Automated-Entry Patient-Generated Health Data for Chronic Conditions: The Evidence on Health Outcomes

Prepared for:
Agency for Healthcare Research and Quality
US Department of Health and Human Services
5600 Fishers Lane
Rockville, MD 20857
www.ahrq.gov

Contract No.
To be included in the final report

Prepared by:
To be included in the final report

Investigators:
To be included in the final report

AHRQ Publication No.
<Month Year>
Preface

The Agency for Healthcare Research and Quality (AHRQ), through its Evidence-Based Practice Centers (EPCs), sponsors the development of evidence reports and technology assessments to assist public- and private-sector organizations in their efforts to improve the quality of healthcare in the United States. The reports and assessments provide organizations with comprehensive, science-based information on common, costly medical conditions and new healthcare technologies and strategies. The EPCs systematically review the relevant scientific literature on topics assigned to them by AHRQ and conduct additional analyses when appropriate before developing their reports and assessments.

This EPC evidence report is a Technical Brief. A Technical Brief is a rapid report, typically on an emerging medical technology, strategy, or intervention. It provides an overview of key issues related to the intervention—for example, current indications, relevant patient populations and subgroups of interest, outcomes measured, and contextual factors that may affect decisions regarding the intervention. Although Technical Briefs generally focus on interventions for which limited published data exist and too few completed protocol-driven studies to support definitive conclusions, the decision to request a Technical Brief is not solely based on the availability of clinical studies. The Technical Brief’s goals are to provide an early objective description of the state of the science, a potential framework for assessing the applications and implications of the intervention, a summary of ongoing research, and information on future research needs. In particular, through the Technical Brief, AHRQ hopes to gain insight on the appropriate conceptual framework and critical issues that will inform future research.

AHRQ expects that the EPC evidence reports and technology assessments will inform individual health plans, providers, and purchasers as well as the health care system as a whole by providing important information to help improve health care quality.

If you have comments on this Technical Brief, they may be sent by mail to the Task Order Officer named below at: Agency for Healthcare Research and Quality, 5600 Fishers Lane, Rockville, MD 20857, or by email to epc@ahrq.hhs.gov.

Gopal Khanna, M.B.A.  
Director  
Agency for Healthcare Research and Quality

Arlene S. Bierman M.D., M.S.  
Director  
Center for Evidence and Practice Improvement  
Agency for Healthcare Research and Quality

Stephanie Chang M.D., M.P.H.  
Director  
Evidence-based Practice Center Program  
Center for Evidence and Practice Improvement  
Agency for Healthcare Research and Quality

Elise Berliner, Ph.D.  
Task Order Officer  
Center for Evidence and Practice Improvement  
Agency for Healthcare Research and Quality
Acknowledgments
To be included in the final report

Key Informants
In designing the study questions, the EPC consulted a panel of Key Informants who represent subject experts and end-users of research. Key Informant input can inform key issues related to the technical brief’s topic. Key Informants are not involved in the analysis of the evidence or the writing of the report. Therefore, in the end, study questions, design, methodologic approaches, and/or conclusions do not necessarily represent the views of individual Key Informants.

Key Informants must disclose any financial conflicts of interest greater than $5,000 and any other relevant business or professional conflicts of interest. Because of their role as end-users, individuals with potential conflicts may be retained. The TOO and the EPC work to balance, manage, or mitigate any conflicts of interest.

The list of Key Informants who provided input to this report follows:
To be included in the final report
**Peer Reviewers**

Before publication of the final evidence report, EPCs sought input from independent Peer Reviewers without financial conflicts of interest. However, the conclusions and synthesis of the scientific literature presented in this report do not necessarily represent the views of individual reviewers.

Peer Reviewers must disclose any financial conflicts of interest greater than $5,000 and any other relevant business or professional conflicts of interest. Because of their unique clinical or content expertise, individuals with potential nonfinancial conflicts may be retained. The TOO and the EPC work to balance, manage, or mitigate any potential nonfinancial conflicts of interest identified.

