as possible; (2) (posture test) stand upright unaided for thirty seconds; (3) (gait test) walk twenty steps forward, turn around, and return back to the starting position; (4) (finger tapping test) tap the screen alternately keeping a regular rhythm; and (5) (reaction time test) press and hold the on-screen button as soon as it appears and release it as soon as it disappears. The participants were asked to conduct the above specified tests four times daily: just before taking their first (morning) dose of levodopa (or in one case, rasagiline), one hour later, mid- afternoon, and before going to bed.
Wadhawan	2011	Case-control	Convenient	Image artefacts	not reported	In this study, we use images from a large library of skin cancer images that were uploaded to the phone.	not reported	not reported	Melanoma	A support vector machine (SVM) is trained using a subset (training set) of the total images available, and the resulting classifier is used to determine whether the rest of the images (test set) are malignant or benign.

	Wadhawan	2011	Case-control	Convenient	Image artefacts	not reported	In this study, we used images from a large commercial library of skin cancer images annotated by expert dermatologists that were uploaded to the phone. Intra-observer and inter-observer agreement could be low for certain criteria. To demonstrate the feasibility of our automated system, we only chose images considered as low difficulty by the experts. There were 385 low difficulty images in the database and our segmentation methods could provide a satisfactory boundary for 347 (90.13%) of them.	not reported	not reported	Melanoma	A skin lesion image can be acquired using the smartphone camera (with or without an external attachment which can provide illumination and magnification) or can be loaded from the photo library to provide the diagnosis in real time. To identify a region of interest (ROI), an image is first converted to greyscale, and then fast median filtering for noise removal is performed, and followed by ISODATA segmentation, and several morphological operations.
	Daneault	2012	Observational	Random	not reported	not reported	not reported	not reported	not reported	Tremor in Essential tremor, Parkinson's disease and multiples Sclerosis	Algorithms for tremor recording and online analysis
	Doukas	2012	Case-Control	Convenient	not reported	not reported	For the initial evaluation of the system a dataset consisting of over 3000 skin lesion image sets of manually classified images has been utilized. The dataset contained about 800 images with melanoma, 600 with dysplastic nevus and the rest 1600 images with benign nevi.	not reported	not reported	Melanoma	SVM algorithm / automated skin lesion assessment system based on mobile technologies that can be used by patients for an early characterization of lesions and estimation for further assessment.
	Doukas	2012	Usability Questionnaire	Convenient	not reported	not reported	For the initial evaluation of the system a dataset consisting of over 3000 skin lesion image sets of manually classified images has been utilized. The dataset contained about 800 images with melanoma, 600 with dysplastic nevus and the rest 1600 images with benign nevi.	not reported	not reported	Melanoma	SVM algorithm / automated skin lesion assessment system based on mobile technologies that can be used by patients for an early characterization of lesions and estimation for further assessment.
	Fuadah	2015	Cross-sectional study	Convenient	not reported	not reported	In this research, we use 160 eyes images. We divided them into two categories: cataract (80 images) and normal (80 images). The training data set consist of 40 normal image and 40 cataract images.	not reported	not reported	Cataract	The interface of MCataract application consists of main activity and start activity that has functions, taking a picture or loading image from the gallery, cropping to obtain the pupil area as RoI, showing the pupil area obtained, and diagnosing the condition of eye image. K-Nearest Neighbor (k-NN) as classification method.

Do	2014	Cross-sectional study	Convenient	not reported	not reported	The database includes 81 color images provided by National Skin Center, Singapore	not reported	not reported	Melanoma	In this paper, we proposed (i) an efficient segmentation scheme by combining fast skin detection and fusion of two fast segmentation results; (ii) new features which efficiently capture the color variation and border irregularity of segmented lesion and (iii) an efficient criterion for selecting features. From selected features by proposed criterion, an automatic melanoma diagnosis system using mobile is developed.
Ramlakhan	2011	Case-control study	Convenient	not reported	not reported	As the lesion image dataset, 37 images of benign skin lesions, and 46 images of malignant lesions were obtained from various Internet sites.	not reported	not reported	Melanoma	The system consists of three major components: image segmentation, feature calculation, and classification. It is designed to run on a mobile device with a camera, such as a smartphone or a tablet PC. A skin lesion image is converted to a monochrome image for outline contour detection. Color and shape features of the lesion are extracted and used as input to a kNN classifier.
Raknim	2016	Case study	Convenient	not reported	not reported	not reported	not reported	1 year of data collection.	Parkinson's disease / changes in the walking patterns of PD patients	PDR-based method to continuously monitor and record the patient's gait characteristics using a smartphone / identify changes in the walking patterns of a patient
Arora	2011	Cohort study	Convenient	not reported	not reported	Individuals with Parkinson's diagnosed clinically by a movement disorder specialist and control participants were recruited from an academic movement disorder clinic (Johns Hopkins).	not reported	1 month	Parkinson's disease	Using tri-axial accelerometry data for self-administered tests of gait and postural sway recorded via consumer-grade smartphones to accurately distinguish Parkinson's disease participants from healthy controls.
Arora	2011	Cohort study	Convenient	not reported	not reported	Individuals with Parkinson's diagnosed clinically by a movement disorder specialist and control participants were recruited from an academic movement disorder clinic (Johns Hopkins).	not reported	1 month	Age- and gender-matched participants	Using tri-axial accelerometry data for self-administered tests of gait and postural sway recorded via consumer-grade smartphones to accurately distinguish Parkinson's disease participants from healthy controls.

Garca	2014	unclear	unclear	not reported	not reported	All the PD patients performed the tests in Hospital S. João, Oporto, Portugal, while they were waiting for the medical consult. Only the ones indicated by a neurologist helped in this project since a few requisites were necessary (motor capabilities and no medication related symptoms).	not reported	not reported	Parkinson's disease	The smartphone application has four different components: two of them require active interaction of the patient with the smartphone (spiral and tap analysis), one requires passive interaction (gait analysis) and the last component is used by the health care professional (simple questions with simple observation skills).
Garca	2014	unclear	unclear	not reported	not reported	The control group was assembled by visiting social centers or inviting seniors to Fraunhofer's facilities.	not reported	not reported	Healthy controls	The smartphone application has four different components: two of them require active interaction of the patient with the smartphone (spiral and tap analysis), one requires passive interaction (gait analysis) and the last component is used by the health care professional (simple questions with simple observation skills).

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.	3, 4
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	4
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.	5
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.	5
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.	appendix
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).	5
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.	5, 6
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	6
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.	5,6
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	6
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.	6

PRISMA 2009 Checklist

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).	-
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	n.a.
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.	7
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.	11,12, 14
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.	13, 14
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	13
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]).	n.a.
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).	16
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).	16
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	17
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.	6

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed1000097

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

BMJ Open

Evidence assessing the diagnostic performance of medical smartphone apps - A Systematic Review and exploratory meta-analysis

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-018280.R2
Article Type:	Research
Date Submitted by the Author:	22-Oct-2017
Complete List of Authors:	Buechi, Rahel; Cantonal Hospital Lucerne, Eye Clinic Faes, Livia; medignition Research Consultants, Bachmann, Lucas; medignition Research Consultants Thiel, Michael; Eye Clinic, Cantonal Hospital of Lucerne, Bodmer, Nicolas; medignition Research Consultants Schmid, Martin; Eye Clinic, Cantonal Hospital of Lucerne, Job, Oliver; Eye Clinic, Cantonal Hospital of Lucerne, Lienhard, Kenny; Faculty of Business and Economics, University of Lausanne, Information Systems
Primary Subject Heading:	Evidence based practice
Secondary Subject Heading:	Health informatics, Health services research
Keywords:	mobile health apps, evidence-based medicine, systematic review, diagnostic research

Evidence assessing the diagnostic performance of medical smartphone apps - A Systematic Review and exploratory meta-analysis

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

medignition Inc, Research Consultants, 8004 Zurich, Switzerland, Rahel Büchi research fellow, Livia Faes, research fellow, Lucas M Bachmann, professor of epidemiology, director, Nicolas S Bodmer Research Fellow, Eye Clinic, Cantonal Hospital of Lucerne Michael A Thiel, professor of ophthalmology, director, Martin K Schmid co-director, Oliver Job, head neuro-ophthalmology, Department of Information Systems, Faculty of Business and Economics, University of Lausanne Kenny R Lienhard, research fellow

Correspondence to: Lucas M. Bachmann: bachmann@medignition.ch

Rahel Büchi and Livia Faes contributed equally.

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 in-built 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 December 15th, 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 in-built sensors of a smartphone for diagnosis of disease. The methodological quality of 11 studies used in an exploratory meta-analysis was assessed with the QUADAS-2 tool and the reporting quality with the STARD statement. Sensitivity and specificity of studies reporting two-by-two tables were calculated and summarized.

Results We screened 3'296 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 1,048 subjects (758 with the target conditions and 290 healthy volunteers). Overall, the summary estimate for sensitivity was 0.82 (95 % confidence interval (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.

This systematic review was prospectively registered at PROSPERO under the number 42016033049.

Strength and Limitations of this Study

Strength

- A comprehensive literature search to retrieve the published evidence, applying stringent inclusion criteria and assessed the methodological quality of the studies systematically.

Limitations

- 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, nine percent 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 1,000 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 health care 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 dialog is of particular importance for patients with chronical 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 randomized 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 fact that most apps target at a diagnostic problem, it would be helpful to gauge the scientific basis of them. In this comprehensive systematic review we thus summarized the currently available papers assessing diagnostic properties of mobile health apps.

Methods

This review was conducted according to the PRISMA(10) statement recommendations.

