Applications of Machine Learning in Real-life Digital Health Interventions: A Review of the Literature
Andreas K Triantafyllidis\textsuperscript{a,b}, and Athanasios Tsanas\textsuperscript{c,d}

\textsuperscript{a} Information Technologies Institute, Centre for Research and Technology Hellas (CERTH), Greece

\textsuperscript{b} Lab of Computing, Medical Informatics and Biomedical Imaging Technologies, School of Medicine, Aristotle University of Thessaloniki, Greece

\textsuperscript{c} Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, UK

\textsuperscript{d} Oxford Centre for Industrial and Applied Mathematics, University of Oxford, UK

\textbf{Address for Correspondence:}
Dr Andreas Triantafyllidis
Information Technologies Institute
Centre for Research and Technology Hellas
Thessaloniki 57001, GR.
Tel.: +30 2311 257610
E-mail: atriand@iti.gr
Abstract

**Background:** Machine learning has attracted considerable research interest towards developing smart digital health interventions. These interventions have the potential to revolutionize healthcare and lead to substantial outcomes for patients and medical professionals.

**Objective:** Provide a literature review of applications of machine learning in real-life digital health interventions, aiming to improve the understanding of researchers, clinicians, engineers, and policy makers, in developing robust and impactful data-driven interventions in the healthcare domain.

**Methods:** The bibliographic databases of PubMed and Scopus were searched with terms related to machine learning, to identify real-life studies of digital health interventions incorporating machine learning algorithms. We grouped those interventions according to their target (i.e., target condition), the study design, the number of enrolled participants, the follow-up duration, the primary outcome and whether this had been statistically significant, the machine learning algorithms used in the intervention, as well as the outcome of the algorithms (e.g., prediction).

**Results:** Our literature search identified 8 interventions incorporating machine learning in a real-life research setting, of which 3 (37%) were evaluated in a randomized controlled trial, and 5 (63%) in a pilot or experimental single-group study. The interventions targeted depression prediction and management, speech recognition for people with speech disabilities, self-efficacy for weight loss, detection of changes in biopsychosocial condition of patients with multiple morbidity, stress management, treatment of phantom limb pain, smoking cessation, and personalized nutrition based on glycemic response. The average number of enrolled participants in the studies was 71 (range 8 - 214), and the average follow-up study duration was 69 days (range 3 - 180). 6 interventions (75%) showed statistical significance (at p=0.05 level) in health outcomes.
**Conclusions:** This review found that digital health interventions incorporating machine learning algorithms in real-life studies can be useful and effective. Given the low number of studies identified in this review and that they do not follow a rigorous machine learning evaluation methodology, we urge the research community to conduct further studies in intervention settings following evaluation principles and demonstrating the potential of machine learning in clinical practice.

**Keywords:** Machine learning, data mining, artificial intelligence, digital health, review

**Introduction**

Digital health interventions [1], including modalities such as telemedicine, web-based strategies, e-mail, mobile phones, mobile applications, text messaging, and monitoring sensors, have enormous potential to support independent living and self-management [2], reduce healthcare costs [3], and have shown great promise towards improving health [4]. With the advent of new tools and algorithms for machine learning, a new class of smart digital health interventions can be developed, which could revolutionize effective healthcare delivery [5].

The term “machine learning” is widely used across disciplines, but has no universally accepted definition [6]. This is in part explained because of the breadth of the areas it covers, and also because researchers from diverse disciplines have historically (and still do) contributed to its development. Broadly, it refers to an algorithmic framework which can provide insights into data, whilst facilitating inference and providing a tentative setting to determine functional relationships.

Machine learning has been applied in multiple healthcare domains, including diabetes [7], cancer [8], cardiology [9], and mental health [10]. Most of the developed machine learning models and tools in research settings investigate the potential of prognosis [11], diagnosis [12], or differentiation of clinical groups (e.g. a group with a pathology and a healthy control group or
groups with pathologies) [13], thus demonstrating promise towards the development of computerized decision support tools [14]. The key requirement for the development of these tools is both sufficiently large datasets (both in terms of number of participants and explanatory variables to explore), and also accurate labels, typically provided by expert clinicians. The premise is the identification of those data structures or variables (e.g., clinical, behavioral, or demographic variables) which are associated with the target outcome (e.g., whether a person has cancer). In this regard, useful knowledge can be derived from the available data, which can empower patients to monitor their health status longitudinally, and support health professionals in decision making with regards to management, treatment, and follow-up interventions where required.