The list of Peer Reviewers follows:
To be included in the final report
Automated-Entry Patient-Generated Health Data for Chronic Conditions: The Evidence on Health Outcomes

Structured Abstract

**Background.** Automated-entry consumer devices that collect and transmit patient-generated health data (PGHD) are currently being evaluated as potential tools to aid in the management of chronic diseases. The majority of consumer health technologies entering the U.S. market have not undergone FDA scrutiny. Manufacturers of non-FDA-regulated devices cannot legally make claims that their devices meet any stated performance/clinical standards, although they will often allude to the performance of their devices through carefully worded marketing. The need exists to evaluate the evidence regarding consumer PGHD technologies, particularly for devices that have not gone through FDA evaluation.

**Purpose.** To summarize the research related to automated-entry consumer health technologies that provide PGHD for the prevention or management of 11 chronic diseases.

**Methods.** The project scope was determined through discussions with Key Informants. KIs agreed with our proposal to focus on technologies available to consumers without a prescription that provide automated data capture, as well as our proposal to not require that data be automatically transmitted to health care staff. We searched MEDLINE and EMBASE (via EMBASE.com), In-Process MEDLINE and PubMed unique content (via PubMed.gov), and the Cochrane Database of Systematic Reviews for systematic reviews or controlled trials. We also searched ClinicalTrials.gov for ongoing studies. We used eligibility criteria to screen abstracts and full articles before study-level data abstraction. We assessed risk of bias and extracted data on health outcomes, surrogate outcomes, usability, sustainability, cost-effectiveness outcomes (quantifying the trade-offs between health effects and cost), process outcomes, and other characteristics related to PGHD technologies. To assess device similarity with devices currently marketed by the same manufacturer, device engineers used a 1-4 scale (1 being similar to current models, 2 being somewhat different to current models, 3 being very different to current models, and 4 being unable to reliably assess similarity to current models). For isolated effects on health outcomes, we classified the results in one of four categories: 1) likely no effect, 2) unclear, 3) possible positive effect, or 4) likely positive effect. When we categorized the data as “unclear” based solely on health outcomes, we then examined surrogate outcomes for that particular clinical condition. We then used these surrogate data, along with the direct effects on health outcomes, to determine which of the four categories was most appropriate.

**Findings.** We identified 111 unique studies that met inclusion criteria. The largest number of studies addressed patients with hypertension (50 studies) and obesity (43 studies). Eighty-three trials used a single PGHD device, 22 used 2 PGHD devices, and the other 6 used 3 or more PGHD devices. Pedometers, blood pressure (BP) monitors, and scales were commonly used in the same studies. Overall, we found a “possible positive effect” of PGHD interventions on health outcomes for hypertension, coronary artery disease, heart failure, and asthma. We rated PGHD interventions for obesity as having “likely no effect” on health outcomes. We considered the findings “unclear” regarding PGHD interventions for diabetes prevention, sleep apnea, cardiac
arrhythmias or conduction abnormalities, stroke, Parkinson’s disease, and chronic obstructive pulmonary disease. For these conditions, most studies lacked adequate statistical power to detect a clinically important between-group difference on health outcomes. Most studies did not report harms related to PGHD interventions; the relatively few harms reported were minor and transient, with event rates usually comparable to harms in the control groups. Few studies reported cost-effectiveness analyses, and only for PGHD interventions for hypertension, coronary artery disease and chronic obstructive pulmonary disease; the findings were variable across different chronic conditions and devices. Patient adherence to PGHD interventions was highly variable across studies, but patient acceptance/satisfaction and usability was generally fair to good. However, device engineers independently evaluated consumer wearable and handheld BP monitors and considered the user experience to be poor, while their assessment of smartphone-based ECG monitors found the user experience to be good. Student volunteers involved in device usability testing of the Weight Watchers Online app found it well-designed and relatively easy to use.