Data Sources

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

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 in-built 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, i.e. the possibility to calculate sensitivity and specificity (O). We accepted all reference tests (C) used 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 2x2 table data that were summarized in the meta-analysis was made using the QUADAS-2 tool. Reporting quality was assessed using the STARD statement (22, 23). Quality assessment involved scrutinizing 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 standardized 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 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(24) pooling, at least 4 studies on the same target condition had to be available. Therefore, no separate analysis for health apps on Parkinson's disease, falling in chronic stroke patients and atrial fibrillation was possible.

All analyses were done using Stata 14.1 statistics software package (StataCorp LP,College Station, TX, USA).

Role of the Funding Source

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 sector, via an unrestricted research grant. LMB holds shares of medignition. LMB was responsible for the design and the statistical analysis of the study.

Results

Study selection

Electronic searches retrieved 4,010 records. After excluding duplicates, 3,296 records remained and were screened based on title and abstract. Subsequently, 3,209 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, thirty studies provided some clinical data (2, 11-21, 25-42). Details on these studies are available in the Appendix. Eleven studies reporting seventeen 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.

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 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 2015	Parkinson's Disease	Diagnostic Case Control Study	No	10	65.1 (9.8)	30%	Not reported.	Not reported.
Chadwick et al 2014	Melanoma	Diagnostic Case Control Study	No	15	Not applicable	Not applicable	Not reported.	Not reported.
Kostikis et al 2015	Parkinson's Disease	Diagnostic Case Control Study	No	23	78	52%	Not reported.	Not reported.
Lagido et al 2014	Atrial Fibrillation	Prospective Cohort Study	No	43	Not reported	Not reported	Not reported.	Not reported.
Maier et al 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 e.g. hair, images containing more than one lesion, incomplete imaged lesions, non- melanocytic lesions, two-point differences cases.
Ramlakhan et al 2011	Melanoma	Diagnostic Case Control Study	No	46	Not applicable	Not applicable	Not reported.	Not reported.
Takuya et al 2015	Falling in Chronic Stroke Patients	Diagnostic Case Control Study	No	11	70.5 (12.5)	Not reported	More than 12 months since stroke onset and ability to walk 16 meters independently with or without a single-point cane and/or an orthosis.	Severe cardiovascular, respiratory, musculoskeletal, or neurologic disorder other than stroke that affected gait performance; unable to understand the instructions because of communication problem or moderate to severe cognitive dysfunction (i.e., 5 or more errors on the Short Portable Mental Status Questionnaire [SPMSQ]); household ambulators walked only indoors or only mobilized during rehabilitation sessions.

Wadhawan et al 2011	Melanoma	Diagnostic Case Control Study	No	1,300	Not applicable	Not applicable	Not reported.	Image artefacts.
Wadhawan et al 2011	Melanoma	Diagnostic Case Control Study	No	347	Not applicable	Not applicable	Not reported.	Image artefacts.
Wolf et al 2013	Melanoma	Diagnostic Case Control Study	No	188	Not applicable	Not applicable	Images for which there was a clear histologic 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.

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 stroke patients (n=1)(20). The studies altogether involved 1,048 subjects, 758 subjects with the target condition and 290 healthy volunteers or controls. One paper reported on approximately 3,000 skin lesions of an unknown number of patients (28). The complete data abstraction of these studies is available in the **appendix**.

Eleven studies (11-21) that investigated 13 diagnostic health apps allowing the construction of 17 two-by-two tables qualified for the meta-analysis. 12 tables reported on diagnosis of melanoma, three on Parkinson's disease, one assessed falling in chronic stroke patients, and another atrial fibrillation.

Methodological Quality

A summary of the methodological quality is shown in **Table 2**

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 2014	*Yes*	*Yes*	*Yes*	*Yes*
Arora et al 2015	*Yes*	*Yes*	*No*	*Yes*
Chadwick et al 2014	*Yes*	*Yes*	*No*	*Yes*
Kostikis et al 2015	*Yes*	*Yes*	*No*	*Yes*
Lagido et al 2014	*Yes*	*Yes*	*Yes*	*Yes*
Maier et al 2015	*Yes*	*Yes*	*No*	*Yes*
Ramlakhan et al 2011	*Yes*	*Yes*	*Yes*	*Yes*
Takuya et al 2015	*Yes*	*Yes*	*Yes*	*Yes*
Wadhawan et al 2011	*Yes*	*Yes*	*No*	*Yes*
Wadhawan et al 2011	*Yes*	*Yes*	*Yes*	*Yes*
Wolf et al 2013	*Yes*	*Yes*	*Yes*	*Yes*

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 i.e. reasons for non-adherence, simplicity of use and difficulties and comprehensibility.

Exploratory analyses of diagnostic accuracy

The summary estimate for sensitivity was 82 percent (95 % confidence interval (CI); 0.56 to 0.94) and pooled specificity was 89 percent (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 chronic stroke patients 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 (HSROC) is seen in **Figure 2.**

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 2014	100.0%	100.0%	10	0	0	10	Not reported	Not reported	Parkinson's Disease
Arora et al 2015	100.0%	100.0%	10	0	0	10	Not reported	Not reported	Parkinson's Disease
Chadwick et al 2014	0.0%	100.0%	0	0	5	10	Not reported	Skin Scan	Melanoma
Chadwick et al 2014	0.0%	100.0%	0	0	4	5	Not reported	Mel App	Melanoma
Chadwick et al.2014	80.0%	20.0%	4	8	1	2	Not reported	Mole Detective	Melanoma
Chadwick et al 2014	80.0%	60.0%	4	4	1	6	Not reported	SpotMole Plus	Melanoma
Chadwick et al 2014	80.0%	60.0%	4	4	1	6	Not reported	Dr.Mole Premium	Melanoma
Kostikis et al 2015	82.6%	90.0%	19	2	4	18	0.94	Not reported	Parkinson's Disease
Lagido et al 2014	75.0%	97.1%	6	1	2	34	Not reported	Not reported	Atrial Fibrillation
Maier et al 2014	73.1%	83.1%	19	20	7	98	Not reported	Not reported	Melanoma
Ramlakhan et al 2011	91.3%	48.6%	42	19	4	18	Not reported	Not reported	Melanoma
Takuya et al 2015	72.7%	84.6%	8	2	3	11	0.75	Not reported	Falling in Chronic Stroke Patients
Wadhawan et al 2011	81.1%	86.2%	30	12	7	75	0.91	Skin Scan	Melanoma
Wadhawan et al 2011	87.3%	71.3%	96	68	14	169	Not reported	7-Point Checklist	Melanoma
Wolf et al 2013	70.0%	39.3%	42	74	18	48	Not reported	Not reported	Melanoma
Wolf et al 2013	68.3%	36.8%	41	79	19	46	Not reported	Not reported	Melanoma
Wolf et al 2013	6.7%	93.6%	4	7	56	103	Not reported	Not reported	Melanoma

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, 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 co-workers, who systematically summarized 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 randomized controlled trials to validate mental mobile health apps in clinical care (8). Likewise, Majeed-Ariss and co-authors, 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 over-inclusive 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 percent 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 found, had 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, they usually lead 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 US, 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.

Competing interests

This study was supported via an unrestricted research grant of medignition Inc. LMB holds shares of medignition. LMB was responsible for the design and the statistical analysis of the study. All other authors declare: no support from any organization for the submitted work; no financial relationships with any organizations that might have an interest in the submitted work in the previous three years, no other relationships or activities that could appear to have influenced the submitted work.

other languages, create adaptations, reprints, include within collections and create summaries, extracts and/or, abstracts of the Contribution, iii) create any other derivative work(s) based on the Contribution, iv) to exploit all subsidiary rights in the Contribution, v) the inclusion of electronic links from the Contribution to third party material where-ever it may be located; and, vi) license any third party to do any or all of the above.

Contributions

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 co-authors provided intellectual input and approved the final manuscript. LMB is the study guarantor.

Transparency declaration

LMB affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

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 https://www.tractica.com/resources/white-papers/enterprise-wearable-technology-case-studies/: Tractica; 2015 [

2. Kaiser PK. Emerging therapies for neovascular age-related macular degeneration: drugs in the pipeline. Ophthalmology. 2013;120(5 Suppl):S11-5.

3. Institute I. Patient Adoption of mHealth. http://www.imshealth.com/en/thought-leadership/ims-institute/reports/patient-adoption-of-mhealth; 2013.

4. https://www.ncbi.nlm.nih.gov/mesh/?term=mobile+application aa. last visited: 2/2/2017.

5. Misra S. New report finds more than 165,000 mobile health apps now available, takes close look at characteristics & use 2015 [Available from: http://www.imedicalapps.com/2015/09/ims-health-apps-report/#.

6. F. BSaN. Healthcare's digital future http://www.mckinsey.com/industries/healthcare-systems-and-services/our-insights/healthcares-digital-future: McKiinsey & Company; 2014 [

7. Majeed-Ariss R, Hall AG, McDonagh J, Fallon D, Swallow V. 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(2):e40.

8. Bakker D, Kazantzis N, Rickwood D, Rickard N. Mental Health Smartphone Apps: Review and Evidence-Based Recommendations for Future Developments. JMIR Mental Health. 2016;3(1):e7.

9. Anderson K. BO, Emmerton L. Mobile Health Apps to Facilitate Self-Care: A Qualitative Study of User Experiences. PLOS ONE. 2016.

10. Moher D LA, Tetzlaff J, Altman DG, The PRISMA Group. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. . BMJ 2009.

11. Arora S, Venkataraman V, Zhan A, Donohue S, Biglan K, Dorsey E, et al. Detecting and monitoring the symptoms of Parkinson's disease using smartphones: a pilot study. Parkinsonism & related disorders. 2015;21(6):650-3.