Despite a considerably growing body of research literature in the use of machine learning in healthcare applications [15], it is astonishing how few of these suggestions are actually translated into clinical practice [16]. There is remarkably limited empirical evidence on the effectiveness of machine learning applications in digital health interventions. This is rather surprising, since any proposed healthcare solutions would reach their full potential only if they are embraced by the medical community, becoming integrated within properly designed digital health interventions and tested in real-life studies with patients and health professionals.

Considering that machine learning models and tools have not been widely and reliably used in clinical practice, whereas the peer-reviewed literature in the field grows exponentially, we wanted to assess the progress made in smart data-driven health interventions applied in real-life research settings, i.e., the real-world in which constraints in terms of available resources or opportunities to collect reliable data may exist as opposed to simulation or laboratory-based studies [17]. In this direction, we present a systematic literature review of digital health interventions incorporating machine learning algorithms, by identifying and mapping out their features and
outcomes, with the aim to improve our knowledge in the design and development of impactful intelligent interventions.

Methods

The bibliographic databases of PubMed and Scopus were searched to identify digital health non-pharmacological interventions incorporating machine learning, which were assessed in pragmatic studies published after 2008. In this context, the inclusion criteria for study selection were: a) the study should be conducted with patients, health professionals or both, in a real-life setting, b) machine learning algorithms/models used in the digital health intervention should have been used (rather than merely reporting statistical hypothesis testing results and/or statistical associations), c) quantitative outcomes of the study should be presented, and d) the paper describing the study should be written in English. Retrospective studies, case reports, ongoing studies, surveys or reviews, laboratory or simulation studies, studies describing protocols, qualitative studies, and all studies published before 2008 were excluded from the review because we wanted to get the pulse of recent research developments in the field which have been used in clinical interventional settings.

We used the string “(machine learning) OR (data mining) OR (artificial intelligence) AND health” for search within the title, abstract, and keywords of the manuscripts. “Species” in PubMed was limited to include only humans. Authors AKT and AT screened the identified papers following the literature search independently, in order to minimize bias in the selection process. Any disagreements were resolved by discussion between the authors and reaching consensus. We screened the abstracts of the candidate articles for inclusion, and subsequently read the full manuscripts of the ones found being eligible according to the inclusion criteria. Subsequently, we excluded the articles not providing sufficient information in regard with the application of machine
learning, or being ineligible. We used the Effective Public Health Practice Project (EPHPP) tool to assess the methodological quality of the included studies, which has been found to be reliable [18]. The studies that focused on interventions were synthesized (AKT) according to their target (i.e., target condition), the study design, the number of enrolled participants, the follow-up duration, the primary outcome and whether this has been significantly positive, the machine learning algorithms used in the intervention, as well as the outcome of the algorithms (e.g., prediction of a target outcome). The systematic review was conducted following the PRISMA guidelines [19]. We have included a completed PRISMA checklist as Supplementary Material.

Figure 1 Flow diagram for study inclusion following the PRISMA format.
Results

Literature Search Outcomes

Our last search in November, 2018 returned 1386 manuscripts from the PubMed database, and 7024 manuscripts from Scopus. All the retrieved records were imported in the Mendeley© bibliography management software [20], which identified 1093 duplicates. The abstracts of the remaining 7317 results were screened according to our inclusion and exclusion criteria, and 21 eligible articles were identified. The reviewers read the full-text of the 21 manuscripts, and agreed to include 8 in total eligible manuscripts. The reasoning for excluding research articles is summarized in the flow diagram for study inclusion following the PRISMA format (Fig. 1).

Quality Assessment

Based on the EPHPP criteria for selection bias, design, confounders, blinding, data collection, and drop-outs, the methodological quality was found to be moderate for 2 out of 8 studies [21,22] (25%), and weak for the remaining 6 studies [23–28] (75%) (Table 1). Most studies were poorly rated because of selection bias, insufficient care in controlling for confounders, and