**Implications.** Multiple randomized controlled trials (RCTs) have evaluated some PGHD technologies (e.g., pedometers, scales, BP monitors), particularly for obesity and hypertension, but health outcomes were generally underreported. Despite this limitation, we found evidence suggesting a possible positive effect of PGHD interventions on health outcomes for 4 chronic conditions. Lack of reporting of health outcomes and insufficient statistical power to assess these outcomes comprise the main reasons we considered the effect of PGHD on health outcomes as “unclear” for some chronic conditions. In general, the majority of studies on PGHD technologies still focus on non-health-related outcomes, even among chronic conditions with the largest evidence bases. Future RCTs should focus on measuring the effects of these technologies on health outcomes in patients with chronic disease. Furthermore, many current RCTs did not isolate the effect of the PGHD intervention from other components in a multicomponent intervention. Future studies should be designed to isolate the effect of the PGHD intervention (e.g. PGHD + X versus X alone).
Evidence Summary

Main Points

- 111 controlled trials on 11 chronic conditions studied the use of automated-entry consumer devices that capture patient generated health data (PGHD)
- Possible positive effect of PGHD technologies for three chronic conditions (coronary artery disease, heart failure, and asthma), based on direct health outcomes data
- Possible positive effect on hypertension health outcomes, based on consistent surrogate data (systolic blood pressure, diastolic blood pressure)
- Likely no effect on obesity health outcomes, based on consistent surrogate data (BMI, weight)
- Potential benefit was unclear for other conditions.
- Patient adherence to PGHD interventions was highly variable and patient acceptance/satisfaction and usability was generally fair to good.
- Future studies should measure long-term health outcomes and attempt to isolate the effect of PGHD technologies.

Background and Purpose
Automated-entry consumer devices that collect and transmit patient-generated health data (PGHD) are currently being evaluated as potential tools to aid in the management of chronic diseases. The majority of consumer health technologies entering the U.S. market have not undergone FDA scrutiny. While these technologies provide much information to patients and providers, we focused on whether they improve health outcomes for patients with chronic conditions (e.g., hypertension, obesity, coronary artery disease). In summary, we examined the evidence on health outcomes related to automated-entry consumer technologies that provide PGHD for the prevention or management of 11 chronic diseases.

Methods
We employed methods consistent with those outlined in the AHRQ EPC Program Methods Guidance (https://effectivehealthcare.ahrq.gov/products/cer-methods-guide/overview), and we describe these in the full report. Our searches covered publication dates up to December 23, 2019. We determined whether each study had constructed comparisons group(s) that isolated the effect of the PGHD technology. For isolated effects on health outcomes, we classified the results in one of four categories: 1) likely no effect, 2) unclear, 3) possible positive effect, or 4) likely positive effect. If the results consistently demonstrate the lack of an effect (via narrow confidence intervals around a null effect), we coded it as “likely no effect.” We examined key surrogate outcomes (e.g., blood pressure for hypertension, BMI for obesity) whenever the direct health outcome data were unclear. We also tabulated data on the frequency of device usage, ease of use, technical problems, and cost-effectiveness.
## Results

<table>
<thead>
<tr>
<th>Clinical Condition</th>
<th>Results Categorization for Isolated Health Outcomes</th>
<th>Comments</th>
</tr>
</thead>
</table>
| Obesity                                  | Likely no effect                                    | 3 of 43 included trials reported whether there were isolated effects on health outcomes (specifically, quality of life), and overall results were unclear.  
14 trials reported whether there were isolated effects of device presence on surrogate outcomes (BMI or weight), and all 25 point estimates were less than the minimal important difference (5% body weight). |
| Diabetes prevention                      | Unclear                                             | None of the three trials reported whether there were isolated effects on health outcomes.  
One trial reported a surrogate outcome (metabolic syndrome risk) and it found an advantage of PGHD, however it was at high risk of bias. |
| Sleep apnea                              | Unclear                                             | None of the three trials reported whether there were isolated effects on health outcomes.  
One trial reported a surrogate outcome (number of days on which apnea events were witnessed) and it found no statistically significant difference and was at high risk of bias. |
| Hypertension                             | Possible positive effect                            | Six of the 50 included studies reported whether there were isolated effects on health outcomes (including quality of life, mortality, and hospitalizations), and overall results were unclear.  
Seventeen studies reported whether there were isolated effects of device presence on surrogate outcomes (SBP, DBP, and BP control), and results generally favored PGHD arms. |
| Coronary artery disease                   | Possible positive effect                            | Mortality was significantly lower in the PGHD arm in the only study that reported it. Re-hospitalization was also lower but did not reach statistical significance. |
| Heart failure                            | Possible positive effect                            | Different quality of life measures favored the PGHD intervention group in two studies that isolated the effect of PGHD.                                                                                   |
| Cardiac arrhythmias or conduction         | Unclear                                             | There were no statistically significant between-group differences in health outcomes that favored the PGHD intervention. One low risk study found a higher rate of emergency room visits in the AliveCor group, but a positive effect of PGHD for time to arrhythmia detection. Overall, the effect of PGHD on health outcomes is unclear. |
| Parkinson's Disease                      | Unclear                                             | No studies met inclusion criteria                                                                                                                                                                     |
| COPD                                     | Unclear                                             | 3 of 9 RCTs reported isolated effects on health outcomes, but data were unclear. Data on surrogate outcomes (e.g., lung function) were also unclear.                                                           |
| Asthma                                   | Possible positive effect                            | 1 study met inclusion criteria (moderate risk of bias), and it found better symptom control in the PGHD group overall and in the pediatric population alone.                                             |