12. Chadwick X, Loescher, L. J., Janda, M., & Soyer, H. P. 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, Kotsavasiloglou C. A Smartphone-Based Tool for Assessing Parkinsonian Hand Tremor. IEEE journal of biomedical and health informatics. 2015;19(6):1835-42.

14. Lagido RB, Lobo J, Leite S, Sousa C, Ferreira L, Silva-Cardoso J, editors. 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: IEEE.

15. Maier T, Kulichova D, Schotten K, Astrid R, Ruzicka T, Berking C, et al. Accuracy of a smartphone application using fractal image analysis of pigmented moles compared to clinical diagnosis and histological result. Journal of the European Academy of Dermatology and Venereology. 2015;29(4):663-7.

16. Ramlakhan K, Shang Y, editors. A mobile automated skin lesion classification system. 2011 IEEE 23rd International Conference on Tools with Artificial Intelligence; 2011: IEEE.

17. Wadhawan T, Situ N, Lancaster K, Yuan X, Zouridakis G. SkinScan©: A PORTABLE LIBRARY FOR MELANOMA DETECTION ON HANDHELD DEVICES. Proceedings / IEEE International Symposium on Biomedical Imaging: from nano to macro IEEE International Symposium on Biomedical Imaging. 2011;2011:133-6.

18. Wadhawan T, Situ N, Rui H, Lancaster K, Yuan X, Zouridakis G. Implementation of the 7-Point Checklist for Melanoma Detection on Smart Handheld Devices. Conference Proceedings. 2011;2011:3180-3.

19. Wolf JA, Moreau JF, Akilov O, Patton T, English JC, Ho J, et al. Diagnostic inaccuracy of smartphone applications for melanoma detection. JAMA dermatology. 2013;149(4):422-6.

20. Takuya I. HT, Shigeru U. Accelerometry-based gait characteristics evaluated using a smartphone and their association with fall risk in people with chronic stroke. Journal of Stroke and Cerebrovascular Diseases. 2015;24(6):1305-11.

21. Arora S. VV, Donohue S., Biglan K.M., Dorsey E.R., Little M.A., editor High Accuracy Discrimination of Parkinson's Disease Participants from Healthy Controls Using Smartphones. 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2014 2014.

22. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011;155(8):529-36.

23. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, 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, Egger M, Deeks JJ, Shang A, et al. An empirical comparison of methods for meta-analysis of diagnostic accuracy showed hierarchical models are necessary. J Clin Epidemiol. 2008;61(11):1095-103.

25. Bove R, White CC, Giovannoni G, Glanz B, Golubchikov V, Hujol J, et al. Evaluating more naturalistic outcome measures A 1-year smartphone study in multiple sclerosis. Neurology-Neuroimmunology Neuroinflammation. 2015;2(6):e162.

26. Daneault J-F, Carignan B, Codère CÉ, Sadikot AF, Duval C. Using a smart phone as a standalone platform for detection and monitoring of pathological tremors. Frontiers in human neuroscience. 2013;6:357.

27. Do T-T, Zhou Y, Zheng H, Cheung N-M, Koh D, editors. Early melanoma diagnosis with mobile imaging. 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society; 2014: IEEE.

28. Doukas C. SP, Kiranoudis C.T., Maglogiannis I. Automated skin lesion assessment using mobile technologies and cloud platforms. Conf Proc IEEE Eng Med Biol Soc. 2012:2444-7.

29. Ellis RJ, Ng YS, Zhu S, Tan DM, Anderson B, Schlaug G, et al. A Validated Smartphone-Based Assessment of Gait and Gait Variability in Parkinson's Disease. PLoS one. 2015;10(10):e0141694.

30. Kostikis N, Hristu-Varsakelis D, Arnaoutoglou M, Kotsavasiloglou C, editors. Smartphone-based evaluation of parkinsonian hand tremor: Quantitative measurements vs clinical assessment scores. 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society; 2014: IEEE.

31. Lee J, Reyes BA, McManus DD, Maitas O, Chon KH. Atrial fibrillation detection using an iPhone 4S. IEEE Transactions on Biomedical Engineering. 2013;60(1):203-6.

32. McManus DD, Lee J, Maitas O, Esa N, Pidikiti R, Carlucci A, 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(3):315-9.

33. Nishiguchi S, Ito H, Yamada M, Yoshitomi H, Furu M, Ito T, et al. Self-assessment tool of disease activity of rheumatoid arthritis by using a smartphone application. TELEMEDICINE and e-HEALTH. 2014;20(3):235-40.

34. Printy BP, Renken LM, Herrmann JP, Lee I, Johnson B, Knight E, et al., editors. Smartphone application for classification of motor impairment severity in parkinson's disease. 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society; 2014: IEEE.

35. Raknim P, Lan K-c. Gait Monitoring for Early Neurological Disorder Detection Using Sensors in a Smartphone: Validation and a Case Study of Parkinsonism. Telemedicine and e-Health. 2016;22(1):75-81.

36. Shinohara A, Ito T, Ura T, Nishiguchi S, Ito H, Yamada M, et al., editors. 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 Y-Z, He Y-G, Mitzel G, Zhang S, Bartlett MB. Handheld Shape Discrimination Hyperacuity (hSDH) Test on a Mobile Device for Remote Monitoring of Visual Function in Maculopathy. Investigative Ophthalmology & Visual Science. 2012;53(14):2914-.

38. Winther C, Frisen L. Self-Testing of Vision in Age-Related Macula Degeneration: A Longitudinal Pilot Study Using a Smartphone-Based Rarebit Test. J Ophthalmol. 2015;2015:285463.

39. Woods AM, Nowostawski M, Franz EA, Purvis M. Parkinson's disease and essential tremor classification on mobile device. Pervasive and Mobile Computing. 2014;13:1-12.

40. Fuadah YN, Setiawan AW, Mengko TL, editors. Mobile cataract detection using optimal combination of statistical texture analysis. 2015 4th International Conference on Instrumentation, Communications, Information Technology, and Biomedical Engineering (ICICI-BME); 2015: IEEE.

41. Yamada M, Aoyama T, Mori S, Nishiguchi S, Okamoto K, Ito T, et al. Objective assessment of abnormal gait in patients with rheumatoid arthritis using a smartphone. Rheumatology international. 2012;32(12):3869-74.

42. Zhu S, Ellis RJ, Schlaug G, Ng YS, Wang Y, editors. Validating an iOS-based Rhythmic Auditory Cueing Evaluation (iRACE) for Parkinson's Disease. Proceedings of the 22nd ACM international conference on Multimedia; 2014: ACM.

43. 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.

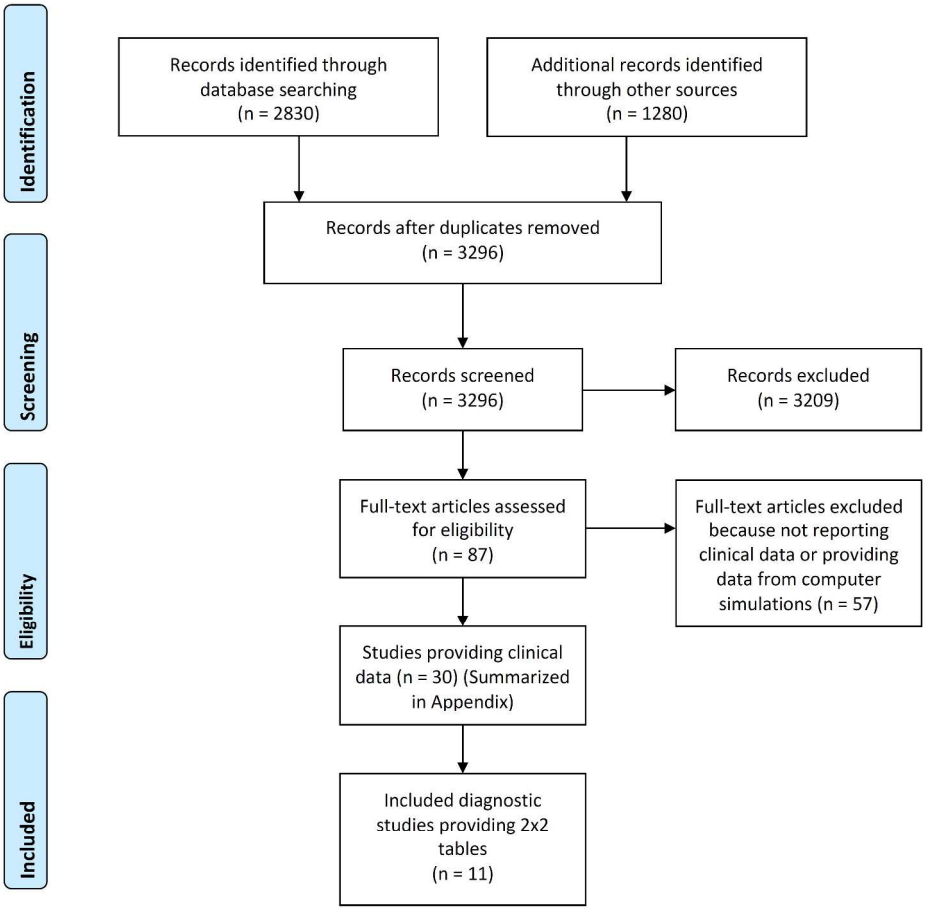
44. Bachmann LM, Coray R, Estermann P, Ter Riet G. Identifying diagnostic studies in MEDLINE: reducing the number needed to read. J Am Med Inform Assoc. 2002;9(6):653-8.