<table>
<thead>
<tr>
<th>Study</th>
<th>EPHPP Criteria</th>
<th>Global Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>Burns et al. [23]</td>
<td>W</td>
<td>M</td>
</tr>
<tr>
<td>Hawley et al. [24]</td>
<td>W</td>
<td>W</td>
</tr>
<tr>
<td>Manuvinakurike et al. [22]</td>
<td>W</td>
<td>S</td>
</tr>
<tr>
<td>Morrison et al. [26]</td>
<td>M</td>
<td>M</td>
</tr>
<tr>
<td>Ortiz-Catalan et al. [27]</td>
<td>M</td>
<td>M</td>
</tr>
<tr>
<td>Sadasivam et al. [21]</td>
<td>W</td>
<td>S</td>
</tr>
<tr>
<td>Zeevi et al. [28]</td>
<td>W</td>
<td>S</td>
</tr>
</tbody>
</table>
the high percentage of withdrawals or drop-outs (or the absence of their description). The design of a randomized or controlled clinical trial was described in 3 studies [21,25,28] (37%), and 5 interventions (63%) were evaluated in a pilot or experimental single-group study.

**Type of Interventions and Target Population**

The interventions targeted depression prediction and management [23], speech recognition for people with speech disabilities [24], self-efficacy for weight loss [22], detection of

Table 2 Characteristics of included studies and implications for clinical practice (IT: Intervention target, SD: Study design, NP: Number of enrolled participants, FU: Follow-up duration, PO: Primary outcome, SO: Significantly positive outcome reported, MA: Machine learning algorithm(s) used, MO: Machine learning algorithm outcome, IC: Implications for clinical practice).

<table>
<thead>
<tr>
<th>IT</th>
<th>SD</th>
<th>NP</th>
<th>FU</th>
<th>PO</th>
<th>SO</th>
<th>MA</th>
<th>MO</th>
<th>IC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Burns et al. [23]</td>
<td>Depression prediction and management</td>
<td>Single group interventional study</td>
<td>8</td>
<td>8 weeks</td>
<td>Severity of depressive symptoms</td>
<td>Yes</td>
<td>Regression and decision trees</td>
<td>Prediction of sadness (7-point Likert scale) and location</td>
</tr>
<tr>
<td>Hawley et al. [24]</td>
<td>Speech recognition for people with speech disabilities</td>
<td>Observational study</td>
<td>9</td>
<td>2-4 weeks</td>
<td>Recognition accuracy in real settings</td>
<td>No</td>
<td>Hidden Markov Models</td>
<td>Closeness of fit of user attempt in speaking a word to own speech recognition model</td>
</tr>
<tr>
<td>Manuvinakurike et al. [22]</td>
<td>Self-efficacy for weight loss</td>
<td>2x2 between-subjects randomized pilot study</td>
<td>103</td>
<td>N/A</td>
<td>Self-efficacy for weight loss</td>
<td>Yes (No for decisional balance)</td>
<td>Adaptive boosting</td>
<td>Relevance of stories to the users</td>
</tr>
<tr>
<td>Martin et al. [25]</td>
<td>Detection of changes in biopsychosocial condition of patients with multiple morbidity (e.g., lung disease and cardiovascular disease)</td>
<td>Randomized Controlled Trial</td>
<td>214</td>
<td>6 months</td>
<td>Unplanned hospital visits</td>
<td>Yes</td>
<td>Decision trees</td>
<td>Prediction of unplanned events</td>
</tr>
<tr>
<td>Morrison et al. [26]</td>
<td>Stress management</td>
<td>Exploratory mixed-methods study</td>
<td>77</td>
<td>3 days on average</td>
<td>Patterns of use and notification response</td>
<td>No</td>
<td>Naive Bayesian classifier</td>
<td>Notification response (yes/no)</td>
</tr>
</tbody>
</table>
| Ortiz-Catalan et al. [27] | Treatment of phantom limb pain | Single group interventional study | 14 | 6 months | Phantom limb pain | Yes | Linear Discriminant Analysis, Multi-Layer Perceptron (supervised) | Prediction of individual and simultaneous | To be utilized in novel treatment methods which could be used in case evidence-
changes in biopsychosocial condition of patients with multiple morbidity [25], stress management [26], treatment of phantom limb pain [27], smoking cessation [21], and personalized nutrition based on glycemic response [28] (Table 2).

3 out of 8 interventions (38%) targeted patients; individuals diagnosed with depression [23], multiple morbidities such as lung disease and cardiovascular disease [25], and phantom limb pain [27]. One intervention (13%) targeted people with speech disabilities [24]. 4 interventions [21,22,26,28] (50%), targeted individuals who have not explicitly been diagnosed with a disease or impairment. All target groups involved adults. The average number of enrolled participants in the studies was 71 (range 8 - 214), and the average follow-up study duration was 69 days (range 3 - 180).