Also, student volunteers involved in usability testing of the Weight Watchers Online app found it well-designed and relatively easy to use. Device engineers independently evaluated consumer wearable and handheld BP monitors and considered the user experience to be poor, while their assessment of smartphone-based ECG monitors found the user experience to be good.

## Strengths and Limitations

The above findings summarize studies that used an isolated-effect design (e.g., PGHD alone vs no intervention, or PGHD + X vs. X alone). Many other studies used multicomponent
interventions (of which PGHD was one component), making it impossible to determine the impact of PGHD. Furthermore, many included studies reported only surrogate outcomes, or only followed patients for a short amount of time (e.g., 3 months).

**Implications and Conclusions**

Automated-entry consumer PGHD technologies provide a wealth of information for patients and providers, and for four chronic conditions, evidence suggests a possible positive effect on health outcomes. Clinicians might consider recommending that patients with these conditions use a consumer technology for self-monitoring, such as recommending a blood pressure monitor for patients with hypertension.

For obesity, we found consistent evidence of no important effect on either BMI or weight, suggesting that there is likely no effect on health outcomes. The most common devices in the obesity studies were accelerometers and pedometers, therefore measuring one’s steps per day does not appear to reduce weight or improve health.

For many other conditions, however, the available evidence is unclear on the PGHD effect on health outcomes. More studies with an isolated-effect design, measuring health outcomes in the long term, may help reduce the uncertainty of the impact of PGHD technologies.
Contents

Introduction .................................................................................................................................. 1
Background .................................................................................................................................. 1
Guiding Questions .................................................................................................................... 2
Methods ....................................................................................................................................... 3
  Discussion with Key Informants ............................................................................................. 3
  PICOTS and Inclusion Criteria ............................................................................................... 4
  Search Strategies ................................................................................................................... 5
  Study Selection ...................................................................................................................... 5
  Study-Level Data Extraction ............................................................................................... 5
  Risk of Bias ........................................................................................................................... 6
  Results Classification .......................................................................................................... 6
  Evaluation of Economic Evaluations ................................................................................ 7
  Device Similarity and Usability Testing ............................................................................. 7
  Data Presentation ................................................................................................................ 8
  Peer Review and Public Commentary ............................................................................... 8
Findings ..................................................................................................................................... 8
  Results of Literature Searches ............................................................................................ 8
  Overview of Evidence .......................................................................................................... 8
  Obesity .................................................................................................................................. 11
  Diabetes Prevention ............................................................................................................ 22
  Sleep Apnea ......................................................................................................................... 27
  Hypertension ....................................................................................................................... 32
  Coronary Artery Disease ...................................................................................................... 45
  Heart Failure ....................................................................................................................... 50
  Cardiac Arrhythmias .......................................................................................................... 55
  Stroke ................................................................................................................................... 61
  Parkinson’s Disease ........................................................................................................... 66
  Chronic Obstructive Pulmonary Disease ......................................................................... 66
  Asthma .................................................................................................................................. 76
Summary and Implications ....................................................................................................... 80
Next Steps ................................................................................................................................. 85
References ................................................................................................................................. 87
Appendix A. Search Strategies ................................................................................................. A-1
Appendix B. Excluded Publications ....................................................................................... B-1
Appendix C. Evidence Tables ................................................................................................. C-1