45. Fryback DG, Thornbury JR. The efficacy of diagnostic imaging. Med Decis Making. 1991;11(2):88-94.

Figure Legends

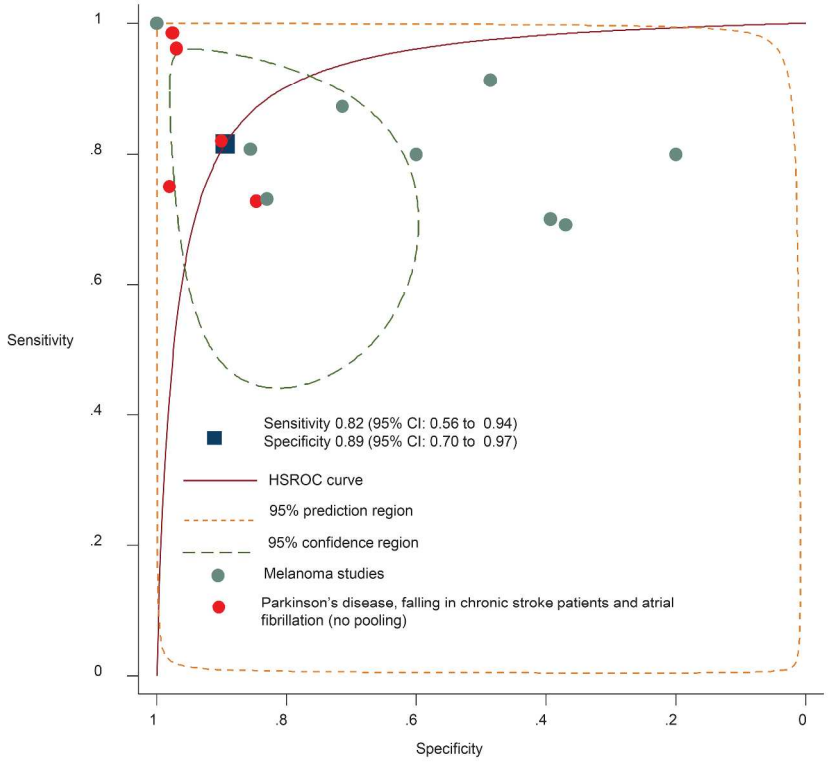
Figure 1. Flow Chart According to the PRISMA Statement

Figure 2 Hierarchical Summary Receiver Operating Characteristic Curve (HSROC)



Flow Chart According to the PRISMA Statement

391x366mm (300 x 300 DPI)



Hierarchical Summary Receiver Operating Characteristic Curve (HSROC)

391x366mm (300 x 300 DPI)

Appendix: Search strategies

SCOPUS

TITLE-ABS-KEY ((((**patient*** OR **outpatient*** OR **ambulant** OR **ambulatory**) W/3 **monitor***) OR **self-monitor***) AND (((**cell** OR **cellular** OR **mobile** OR **smart**) W/3 **phone**) OR (**smartphone*** OR **iphone***) OR ((**telemedicine** OR **ehealth** OR **mhealth** OR **telematic**) AND **mobile**)))

Medline (Ovid interface)

1 monitoring, ambulatory/
2 ((patient* or outpatient* or ambulant or ambulatory) adj3 monitor*).ti,ab.
3 self-monitor*.ti,ab.
4 1 or 2 or 3
5 exp cell phones/
6 ((cell or cellular or mobile or smart) adj3 phone).ti,ab.
7 (smartphone* or iPhone*).ti,ab.
8 or/5-7
9 exp telemedicine/ or exp telemetry/
10 (telemedicine or ehealth or mhealth or telematic).ti,ab.
11 9 or 10
12 mobile.ti,ab.
13 11 and 12
14 8 or 13
15 4 and 14

Business Source Premier

S1 ((patient OR outpatient OR ambulant OR ambulatory) N3 monitoring) OR selfmonitoring
S2 (((cell or cellular or mobile or smart) N3 phone)) OR ((smartphone* or iPhone*)) OR ((telemedicine or ehealth or mhealth or telematic) AND mobile))
S3 S1 AND S2

Science Citation Index

1 TS=(((patient* or outpatient* or ambulant or ambulatory) NEAR/3 monitor*) OR self-monitor*)
2 TS=(((cell or cellular or mobile or smart) NEAR/3 phone) OR (smartphone* or iPhone*) OR ((telemedicine or ehealth or mhealth or telematic) AND mobile))
3 #2 AND #1
4 PUBLICATION NAME: (JMIR MHEALTH "AND" UHEALTH OR JMIR MEDICAL INFORMATICS)
5 #4 AND #3
6 #2 AND #1 Refined by: RESEARCH AREAS: (MEDICAL INFORMATICS)
7 #6 AND #5
8 #6 OR #5

author	year	design	sampling	Exclusion criteria	Inclusion criteria	Recruitment	consecutive	duration	clinical condition	Type of measurement
Bove	2015	Cohort study, paired control-group, feasibility study, human observational trial	Convenient	not reported	Informed consent, Multiple sclerosis	Pairs consisting of 1 patient with demyelinating disease and 1 healthy cohabitant, were recruited at the Partners MS Center, a large referral clinical center in the northeastern United States.	no	1 year monitoring	Severity of MS	Questionnaires and visual tests should classify severity of MS
Bove	2015	Cohort study, paired control-group, feasibility study, human observational trial	Convenient	not reported	Informed consent	Pairs consisting of 1 patient with demyelinating disease and 1 healthy cohabitant, all aged 18–55 years, were recruited at the Partners MS Center, a large referral clinical center in the northeastern United States. Cohabitant pairs were recruited to control for common environment.	no	1 year monitoring	Healthy cohabitants of participating patients	Questionnaires and visual tests should provide a control for environmental factors
Kaiser	2013	Open-label, single-arm, multicentre study, Pilot study	Prospective	Patients were excluded if they had concomitant ocular disease in the study eye; neurologic impairment that would interfere with study assessments; use of systemic medications known to be toxic to the lens, retina, or optic nerve; or use of any other investigational agents within 60 days of screening.	Active CNV secondary to AMD (either newly diagnosed and treatment-naive or successfully treated with anti–vascular endothelial growth factor therapy for 1 year) in at least 1 eye, eligible for ranibizumab therapy, with best corrected visual acuity (BCVA) letter score 24 or higher (20/320 Snellen equivalent) by "Early Treatment Diabetic Retinopathy Study chart" at 4 m. Only one eye was required to fulfil entry criteria for a patient to enrol in the study; if both of the patient's eyes had CNV-AMD, then both were included in the analyses.	This study was conducted at 24 centers in the United States (NCT01542866).	no	16 weeks	Wet age-related macular degeneration	Distortion algorithm
Printy	2014	Cohort study, Pilot study	unclear	not reported	not reported	In a movement disorders clinic on the day of an outpatient appointment.	no	not reported	Parkinson's disease and severity	Quantification of the severity of Parkinson's motors symptoms using an application that collects kinematic data and extracts quantitative features using signal processing techniques, support vector machine classifier.
Lagido	2014	Diagnostic case-control	unclear	not reported	not reported	Samples were collected from heart failure patients at rest in Hospital S. Joao in Porto.	no	not reported	Atrial fibrillation	Detect heart rate and heart rate variability using a photoplethysmogram signal with the user's fingertip placed over the smartphone camera.

Maier	2014	Diagnostic case-control	Prospective	Images were excluded from evaluation in case of poor quality images, other elements in the image not belonging to the lesion e.g. hair, images containing more than one lesion, incomplete imaged lesions, non-melanocytic lesions, two-point differences cases (results in non-consecutive risk classes mainly due to inappropriate imaging angle or distance). The cases with an equal number of results in two consecutive risk classes, so-called tie cases (e.g. 1 high risk, 1 medium risk and 1 low risk results), were also excluded.	not reported	We included 195 melanocytic lesions in consecutive patients seen routinely for skin cancer screening at the Department of Dermatology, University Hospital of Munich, Germany after obtaining written informed consent.	yes	not reported	Melanoma	Risk assessment algorithm
Ellis	2015	Diagnostic case-control with age-matched healthy controls	unclear	not reported	A telephone questionnaire was first administered to screen out potential subjects who (1) are not within the age range of 40 to 85; (2) have any problems with their hearing; (3) are not able to walk independently without an aid; (4) have joint problems or other neurological, musculoskeletal or medical problems that can affect walking; (5) have sustained a fall within the past year that continues to affect their walking pattern; (6) have had surgery to implant a device (e.g., deep brain stimulation or pacemaker). Subjects who satisfied all six criteria were invited to participate in the study. Upon arrival at the testing location, four clinical assessments were administered.	All subjects were recruited through the Singapore General Hospital clinics. For safety reasons, the inclusion/exclusion criteria for the present study precluded patients with severe gait dysfunction; most patients in the present sample would be considered to have “moderately advanced” disease. Whether SmartMOVE would perform as well in the case of severe gait dysfunction (e.g., shuffling steps or frequent gait freezing episodes) is thus unknown.	no	not reported	Parkinson's disease	Smartphone’s inertial measurement unit to record gait movements during walking. / The accuracy of smartphone-based gait analysis (utilizing the smartphone’s built-in tri-axial accelerometer and gyroscope to calculate successive step times and step lengths) was validated against two heel contact–based measurement devices: heel-mounted footswitch sensors (to capture step times) and an instrumented pressure sensor mat (to capture step lengths).