Applications of Machine Learning and Outcomes

Overall, 6 out of 8 real-life studies of digital health interventions aided with machine learning algorithms (75%), showed statistical significance (at p=0.05 level) in health outcomes. Different summary measures were used in the identified studies to assess primary outcomes, which reflects both the lack of standardization in methodology, and also in the metrics used in research.
fields. Where possible, we aimed to use accuracy of used algorithms and p value (e.g., for showing statistical significance of outcomes in an intervention group compared to a control group) as the principal summary measures. All included studies, in terms of intervention purpose and content, evaluation outcomes, and implications for clinical practice, are briefly described below.

Burns et al. [23] described a multi-component mobile-based intervention in which machine learning models to predict the mood, emotions, cognitive/motivational states, activities, environmental and social context of patients with depression were used, along with feedback graphs for self-reflection on behaviour and coaching provided by caregivers. The predictive models were based on phone sensor derived variables (e.g., global positioning system, ambient light, phone calls), and regression along with decision trees were used. The accuracy of the models was promising for location prediction (60% to 91%), but prediction was very poor for emotions such as sadness. Overall, the 8 participants of the study, became less likely to meet the criteria for diagnosis of major depressive disorder (p=0.03), and their symptoms of depression and anxiety were decreased by the end of the study (p<0.001). Patients were also satisfied with the intervention (5.71 average rating on a scale 1 to 7), and 6 out of 7 treatment completers (86%) indicated the intervention was helpful in understanding triggers for negative moods. Despite the benefits of self-reflection on behavior through the use of a multi-component mobile health monitoring system, and the clinical improvements shown in the study, the authors reported that the clinical utility of the used prediction models should be improved, since the prediction outcomes (e.g., location and mood) were merely displayed to the users, and there were no direct interventions based on them.

Hawley et al. [24] described the use of a device capable of recognizing the speech of people with dysarthria and generating voice messages. The authors used Hidden Markov Models (HMM), to determine the proximity of a spoken word to a personalized speech model for that individual. However, only 67% recognition accuracy was achieved in a real-life observational study with 9
participants. Participants noticed that there was a reduced ease of communication through the device compared to their usual communication method of either speaking or speaking supported by a conventional voice-output communication aid, mainly due to the low accuracy of speech recognition. Nevertheless, feedback from participants was positive about the device concept, given that speech recognition is improved.

Manuvinakurike et al. [22] focused on changes in self-efficacy for weight loss, through the provision of personal health behavior change stories found in Internet. An algorithm based on adaptive boosting was developed to find the most relevant story based on the stage of change and the demographic characteristics of a user, along with the emotional tone and overall quality of the story (accuracy between 84% and 98% for the classification of 5 stages of change). Testing of the algorithm with 103 users revealed significantly greater increases in self-efficacy for weight loss (p=0.02), and an insignificant effect on change in decisional balance (p=0.83). In addition, the medium used to tell the stories, being either text or an animated conversational agent, had no effect in health behavior change. The authors concluded that their approach could maximize participants’ engagement in longitudinal health behavior change interventions.

Martin et al. [25] utilized a system in which decision trees could predict unplanned hospital visits of patients with multiple morbidities such as lung disease or cardiovascular disease, and alerts were sent to health professionals which were acted upon according to agreed guidelines. The system was based on information received via patient phone calls with lay care guides. Linguistic and meta-linguistic features were extracted together with current patient status, in order to train the prediction models (positive predictive value of 70% for predicting unplanned events). A randomized controlled trial with 214 patients for 6 months (the largest trial found in the review in terms of number of enrolled participants and duration) showed a reduction of 50% in the number of unplanned hospital events of participants in the intervention group compared to control. The
The most common response to an alert indicating that a patient needs attention (red alert) was to phone the patient the next day to reassess the situation and contact the GP (3% of calls), suggest/plan a GP visit (11% of calls), or call an ambulance (< 0.01% of calls). In summary, the authors reported that predictive analytics on an ongoing basis could be used to signify risk of hospitalization and guide the healthcare system to take appropriate actions.

Morrison et al. [26] used push notifications in order to enhance engagement of smartphone users for stress management. A Naïve Bayes classifier was used to predict whether a user will respond to a notification, thereby building a personalized intelligent mechanism for notification delivery, based on the times within a day a user is more likely to view and react on the received messages. However, the exploratory study with 77 participants, showed no statistically significant difference between participants receiving the messages sent “intelligently”, compared to participants receiving a message daily or occasionally within 72 hours (Cohen’s d value 0.14 for intelligent vs daily group and 0.5 for intelligent vs occasional group, for action over messages). Although notification delivery based on time had no effect in the study groups (i.e., response on notifications was no different), the authors concluded that frequent daily messages may not deter users from engaging with digital health interventions.