Tables
Table 1. Inclusion Criteria ......................................................................................................... 4
Table 2. Overview of Evidence ................................................................................................. 8
Table 3. Obesity: Categorization of Isolated Effects Due to Device Presence/Absence .......... 19
Table 4. Diabetes Prevention: Categorization of Isolated Effects Due to Device Presence/Absence .............................................................. 26
Table 5. Sleep Apnea: Categorization of Isolated Effects Due to Device Presence/Absence .. 31
Table 6. Hypertension: Categorization of Isolated Effects Due to Device Presence/Absence . 41
Table 7. Coronary Artery Disease: Categorization of Isolated Effects Due to Device Presence/Absence ....................................................................................................... 49
Table 8. Heart Failure: Categorization of Isolated Effects Due to Device Presence/Absence ....................................................................................................... 54
Table 9. Cardiac Arrhythmias: Categorization of Isolated Effects Due to Device Presence/Absence ....................................................................................................... 60
Table 10. Stroke: Categorization of Isolated Effects Due to Device Presence/Absence ....................................................................................................... 64
Table 11. COPD: Categorization of Isolated Effects Due to Device Presence/Absence ....................................................................................................... 72
Table 12. Asthma: Categorization of Isolated Effects Due to Device Presence/Absence ....................................................................................................... 79
Table 13. Primary Findings......................................................................................................................................................... 80

Figures
Figure 1. Counts of Unique Devices in Categories ................................................................................................................ 10
Figure 2. BMI differences in studies of isolated effects of device presence/absence ................................................................. 17
Figure 3. Weight differences in studies of isolated effects of device presence/absence ................................................................. 17
Figure 4. SBP differences in studies of isolated effects of device presence/absence ................................................................. 39
Figure 5. DBP differences in studies of isolated effects of device presence/absence ................................................................. 40
Figure 6. Graphical Summary of Findings .................................................................................................................................. 82
Introduction

Background

Chronic diseases (including coronary artery disease [CAD], hypertension, diabetes mellitus, obesity, and stroke) increase morbidity, mortality, and cost to the health care system and society, thereby creating a substantial public health burden. According to the Centers for Disease Control and Prevention (CDC), 6 in 10 adults have 1 or more chronic conditions and 4 in 10 adults have 2 or more conditions (https://www.cdc.gov/chronicdisease/about/index.htm). Self-management strategies that educate and assist patients to manage and monitor their chronic disease have the potential to improve health outcomes, maintain or slow the worsening of progressive conditions, and reduce costs.

Consumer devices that collect and transmit patient-generated health data (PGHD) are currently being evaluated as potential tools to aid self-management of chronic diseases. These technologies, which include mobile apps and wearable monitors, have the potential to reduce overutilization of the healthcare system by patients with chronic diseases. Monitoring enabled by PGHD devices could enhance self-management and thereby lower the number of ambulatory clinic visits, emergency room visits, and hospitalizations associated with several chronic diseases. This would save time and resources for both patients and providers as well as lower overall costs.

PGHD is a rapidly growing field in which the availability and development of the technologies have, in many instances, outpaced the publication of trials designed to evaluate health outcomes, usability, interoperability, and benefits and harms of these technologies. It is important to determine which of the available technologies have been assessed to determine efficacy related to health outcomes for consumers with (or at risk for) chronic diseases.

Over the past few years, consumer health technologies entering the U.S. market has surged, with the majority being products not undergoing FDA scrutiny. Consumer health technologies marketed in the United States can be broadly divided into those FDA considers medical devices that are regulated (i.e., devices that claim to diagnose, prevent, or treat medical conditions) and those that FDA does not consider to be medical devices. Manufacturers of FDA-regulated devices typically are required to demonstrate, through testing, that their devices meet any stated performance/clinical claims. In contrast, manufacturers of non-FDA-regulated devices cannot legally make these claims, although manufacturers will often allude to the performance of their devices through carefully worded marketing. Consumers and providers may find it challenging to distinguish between devices that are FDA regulated and those that are not.