Ellis	2015	Diagnostic case-control with age-matched healthy controls	unclear	not reported	A telephone questionnaire was first administered to screen out potential subjects who (1) are not within the age range of 40 to 85; (2) have any problems with their hearing; (3) are not able to walk independently without an aid; (4) have joint problems or other neurological, musculoskeletal or medical problems that can affect walking; (5) have sustained a fall within the past year that continues to affect their walking pattern; (6) have had surgery to implant a device (e.g., deep brain stimulation or pacemaker). Subjects who satisfied all six criteria were invited to participate in the study. Upon arrival at the testing location, four clinical assessments were administered.	All subjects were recruited through the Singapore General Hospital clinics. For safety reasons, the inclusion/exclusion criteria for the present study precluded patients with severe gait dysfunction; most patients in the present sample would be considered to have "moderately advanced" disease. Whether SmartMOVE would perform as well in the case of severe gait dysfunction (e.g., shuffling steps or frequent gait freezing episodes) is thus unknown.	no	not reported	Healthy subjects	Smartphone's inertial measurement unit to record gait movements during walking. / The accuracy of smartphone-based gait analysis (utilizing the smartphone's built-in tri-axial accelerometer and gyroscope to calculate successive step times and step lengths) was validated against two heel contact–based measurement devices: heel-mounted footswitch sensors (to capture step times) and an instrumented pressure sensor mat (to capture step lengths).
Nishiguchi	2014	Open-label, follow-up study	unclear	Other musculoskeletal disorders, cognitive disorders, Parkinson's disease, stroke, or unable to walk unassisted over 15m using walking aids. Patients with previous surgery in the lower extremities were also excluded.	Patients with rheumatoid arthritis defined by the 1987 or 2010 American College of Rheumatology criteria were included.	not reported	no	unclear	Rheumatoid arthritis, disease activity	The modified Health Assessment Questionnaire (mHAQ), self- assessed TJC (self-assessed tender joint count, out of 49 joints), and self-assessed SJC (self-assessed swollen joint count, out of 46 joints) were recorded on the smartphone application that we developed. The mHAQ, a self-reported measure of physical function to quantify functional disability. The mHAQ is expressed on a scale ranging from 0 to 3, where 0 = no disability and 3 = severe functional disability. Gait Analysis: The participants were instructed to walk along a 15-m walkway at their preferred speed. Trunk linear accelerations were measured by participants themselves with the smartphone as they walked on the walkway. The smartphone was kept adjacent to the L3 spinous process, which is close to where the body's center of mass is believed to be located during quiet standing using a semi-elastic belt.

Shinohara	2013	Feasibility study	unclear	Other musculoskeletal disorders, cognitive disorders, Parkinson's disease, stroke, or unable to walk over 10 m unassisted. Patients with previous surgery in the lower extremities were also excluded.	Patients with rheumatoid arthritis as defined by the American College of Rheumatology 1987 or 2010 criteria were included.	The participants were patients who attended the rheumatology outpatient clinic of Kyoto University Hospital.	no	Until the next hospital visit: mean duration, 35.0 ± 11.3 days	Rheumatoid arthritis, disease activity	Linear trunk accelerations are gathered by the participants' smartphones (kept in a waist pouch) as they walked for 10 seconds at their preferred speed. Peak frequency (PF), autocorrelation peak (AC), and coefficient of variance (CV) of the acceleration peak intervals. The PF value indicates the gait cycle, which is the time taken for 1 step. The AC value indicates the degree of gait balance, so a higher AC value indicates a greater degree of balance. The CV value indicates the degree of gait variability, i.e., the variability in the elapsed time between the first contacts of 2 consecutive footfalls. The modified Health Assessment Questionnaire (mHAQ), self- assessed TJC (self-assessed tender joint count, out of 49 joints), and self-assessed SJC (self-assessed swollen joint count, out of 46 joints) were recorded on the smartphone application that we developed. The mHAQ, a self-reported measure of physical function to quantify functional disability in RA. The mHAQ is expressed on a scale ranging from 0 to 3, where 0 = no disability and 3 = severe functional disability. General health condition and pain condition were recorded on the smartphone using a visual analogue scale (VAS). Gait Analysis: The participants were instructed to walk along a 15-m walkway at their preferred speed. Trunk linear accelerations were measured by participants themselves with the smartphone as they walked on the walkway. The smartphone was kept adjacent to the L3 spinous process, which is close to where the body's center of mass is believed to be located during quiet standing using a semi-elastic belt.

Yamada	2011	Cross-sectional study	unclear	We excluded participants based on the following exclusion criteria: other musculoskeletal disorders, cognitive disorders, Parkinson's disease, stroke, or unable to walk unassisted over 15 m using current walking aids.	Patients with rheumatoid arthritis defined by the American College of Rheumatology 1987 criteria were included.	This was a cross-sectional study performed between April 2011 and May 2011 in the rheumatology outpatient clinics of Kyoto University Hospital. A total of 39 RA patients (mean age, 65.9 ± 10.0 years) participated.	not reported	not reported	Rheumatoid arthritis, disease activity	The smartphone used in this study includes an acceleration sensor, a recording device, and a computer program for processing the acceleration signals. Trunk linear accelerations were measured using the smartphone while the subject walked on the walkway. The smartphone was attached to the L3 spinous process using a semi-elastic belt. Before measurements, the accelerometer of the smartphone was calibrated statically against gravity. The accelerometer of the smartphone sampled at 33 Hz. The recorded signals were analysed by the application developed in the android environment. Gait analysis: The participants were instructed to walk on a 20-m walk- way at their preferred speed. All participants wore their usual walking shoes, avoiding high heels and hard-soled shoes. The mid (10-m) walking time was measured using an electronic stopwatch.
Yamada	2011	Cross-sectional study	unclear	We excluded participants based on the following exclusion criteria: other musculoskeletal disorders, cognitive disorders, Parkinson's disease, stroke, or unable to walk unassisted over 15 m using current walking aids.	not reported	Twenty older individuals also took part in this experiment as control participants.	not reported	not reported	Healthy Control	The smartphone used in this study includes an acceleration sensor, a recording device, and a computer program for processing the acceleration signals. Trunk linear accelerations were measured using the smartphone while the subject walked on the walkway. The smartphone was attached to the L3 spinous process using a semi-elastic belt. Before measurements, the accelerometer of the smartphone was calibrated statically against gravity. The accelerometer of the smartphone sampled at 33 Hz. The recorded signals were analysed by the application developed in the android environment. Gait analysis: The participants were instructed to walk on a 20-m walk- way at their preferred speed. All participants wore their usual walking shoes, avoiding high heels and hard-soled shoes. The mid (10-m) walking time was measured using an electronic stopwatch.

Lee	2013	Case-control study, with-in	Prospective	not reported	not reported	Patients who presented for electrical cardioversion to the University of Massachusetts Medical Center (UMMC) cardiac electrophysiology laboratory were recruited by trained study personnel (McManus, Mathias)	no	not reported	Atrial fibrillation	Detect Atrial fibrillation (AF) and non-sinus rhythm (NSR) using a photoplethysmogram signal with the user's fingertip placed over the smartphone camera. AF and NSR detection is based on threshold values derived from the MIT-BIH AF and MIT-BIH NSR databases using statistical method RMSSD (Root mean square of successive differences).
Lee	2013	Case-control study, with-in	Prospective	not reported	not reported	Patients who presented for electrical cardioversion to the University of Massachusetts Medical Center (UMMC) cardiac electrophysiology laboratory were recruited by trained study personnel (McManus, Mathias)	no	not reported	Atrial fibrillation	Detect Atrial fibrillation (AF) and non-sinus rhythm (NSR) using a photoplethysmogram signal with the user's fingertip placed over the smartphone camera. AF and NSR detection is based on threshold values derived from the MIT-BIH AF and MIT-BIH NSR databases using statistical method RMSSD (Root mean square of successive differences).
Lee	2013	Case-control study, with-in	Prospective	not reported	not reported	Patients who presented for electrical cardioversion to the University of Massachusetts Medical Center (UMMC) cardiac electrophysiology laboratory were recruited by trained study personnel (McManus, Mathias)	no	not reported	Atrial fibrillation	Detect Atrial fibrillation (AF) and non-sinus rhythm (NSR) using a photoplethysmogram signal with the user's fingertip placed over the smartphone camera. AF and NSR detection is based on threshold values derived from the MIT-BIH AF and MIT-BIH NSR databases using statistical method RMSSD (Root mean square of successive differences).
Zhu	2014	unclear	unclear	None of the patients had a coexisting dementia (Mini-Mental State Examination score >24 points) or other diagnosed neurological impairments.	Patients diagnosed with idiopathic Parkinson's disease from Singapore General Hospital; All patients were under stable medication regiments for the preceding four weeks, and were tested at least 30 minutes after taking morning medications.	Patients diagnosed with idiopathic Parkinson's disease were recruited from Singapore General Hospital.	no	not reported	Parkinson's disease	To help “scale up” rhythmic auditory cueing (RAC) for wider distribution, we have developed an iOS-based Rhythmic Auditory Cueing Evaluation (iRACE) mobile application to deliver RAC and assess motor performance in PD patients. The touchscreen of the mobile device is used to assess motor timing during index finger tapping, and the device's built-in tri-axial accelerometer and gyro- scope to assess step time and step length during walking. Novel machine learning-based gait analysis algorithms have been developed for iRACE, including heel strike detection, step length quantification, and left-versus-right foot identification.