Ortiz-Catalan [27] applied myoelectric pattern recognition algorithms for the control of a virtual limb in patients with phantom limb pain, and used gaming along with augmented/virtual reality for treatment. A single-group study with 14 participants revealed that patients’ symptoms of phantom limb pain were significantly decreased (by about 50%) at the end of the provided treatment for 6 months (p=0.0001 for reduction in intensity and quality of pain). The authors suggested that their novel treatment could be used after failure of evidence-based treatments such as the mirror therapy, and before proceeding with invasive or pharmacological approaches.
Sadasivam et al. [21] used a recommender system to send motivational messages to individuals, targeting at smoking cessation. The system was based on the Bayesian probabilistic matrix factorization to predict message rating, through the processing of data from user’s previous ratings of messages, along with other users’ ratings. A randomized controlled trial with 120 users showed that the system was more effective towards influencing people to quit smoking than standard tailored messages (rule-based system) with proven effectiveness (p=0.02), and resulted in similar cessation rate. The authors concluded that their recommender system could be used instead of standard systems for influencing smoking cessation, because it is more personalized (it learns and adapts to person’s behavior) and can incorporate a significantly greater number of variables, but larger trials would be needed to demonstrate effectiveness.

Zeevi et al. [28] used gradient boosting regression to predict the post-meal glycemic response of individuals in real-life, according to blood parameters, dietary habits, anthropometrics, physical activity, and gut microbiota. The results from a randomized controlled study with 24 participants, showed that a personalized diet based on the post-meal glycemic predictions, can significantly modify elevated postprandial blood glucose (p<0.05 for predicting low levels of blood glucose (“good diet”) vs high levels of blood glucose (“bad diet”) which was comparable to diets selected by experts). The authors reported that their approach could be utilized in nutritional interventions for controlling or preventing disorders associated with poor glycemic control, such as obesity, diabetes, and non-alcoholic fatty liver disease. However, evaluation periods of months or even years would be first needed to clearly indicate the effectiveness of the proposed algorithm.
Discussion

Main Findings

This review is, to our knowledge, the first to systematically examine the features and outcomes of digital health interventions incorporating machine learning, which have been implemented and assessed in real-life studies [17]. To this aim, our review differentiates from previous investigations which have merely focused on the broader use of artificial intelligence in medicine, in the context of specific diseases [29,30], machine learning techniques [31,32], or risk prediction models e.g., through mining of electronic health records [33,34], and not considered the real-life evaluation of the relevant interventions. The need to demonstrate evidence of effectiveness of interventions in the real-world has been highlighted in several other studies [35–37]. Our main finding is that the majority of the digital health interventions showed significantly positive health outcomes for patients or healthy individuals, which demonstrates the virtue of machine learning applications in actual clinical practice. However, given the small number of studies identified in this review and their considerable limitations highlighted above, further work is warranted to demonstrate the effectiveness of digital interventions relying on machine learning applications in real-life medical care.

Our review revealed 8 different cases of machine learning applications in a real-life setting; depression prediction and management, speech recognition for people with speech disabilities, self-efficacy for weight loss, detection of changes in biopsychosocial condition of patients with multiple morbidity, stress management, treatment of phantom limb pain, smoking cessation, and personalized nutrition based on glycemic response. The reviewed studies had several implications for the clinical practice including for example better engagement of patients with interventions [22], identification of risk for hospitalization [25], or the introduction of novel treatment methods
From those studies, the studies for speech recognition of people with speech disabilities [24], and notification delivery for stress management [26], clearly reported insignificant outcomes, whereas 6 studies showed significant outcomes, but they were of low-to-moderate methodological quality. Only 3 studies were in the form of a randomized controlled trial, which limited the ability to fully identify the added value of machine learning-enabled interventions compared to standard care. To this end, further rigorous studies with adequately powered samples (recruiting considerably more participants than the average number of 71 participants found in this review) are needed, which will enable to generate the evidence base for the effectiveness of machine learning in clinical practice. To that effect, large trials and publicly accessible databases which have become available over the last few years such as the UK BioBank and the Physionet database provide rich resources which could facilitate insights.