Many studies evaluate multicomponent interventions, with PGHD technologies representing only one component, and they do not separately evaluate the PGHD component’s effect. While many apps or other PGHD technologies are meant to be used in combination with other interventions for chronic disease, the studies still need to be designed to permit isolation of PGHD technology’s effect. Mobile apps that have similar functions have rarely been directly compared in clinical studies to help clinicians identify the most useful apps to recommend for their patients.1

Several issues may influence PGHD’s effectiveness for improving patient outcomes. Concerns surround the accuracy of some of the new devices when measuring patient health data (e.g., Apple Watch). From the patient perspective, some patients may lack access to PGHD technologies or have a low comfort level with these new technologies and, as such, may use them in a suboptimal manner, limiting their effectiveness; they may even abandon use of the
device. Alternatively, even if patients can use the technologies effectively, some may be unwilling to share the data with clinicians due to privacy and security concerns. On the other hand, clinicians may not be equipped to process the information from PGHD technologies.

**Guiding Questions**

1. Which specific consumer automated-entry PGHD applications/technologies/devices have been studied for measurement of health outcomes? What studies are in progress of consumer PGHD devices?
2. What are the characteristics (e.g., interoperability, functions, acceptability/usability, sustainability, feasibility, fidelity, or integration into electronic health records) of these specific consumer automated-entry PGHD technologies?
3. What is the influence of specific consumer automated-entry PGHD technologies on health outcomes? Does this vary across different patient populations, different settings, or other modifiers of effectiveness?
4. What are the harms or adverse events associated with these specific consumer automated-entry PGHD technologies? Which patients in which settings are most at risk of harms?
5. For the technologies demonstrating associations with outcomes of interest, what full economic evaluations provide information on the relative value for consumers?
Methods

We followed established AHRQ processes for Technical Briefs, including interviewing Key Informants (KIs), soliciting additional unpublished materials to inform our review through a Federal Register notice, and using peer and public review. Because AHRQ Technical Briefs focus on emerging and rapidly changing technologies, strength-of-evidence assessments are not typically conducted, and we did not evaluate strength of evidence in this review. This draft report was sent to all KIs and selected peer reviewers who did not serve as KIs; it was also posted to the Effective Healthcare website for public comment.

Discussion with Key Informants

Seven KIs representing diverse perspectives, including FDA, clinical effectiveness, the challenges of telemedicine, policy analysis, and patient advocacy, provided input on this review (listed on p. vi). The intent of KI interviews was to provide context and guidance on areas most important to consider. KIs provided input on the review’s scope, including the proposed populations, interventions, comparators, outcomes, timing/setting (PICOTS). We asked specific questions about use of consumer products, automated data entry, data transmission to health care staff, and confounding factors in interpreting the literature.

We presented to the KIs our list of eight chronic conditions of interest (obesity, hypertension, diabetes prevention, coronary artery disease, heart failure, stroke, chronic obstructive pulmonary disease, asthma), and their input led to the addition of chronic conditions: cardiac arrhythmias or conduction abnormalities, Parkinson’s disease, and sleep apnea. These additions were due to KIs believing that PGHD technologies could potentially provide important benefits for patients with these conditions.

KIs agreed with our proposal to focus on technologies available to consumers without a prescription that provide automated data capture, as well as our proposal to not require that data be automatically transmitted to health care staff. While data transmission is possible with many PGHD technologies (e.g., via a downloadable application), KIs agreed that patients may gain health benefits by being aware that their health data was being monitored in the first place. Some KIs felt that “telemedicine” and “telehealth” should be excluded from the scope; however, because these terms have no standard definitions and because many studies of key PGHD technologies, such as mobile apps and wearables, might use those terms, we decided not to exclude studies for using those terms. Instead, the requirement that the studies use a consumer-purchasable device would help focus the project.