Wolf	2013	Diagnostic case-control	Convenient	Images that contained any identifiable features such as facial features, tattoos, or labels with patient information were either excluded or cropped to remove the identifiable features or information. Lesions with equivocal diagnoses such as “melanoma cannot be ruled out” or “atypical melanocytic proliferation” were excluded, as were Spitz nevi, pigmented spindle cell nevus of Reed, and other uncommon or equivocal lesions. We also excluded lesions with moderate or high-grade atypia given the controversy over their management. Two of the investigators then reviewed all images for image quality and omitted those that were of poor quality or resolution.	Images for which there was a clear histologic diagnosis rendered by a board-certified pathologist. The remaining images were stratified into one of the following categories: invasive melanoma, melanoma in situ, lentigo, benign nevus (including compound, junctional, and low-grade dysplastic nevi), dermatofibroma, seborrheic keratosis, and hemangioma. We only used close-up images of lesions.	not reported	no	not reported	Melanoma	Application 1 uses an automated algorithm to detect the border of the lesion, although it also allows manual input to confirm or change the detected border. It is the only application we tested that has this feature of user input for border detection. The application then analyses the image and gives an assessment of “problematic,” which we considered to be a positive test, “ok,” which we considered to be a negative test, or “error” if the image could not be assessed by the application. We categorized the latter group as unevaluable.
Wolf	2013	Diagnostic case-control	Convenient	Images that contained any identifiable features such as facial features, tattoos, or labels with patient information were either excluded or cropped to remove the identifiable features or information. Lesions with equivocal diagnoses such as “melanoma cannot be ruled out” or “atypical melanocytic proliferation” were excluded, as were Spitz nevi, pigmented spindle cell nevus of Reed, and other uncommon or equivocal lesions. We also excluded lesions with moderate or high-grade atypia given the controversy over their management. Two of the investigators then reviewed all images for image quality and omitted those that were of poor quality or resolution.	Images for which there was a clear histologic diagnosis rendered by a board-certified pathologist. The remaining images were stratified into one of the following categories: invasive melanoma, melanoma in situ, lentigo, benign nevus (including compound, junctional, and low-grade dysplastic nevi), dermatofibroma, seborrheic keratosis, and hemangioma. We only used close-up images of lesions.	not reported	no	not reported	Melanoma	Application 2 uses an automated algorithm to evaluate an image that has been uploaded by the user. The output given is either “melanoma,” which we considered to be a positive test, or “looks good” which we considered to be a negative test. If the image could not be analysed a message of “skin condition not found” was given and we considered the image unevaluable.

Wolf	2013	Diagnostic case-control	Convenient	Images that contained any identifiable features such as facial features, tattoos, or labels with patient information were either excluded or cropped to remove the identifiable features or information. Lesions with equivocal diagnoses such as “melanoma cannot be ruled out” or “atypical melanocytic proliferation” were excluded, as were Spitz nevi, pigmented spindle cell nevus of Reed, and other uncommon or equivocal lesions. We also excluded lesions with moderate or high-grade atypia given the controversy over their management. Two of the investigators then reviewed all images for image quality and omitted those that were of poor quality or resolution.	Images for which there was a clear histologic diagnosis rendered by a board-certified pathologist. The remaining images were stratified into one of the following categories: invasive melanoma, melanoma in situ, lentigo, benign nevus (including compound, junctional, and low-grade dysplastic nevi), dermatofibroma, seborrheic keratosis, and hemangioma. We only used close-up images of lesions.	not reported	no	not reported	Melanoma	Application 3 asks the user to upload an image to the application and then to position it within a box to ensure that the correct lesion is analysed. The output given by the application is “high risk,” which we considered to be a positive test, or “medium risk” or “low risk,” both of which we considered to be a negative test. The presence of a medium risk category in this application presented some difficulty in analysis as it was the only application tested that gave an intermediate output. Thus, we did perform sensitivity and specificity analysis with “medium risk” lesions counting as a positive test as well since it is not clear how a user would interpret such a result. Some lesions generated a message of “error” and these were considered unevaluable.
McManus	2013	Diagnostic case-control, within	Prospective	not reported	not reported	In a prospectively recruited cohort of participants undergoing cardioversion for atrial fibrillation	no	not reported	Atrial fibrillation	Mean square of successive differences (RMSSD): Our application acquired pulsatile signals by illuminating the fingertip using the standard iPhone lamp and recording video signal (30 frames/s) for 2 minutes. The signal was processed by averaging 50x50 green band pixels per frame. We interpolated the pulsatile signal to 30 Hz using a cubic spline algorithm followed by peak detection. As described in prior work, we use a peak detection algorithm that uses a filter bank with estimates of heart rate, variable cut-off frequencies, rank-order nonlinear filters, and decision logic as well as motion noise correction.

	McManus	2013	Diagnostic case-control, within	Prospective	not reported	not reported	In a prospectively recruited cohort of participants undergoing cardioversion for atrial fibrillation	no	not reported	Atrial fibrillation	Shannon entropy (ShE): Our application acquired pulsatile signals by illuminating the fingertip using the standard iPhone lamp and recording video signal (30 frames/s) for 2 minutes. The signal was processed by averaging 50x50 green band pixels per frame. We interpolated the pulsatile signal to 30 Hz using a cubic spline algorithm followed by peak detection. As described in prior work, we use a peak detection algorithm that uses a filter bank with estimates of heart rate, variable cut-off frequencies, rank-order nonlinear filters, and decision logic as well as motion noise correction.
	McManus	2013	Diagnostic case-control, within	Prospective	not reported	not reported	In a prospectively recruited cohort of participants undergoing cardioversion for atrial fibrillation	no	not reported	Atrial fibrillation	Combined statistical method: root mean square of successive RR difference (RMSSD/mean) and Shannon entropy (ShE): Our application acquired pulsatile signals by illuminating the fingertip using the standard iPhone lamp and recording video signal (30 frames/s) for 2 minutes. The signal was processed by averaging 50x50 green band pixels per frame. We interpolated the pulsatile signal to 30 Hz using a cubic spline algorithm followed by peak detection. As described in prior work, we use a peak detection algorithm that uses a filter bank with estimates of heart rate, variable cut-off frequencies, rank-order nonlinear filters, and decision logic as well as motion noise correction.

Takuya	2015	Cross-sectional study	Convenient	Exclusion criteria were (1) severe cardiovascular, respiratory, musculoskeletal, or neurologic disorder other than stroke that affected gait performance; (2) unable to understand the instructions because of communication problem or moderate to severe cognitive dysfunction (i.e., 5 or more errors on the Short Portable Mental Status Questionnaire [SPMSQ]); and (3) household ambulators walked only indoors or only mobilized during rehabilitation sessions.	Inclusion criteria were (1) more than 12 months since stroke onset and (2) ability to walk 16 m independently with or without a single point cane and/or an orthosis.	Community-dwelling adults with chronic stroke receiving day care services were recruited and screened for inclusion and exclusion criteria. conducted in 2 day care centers for elderly adults in Saitama, Japan	no		not reported	Chronic stroke, falls	Gait characteristics
Takuya	2015	Cross-sectional study	Convenient	Exclusion criteria were (1) severe cardiovascular, respiratory, musculoskeletal, or neurologic disorder other than stroke that affected gait performance; (2) unable to understand the instructions because of communication problem or moderate to severe cognitive dysfunction (ie, 5 or more errors on the Short Portable Mental Status Questionnaire [SPMSQ]); and (3) household ambulators walked only indoors or only mobilized during rehabilitation sessions.	Inclusion criteria were (1) more than 12 months since stroke onset and (2) ability to walk 16 m independently with or without a single point cane and/or an orthosis.	Community-dwelling adults with chronic stroke receiving day care services were recruited and screened for inclusion and exclusion criteria. conducted in 2 day care centers for elderly adults in Saitama, Japan	no		not reported	chronic stroke, no falls	Gait characteristics
Chadwick	2014	Case-control study	Convenient	not reported	not reported	The first step in the app testing was the selection of 15 high-quality, full resolution, digital images (.jpg) of melanocytic skin lesions of varying risk (5 melanomas, 10 benign nevi) from the image archive of the study dermatology expert (HPS) to assess the accuracy of the app analysis software.	no		not reported	Melanoma	SkinScan - Asymmetry/Border: Analysis by inbuilt pattern recognition software. / Color: Analysis by input comparison algorithms. / Diameter: no input. / Evolution: no input

Chadwick	2014	Case-control study	Convenient	not reported	not reported	The first step in the app testing was the selection of 15 high-quality, full resolution, digital images (.jpg) of melanocytic skin lesions of varying risk (5 melanomas, 10 benign nevi) from the image archive of the study dermatology expert (HPS) to assess the accuracy of the app analysis software.	no	not reported	Melanoma	MelApp - Asymmetry/Border: Analysis by inbuilt pattern recognition software. Area could be limited to focus the analysis. / Color: Analysis by input comparison algorithms. / Diameter: manual sliding scale input. / Evolution: manual sliding scale input
Chadwick	2014	Case-control study	Convenient	not reported	not reported	The first step in the app testing was the selection of 15 high-quality, full resolution, digital images (.jpg) of melanocytic skin lesions of varying risk (5 melanomas, 10 benign nevi) from the image archive of the study dermatology expert (HPS) to assess the accuracy of the app analysis software.	no	not reported	Melanoma	Mole Detective - Asymmetry/Border: Analysis by inbuilt pattern recognition software. / Color: Analysis by input comparison algorithms. / Diameter: Manual input of <6mm, -mm, or >6mm. / Evolution: No input for analysis. Reminder can be set for future use.
Chadwick	2014	Case-control study	Convenient	not reported	not reported	The first step in the app testing was the selection of 15 high-quality, full resolution, digital images (.jpg) of melanocytic skin lesions of varying risk (5 melanomas, 10 benign nevi) from the image archive of the study dermatology expert (HPS) to assess the accuracy of the app analysis software.	no	not reported	Melanoma	Spot Mole Plus - Asymmetry/Border: Analysis by inbuilt pattern recognition software. Manual adjustment of lesion border available/ Color: Analysis by input comparison algorithms. / Diameter: Manual input of numeric value. / Evolution: no input for past history of change. Can perform serial analysis of lesion images.
Chadwick	2014	Case-control study	Convenient	not reported	not reported	The first step in the app testing was the selection of 15 high-quality, full resolution, digital images (.jpg) of melanocytic skin lesions of varying risk (5 melanomas, 10 benign nevi) from the image archive of the study dermatology expert (HPS) to assess the accuracy of the app analysis software.	no	not reported	Melanoma	Dr. Mole Premium - Asymmetry/Border: Analysis by inbuilt pattern recognition software. Comparison of lesion quadrants for asymmetry./ Color: Analysis by input comparison algorithms. / Diameter: Manual sliding scale input. / Evolution: manual input of "none", "slow" & "fast" with time frames for each.