The delivery of motivational messages [21,26] or stories [22] for health behavior change and engagement, seems to be an emerging area of digital health interventions incorporating machine learning. These studies also demonstrate the latest efforts to promote personalized self-management of citizens and put them at the center of healthcare [38]. Considering the effectiveness of tailored messaging in influencing health behavior change [39], further research in this area is warranted.

The surprisingly small number of identified pragmatic studies in our review might raise some concerns, and indicates the substantial challenge of systematically evaluating digital health interventions which incorporate machine learning [40]. In this context, the retrospective validation of algorithms and models, given the availability of one or more data-sets, constitutes only the first step in the evaluation process [28]. The second step involves the integration of the algorithms and models within a digital health tool, e.g., smartphone-based [23], Internet-based [14], or an aid device [24]. The third step requires the assessment of the developed tool as a digital health
intervention in a real-life research setting (e.g., through a randomized controlled trial), together with patients, health professionals or both [28,41]. The final step would be the monitoring of actual uptake and use of the intervention in real-word and outside of a research setting [42], which is however rarely reported [43]. Admittedly, this process is challenging and anything but trivial. It requires a significant amount of time and resources which might not always be available, and the multidisciplinary collaboration among experts in different fields, e.g., engineering, computer science, behavioral science, and medicine, which might not be straightforward. However, those synergistic collaborative approaches are likely necessary toward the development of evidence-based, sustainable, and impactful digital health interventions [44,45].

Limitations

We used the term “machine learning” along with broader terms such as “data mining” and “artificial intelligence” for our literature search, and not keywords for specific machine learning algorithms or domains relevant to digital health such as telemedicine. This might have resulted in an inadvertent omission of studies which could have contributed to the progress made in machine learning applications for digital health. The afore-mentioned terms were combined with the generic term “health”, aiming to conduct a broad search with provided boundaries, and include the most pertinent articles relevant to digital health. We searched for articles in a limited number of databases, i.e., Pubmed and Scopus, which nevertheless represent two of the most widely-used databases internationally [46]. Hand-search of studies reported in other reviews or the included studies was not conducted, and the inter-rater reliability between the reviewers was not assessed. A meta-analysis was not possible due to the heterogeneity of the included studies.
Conclusion

Our review showed that real-life digital health interventions incorporating machine learning, can be useful and effective. Considering the small number of studies examined in this review and their limitations, further evidence of the clinical usefulness of machine learning in health service delivery is needed. In this direction, researchers are encouraged to move beyond the retrospective validation of machine learning models, by integrating their models within appropriately designed digital health tools, and evaluating their tools in rigorous studies conducted in real-life settings.

Authors’ Contributions

AKT was responsible for the study conduction; Authors AKT and AT reviewed the literature and assessed the quality of the included studies; AKT synthesized the literature according to the described methodology; AKT wrote a first draft of the manuscript and AT contributed to the final version. Both authors have read and agreed to submit the manuscript to be considered for publication.

Acknowledgements

The study was supported by the Wellcome Trust through a Centre Grant No. 098461/Z/12/Z, “The University of Oxford Sleep and Circadian Neuroscience Institute (SCNi). It was also supported by Health Data Research UK which receives its funding from HDR UK Ltd (HDR-5012) funded by the UK Medical Research Council, Engineering and Physical Sciences Research Council, Economic and Social Research Council, Department of Health and Social Care (England), Chief Scientist Office of the Scottish Government Health and Social Care Directorates, Health and Social Care Research and Development Division (Welsh Government), Public Health Agency (Northern
Ireland), British Heart Foundation (BHF) and the Wellcome Trust. The funders had no role in the study and the decision to submit this work to be considered for publication.

Conflicts of Interest

The authors of this manuscript declare no conflicts of interest.

References


40. Monitoring and evaluating digital health interventions. World Health Organization. 2019-


46. Falagas ME, Pitsouni EI, Malietzis GA, Pappas G. Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses. FASEB J [Internet] *Alfa Institute of Biomedical Sciences, Athens, Greece;Department of Medicine, Tufts University School of Medicine, Boston, Massachusetts, USA;andInstitute of Continuing Medical Education of Ioanna, Ioannina, Greece.: Federation of American Societies for Experimental Biology; 2008 Feb;22(2):338–342. PMID:17884971

Abbreviations

**EPHPP**: Effective Public Health Practice Project

**HMM**: Hidden Markov Models