Winther	2015	Longitudinal Pilot study	Convenient	Exclusion criteria were the bilateral presence of currently inactive lesions judged to run a small risk of recurrence, the presence of additional, non- maculopathic causes of visual loss, and inability to participate in conventional acuity testing. Twenty-eight patients partook in the study. Those who had the same type of lesions in both eyes (active or inactive) provided results from the least involved eye only whereas those who had different types of lesions provided results from both eyes.	not reported	Patients were recruited from the wet age-related macular degeneration treatment programme of the Retina Unit at the Sahlgrenska University Hospital, a tertiary-care center.	no	Average monitoring of average of 30 weeks. Another time they report an average of 39	Wet age-related macular degeneration	For formal analysis, all MBT plots were carefully evaluated by subjective inspection, epoch by epoch. Epochs showing trends of decreasing scores, or increasing variation, or both, were rated worse. Epochs showing the opposite evolution were rated better. All other epochs were rated stable. ETDRS results were rated similarly, using direct numerical comparisons of scores. The outcomes of the clinical exam nations, which included biomicroscopy and scrutiny of OCT parameters and maps, were summarized in the same manner.
Winther	2015	Longitudinal Pilot study	Convenient	Exclusion criteria were the bilateral presence of currently inactive lesions judged to run a small risk of recurrence, the presence of additional, non- maculopathic causes of visual loss, and inability to participate in conventional acuity testing. Twenty-eight patients partook in the study. Those who had the same type of lesions in both eyes (active or inactive) provided results from the least involved eye only whereas those who had different types of lesions provided results from both eyes.	not reported	Twenty control subjects were recruited primarily from patients' relatives or other accompanying persons.	no	Average monitoring of average of 30 weeks. Another time they report an average of 39	Healthy relatives or accompanying	not reported

Wang	2013	Cross-sectional study	Convenient	not reported	1) AMD or DR with corrected Early Treatment of Diabetic Retinopathy Study (ETDRS) VA of 20/100 or better in at least one eye, 2) ophthalmic evaluation by retina specialists with clinical and spectral-domain (SD)-OCT documentation, 3) no retinal pathology other than AMD or DR, 4) no concurrent systemic illness affecting the retina, and 5) no dementia or other limitation that would prevent the patient from performing a self-test of visual function. Patients with AMD and DR were recruited at various disease stages, including those under active anti-vascular endothelial growth factor treatment. Patients with epiretinal membrane or pigment epithelial detachment were not excluded.	Patients with AMD and DR were recruited from the clinic of the Department of Ophthalmology, UT Southwestern Medical Center.	no	not reported	Age-related macular degeneration

Wang	2013	Cross-sectional study	unclear	not reported	1) AMD or DR with corrected Early Treatment of Diabetic Retinopathy Study (ETDRS) VA of 20/100 or better in at least one eye, 2) ophthalmic evaluation by retina specialists with clinical and spectral-domain (SD)-OCT documentation, 3) no retinal pathology other than AMD or DR, 4) no concurrent systemic illness affecting the retina, and 5) no dementia or other limitation that would prevent the patient from performing a self-test of visual function. Patients with AMD and DR were recruited at various disease stages, including those under active anti-vascular endothelial growth factor treatment. Patients with epiretinal membrane or pigment epithelial detachment were not excluded.	Patients with AMD and DR were recruited from the clinic of the Department of Ophthalmology, UT Southwestern Medical Center.	no	not reported	Diabetic retinopathy

Wang	2013	Cross-sectional study	unclear	not reported	1) AMD or DR with corrected Early Treatment of Diabetic Retinopathy Study (ETDRS) VA of 20/100 or better in at least one eye, 2) ophthalmic evaluation by retina specialists with clinical and spectral-domain (SD)-OCT documentation, 3) no retinal pathology other than AMD or DR, 4) no concurrent systemic illness affecting the retina, and 5) no dementia or other limitation that would prevent the patient from performing a self-test of visual function. Patients with AMD and DR were recruited at various disease stages, including those under active anti-vascular endothelial growth factor treatment. Patients with epiretinal membrane or pigment epithelial detachment were not excluded.	Healthy subjects were recruited from the normal subject database of the Retina Foundation of the Southwest.	no	not reported	Healthy senior volunteers	
Woods	2014	unclear	unclear	not reported	not reported	Participants diagnosed with Parkinson's disease	no	unclear	Parkinson tremor	Smartphone application that uses discrete wavelet transforms and support vector machines to discriminate between Parkinson's and Essential postural tremors / Triaxial, digital acceleration sensor. The 6 experimental tasks were: (1) tremor with eyes open (Vis+); (2) tremor with eyes closed (Vis−); (3) tremor while attending to the active tremor hand (Bubble); (4) tremor while attending to a laser target at 2 m (Laser2); (5) tremor while attending to a laser target at 1 m (Laser1); (6) tremor while not attending to the hand but while counting backwards by 3 (Counting).

Woods	2014	unclear	unclear	not reported	not reported	Participants with essential tremor	no	unclear	Essential Tremor	Smartphone application that uses discrete wavelet transforms and support vector machines to discriminate between Parkinson's and Essential postural tremors / Triaxial, digital acceleration sensor. The 6 experimental tasks were: (1) tremor with eyes open (Vis+); (2) tremor with eyes closed (Vis−); (3) tremor while attending to the active tremor hand (Bubble); (4) tremor while attending to a laser target at 2 m (Laser2); (5) tremor while attending to a laser target at 1 m (Laser1); (6) tremor while not attending to the hand but while counting backwards by 3 (Counting).
Kostikis	2014	Pilot study	Convenient	not reported	not reported	Subjects participating in this study were all Parkinson's disease patients recruited from the outpatient clinic of the 1st Department of Neurology at the Aristotle University of Thessaloniki.	no	not reported	Parkinson's Tremor	Using a smartphone-based platform, which processes the phone's accelerometr and gyroscope signals to detect and measure hand tremor. / In this work we are initially interested in resting tremor so we asked the subjects to "wear" an iPhone (fitted on a glove as in [3]) on top of their hand while sitting in a chair comfortably and resting both their hands on their lap, keeping that position for 30 seconds. The device was mounted on both their hands alternately, and each test was repeated twice for each subject.

Kostikis	2015	Cross-sectional study	Convenient	not reported	not reported	We recruited patients from the outpatient clinic of the first Department of Neurology at the Aristotle University of Thessaloniki. They all agreed to participate after they were offered a detailed explanation of the study's procedure and goals. All of them were right-handed, under L-DOPA treatment and suffering from Parkinson for more than two years.	no	not reported	Parkinson's Tremor	We attached an iPhone on our volunteer's hands using the same custom-made mounting glove]. It consists of a perforated case into which the phone "locks," and a wrist-supporting glove, both commercially available. The glove fits tightly on the volunteer's hand and the case is tightly sewn on the glove using nonelastic thread, ensuring the stability of the device on top of the hand. With the device attached, each participant had to maintain each of two prescribed postures for 30 s, while acceleration and gyroscope data were recorded by the phone. The two postures we used were the same ones used during the clinical evaluation, 1) "Extended," i.e., seated with both hands extended in front of the torso (Postural Tremor of the Hands, component 3.15 of the MDS-UPDRS) and 2) "Rest," i.e., seated with both hands placed on the arms of the chair (Rest Hand Tremor, component 3.17 of the MDS-UPDRS). The procedure was then repeated for the subject's other hand, in the same two postures. In the following, we will specify the combination of a patient's hand (Right of Left upper extremity) during each position as rR, rL, eR, and eL for rest-right, rest-left, extended-right, and extended-left, respectively.

Kostikis	2015	Cross-sectional study	Random	They were screened for several health conditions which could exclude them from the study, such as hypertension or any movement disorder. They were also notified of the procedure and the purpose of the study before agreeing to participate.	not reported	The control group for the study, contains healthy volunteers, none of whom suffered from a movement disorder, hypertension, or diabetes.	no	not reported	Age-matched healthy volunteers	We attached an iPhone on our volunteer's hands using the same custom-made mounting glove]. It consists of a perforated case into which the phone "locks," and a wrist-supporting glove, both commercially available. The glove fits tightly on the volunteer's hand and the case is tightly sewn on the glove using nonelastic thread, ensuring the stability of the device on top of the hand. With the device attached, each participant had to maintain each of two prescribed postures for 30 s, while acceleration and gyroscope data were recorded by the phone. The two postures we used were the same ones used during the clinical evaluation, 1) "Extended," i.e., seated with both hands extended in front of the torso (Postural Tremor of the Hands, component 3.15 of the MDS-UPDRS) and 2) "Rest," i.e., seated with both hands placed on the arms of the chair (Rest Hand Tremor, component 3.17 of the MDS-UPDRS). The procedure was then repeated for the subject's other hand, in the same two postures. In the following, we will specify the combination of a patient's hand (Right of Left upper extremity) during each position as rR, rL, eR, and eL for rest-right, rest-left, extended-right, and extended-left, respectively.

Kostikis	2015	Cross-sectional study	Convenient	not reported	not reported	We recruited patients from the outpatient clinic of the first Department of Neurology at the Aristotle University of Thessaloniki. They all agreed to participate after they were offered a detailed explanation of the study's procedure and goals. All of them were right-handed, under L-DOPA treatment and suffering from Parkinson for more than two years.	no	not reported	De novo Parkinson's disease: During the study, two of them were hospitalized overnight so that they could be tested in the morning before they received their medication, to approximate de novo patients.	We attached an iPhone on our volunteer's hands using the same custom-made mounting glove]. It consists of a perforated case into which the phone "locks," and a wrist-supporting glove, both commercially available. The glove fits tightly on the volunteer's hand and the case is tightly sewn on the glove using nonelastic thread, ensuring the stability of the device on top of the hand. With the device attached, each participant had to maintain each of two prescribed postures for 30 s, while acceleration and gyroscope data were recorded by the phone. The two postures we used were the same ones used during the clinical evaluation, 1) "Extended," i.e., seated with both hands extended in front of the torso (Postural Tremor of the Hands, component 3.15 of the MDS-UPDRS) and 2) "Rest," i.e., seated with both hands placed on the arms of the chair (Rest Hand Tremor, component 3.17 of the MDS-UPDRS). The procedure was then repeated for the subject's other hand, in the same two postures. In the following, we will specify the combination of a patient's hand (Right of Left upper extremity) during each position as rR, rL, eR, and eL for rest-right, rest-left, extended-right, and extended-left, respectively.

Arora	2015	Pilot study	Convenient	Importantly, this study did not include individuals with other parkinsonian or tremor disorders that may be more difficult to differentiate from Parkinson's disease.	not reported9	Individuals with Parkinson diagnosed clinically by a movement disorder specialist, were recruited from an academic movement disorder clinic (Johns Hopkins) and all participants provided informed consent.	no	1month / study duration: average 34.4 days	Parkinson's disease	(1) (voice test) say the sustained phonation 'aaah' for as long and as steadily as possible; (2) (posture test) stand upright unaided for thirty seconds; (3) (gait test) walk twenty steps forward, turn around, and return back to the starting position; (4) (finger tapping test) tap the screen alternately keeping a regular rhythm; and (5) (reaction time test) press and hold the on-screen button as soon as it appears and release it as soon as it disappears. The participants were asked to conduct the above specified tests four times daily: just before taking their first (morning) dose of levodopa (or in one case, rasagiline), one hour later, mid- afternoon, and before going to bed.
Arora	2015	Pilot study	Convenient	Known neurological disorder	not reported	Were recruited from an academic movement disorder clinic (Johns Hopkins) and all participants provided informed consent. Control participants were spouses, caregivers, relatives, or colleagues of an individual with PD	no	1month / study duration: average 34.4 days	Healthy volunteers (Control participants)	(1) (voice test) say the sustained phonation 'aaah' for as long and as steadily as possible; (2) (posture test) stand upright unaided for thirty seconds; (3) (gait test) walk twenty steps forward, turn around, and return back to the starting position; (4) (finger tapping test) tap the screen alternately keeping a regular rhythm; and (5) (reaction time test) press and hold the on-screen button as soon as it appears and release it as soon as it disappears. The participants were asked to conduct the above specified tests four times daily: just before taking their first (morning) dose of levodopa (or in one case, rasagiline), one hour later, mid- afternoon, and before going to bed.
Wadhawan	2011	Case-control	Convenient	Image artefacts	not reported	In this study, we use images from a large library of skin cancer images that were uploaded to the phone.	not reported	not reported	Melanoma	A support vector machine (SVM) is trained using a subset (training set) of the total images available, and the resulting classifier is used to determine whether the rest of the images (test set) are malignant or benign.

	Wadhawan	2011	Case-control	Convenient	Image artefacts	not reported	In this study, we used images from a large commercial library of skin cancer images annotated by expert dermatologists that were uploaded to the phone. Intra-observer and inter-observer agreement could be low for certain criteria. To demonstrate the feasibility of our automated system, we only chose images considered as low difficulty by the experts. There were 385 low difficulty images in the database and our segmentation methods could provide a satisfactory boundary for 347 (90.13%) of them.	not reported	not reported	Melanoma	A skin lesion image can be acquired using the smartphone camera (with or without an external attachment which can provide illumination and magnification) or can be loaded from the photo library to provide the diagnosis in real time. To identify a region of interest (ROI), an image is first converted to greyscale, and then fast median filtering for noise removal is performed, and followed by ISODATA segmentation, and several morphological operations.
	Daneault	2012	Observational	Random	not reported	not reported	not reported	not reported	not reported	Tremor in Essential tremor, Parkinson's disease and multiples Sclerosis	Algorithms for tremor recording and online analysis
	Doukas	2012	Case-Control	Convenient	not reported	not reported	For the initial evaluation of the system a dataset consisting of over 3000 skin lesion image sets of manually classified images has been utilized. The dataset contained about 800 images with melanoma, 600 with dysplastic nevus and the rest 1600 images with benign nevi.	not reported	not reported	Melanoma	SVM algorithm / automated skin lesion assessment system based on mobile technologies that can be used by patients for an early characterization of lesions and estimation for further assessment.
	Doukas	2012	Usability Questionnaire	Convenient	not reported	not reported	For the initial evaluation of the system a dataset consisting of over 3000 skin lesion image sets of manually classified images has been utilized. The dataset contained about 800 images with melanoma, 600 with dysplastic nevus and the rest 1600 images with benign nevi.	not reported	not reported	Melanoma	SVM algorithm / automated skin lesion assessment system based on mobile technologies that can be used by patients for an early characterization of lesions and estimation for further assessment.
	Fuadah	2015	Cross-sectional study	Convenient	not reported	not reported	In this research, we use 160 eyes images. We divided them into two categories: cataract (80 images) and normal (80 images). The training data set consist of 40 normal image and 40 cataract images.	not reported	not reported	Cataract	The interface of MCataract application consists of main activity and start activity that has functions, taking a picture or loading image from the gallery, cropping to obtain the pupil area as RoI, showing the pupil area obtained, and diagnosing the condition of eye image. K-Nearest Neighbor (k-NN) as classification method.

Do	2014	Cross-sectional study	Convenient	not reported	not reported	The database includes 81 color images provided by National Skin Center, Singapore	not reported	not reported	Melanoma	In this paper, we proposed (i) an efficient segmentation scheme by combining fast skin detection and fusion of two fast segmentation results; (ii) new features which efficiently capture the color variation and border irregularity of segmented lesion and (iii) an efficient criterion for selecting features. From selected features by proposed criterion, an automatic melanoma diagnosis system using mobile is developed.
Ramlakhan	2011	Case-control study	Convenient	not reported	not reported	As the lesion image dataset, 37 images of benign skin lesions, and 46 images of malignant lesions were obtained from various Internet sites.	not reported	not reported	Melanoma	The system consists of three major components: image segmentation, feature calculation, and classification. It is designed to run on a mobile device with a camera, such as a smartphone or a tablet PC. A skin lesion image is converted to a monochrome image for outline contour detection. Color and shape features of the lesion are extracted and used as input to a kNN classifier.
Raknim	2016	Case study	Convenient	not reported	not reported	not reported	not reported	1 year of data collection.	Parkinson's disease / changes in the walking patterns of PD patients	PDR-based method to continuously monitor and record the patient's gait characteristics using a smartphone / identify changes in the walking patterns of a patient
Arora	2011	Cohort study	Convenient	not reported	not reported	Individuals with Parkinson's diagnosed clinically by a movement disorder specialist and control participants were recruited from an academic movement disorder clinic (Johns Hopkins).	not reported	1 month	Parkinson's disease	Using tri-axial accelerometry data for self-administered tests of gait and postural sway recorded via consumer-grade smartphones to accurately distinguish Parkinson's disease participants from healthy controls.
Arora	2011	Cohort study	Convenient	not reported	not reported	Individuals with Parkinson's diagnosed clinically by a movement disorder specialist and control participants were recruited from an academic movement disorder clinic (Johns Hopkins).	not reported	1 month	Age- and gender-matched participants	Using tri-axial accelerometry data for self-administered tests of gait and postural sway recorded via consumer-grade smartphones to accurately distinguish Parkinson's disease participants from healthy controls.

Garca	2014	unclear	unclear	not reported	not reported	All the PD patients performed the tests in Hospital S. João, Oporto, Portugal, while they were waiting for the medical consult. Only the ones indicated by a neurologist helped in this project since a few requisites were necessary (motor capabilities and no medication related symptoms).	not reported	not reported	Parkinson's disease	The smartphone application has four different components: two of them require active interaction of the patient with the smartphone (spiral and tap analysis), one requires passive interaction (gait analysis) and the last component is used by the health care professional (simple questions with simple observation skills).
Garca	2014	unclear	unclear	not reported	not reported	The control group was assembled by visiting social centers or inviting seniors to Fraunhofer's facilities.	not reported	not reported	Healthy controls	The smartphone application has four different components: two of them require active interaction of the patient with the smartphone (spiral and tap analysis), one requires passive interaction (gait analysis) and the last component is used by the health care professional (simple questions with simple observation skills).

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.	3, 4
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	4
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.	5
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.	5
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.	appendix
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).	5
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.	5, 6
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	6
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.	5,6
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	6
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.	6

PRISMA 2009 Checklist

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).	-
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	n.a.
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.	7
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.	11,12, 14
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.	13, 14
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	13
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]).	n.a.
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).	16
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).	16
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	17
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.	6

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed1000097

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