Regarding outcomes, the KIs agreed on the importance of health outcomes (e.g., survival, symptoms, quality of life) but disagreed in the importance of non-health outcomes that define certain medical conditions, such as weight for obesity, blood pressure (BP) for hypertension, and hemoglobin A1c (HbA1c) for diabetes. While these latter outcomes are not “health” outcomes, they do influence clinicians’ decision-making. The review team decided to include and tabulate these data, but to maintain an overall focus on the health outcomes.

Regarding possible confounding variables, KIs felt it was important to consider the degree of patient engagement, noting that patients vary greatly in the degree to which they actually use PGHD devices during a study. Further, study participants may be more motivated to engage than participants outside studies, and this issue can inform a discussion of applicability. Other applicability concerns raised by KIs involved the possibility that study participants have
unrepresentatively high familiarity and comfort with technology (e.g., ready internet access or already have a cell phone) and are more likely to be from urban (not rural) areas.

In summary, KI input helped the review team expand the list of clinical conditions, confirm the focus on consumer technologies without requiring data transmission, refine the approach for surrogate outcomes, and enrich the considerations for applicability.

**PICOTS and Inclusion Criteria**

Table 1 displays the inclusion criteria according to PICOTS.

<table>
<thead>
<tr>
<th>Category</th>
<th>Criteria</th>
</tr>
</thead>
</table>
| **Populations**     | **INCLUDE:** Individuals with one or more of the following 11 chronic conditions: obesity, diabetes prevention, sleep apnea, hypertension, coronary artery disease, heart failure, cardiac arrhythmias or conduction abnormalities, stroke, Parkinson’s disease, chronic obstructive pulmonary disease, or asthma.  
                      | **EXCLUDE:** Individuals with other conditions                                                                                                                                                    |
| **Interventions**   | **INCLUDE:** Consumer health technology, defined as devices consumers use on their own to address health issues and improve quality of life. They are also referred to as direct-to-consumer medical devices, consumer health-based technologies/devices, over-the-counter devices, consumer-based medical devices, and consumer-grade medical devices. Consumer products do not require a health professional, and may be purchased directly by consumers without a prescription. The technology must collect patient data without manual input that can potentially be used by the patient or sent to a healthcare professional (data transmission could be via the same technology or a different technology).  
                      | **EXCLUDE:** PGHD technologies that are not consumer technologies or that rely on manual input.                                                                                                        |
| **Comparators**     | **INCLUDE:** Any comparator is acceptable.                                                                                                                                                           |
| **Outcomes**        | **INCLUDE:** Health outcomes or full economic evaluations. For health outcomes, we defined them differently for different clinical topics:  
                      • We universally included outcomes widely recognized to be important, such as mortality, survival, ER visits, hospital admissions, disease severity, disease progression, and quality of life.  
                      • Outcomes that define a condition (e.g., weight loss for obesity, HbA1c for diabetes, blood pressure for hypertension) were included for studies of that same condition, but were excluded for studies of other conditions (e.g., weight loss for diabetes, blood pressure for obesity).  
                      • Outcomes quantifying the trade-off between cost and effectiveness of interventions, such as cost per QALY, were included if effectiveness was measured using a health outcome, as defined above.  
                      Process outcomes, such as physician-ordered changes in management (e.g., dose alteration, diagnostic testing) were extracted only from studies that reported health outcomes as defined above.  
                      | **EXCLUDE:** Surrogates such as prescription filling behavior, biomarkers that do not define the condition (e.g., blood pressure in patients with obesity), disease knowledge, dietary behavior, steps per day, user satisfaction, or usability. Partial economic evaluations and other cost analyses or descriptions are out of scope as costs are unable to be directly compared across various interventions and are not standardized relative to a consumer or societal outcome of interest. |
| **Timing/Setting**   | **INCLUDE:** No limitations on timing. Setting must be at home or otherwise outside of a hospital or healthcare center.                                                                                      |
| **Study Designs**    | **INCLUDE:** Any study design with a separate comparison group of patients or single-arm registry studies. Systematic reviews will also be included.  
                      | **EXCLUDE:** Narrative reviews, case reports, editorials, comments, letters, meeting abstracts                                                                                                         |
| **Language**        | **INCLUDE:** English                                                                                                                                                                                   |
Search Strategies

We searched MEDLINE and EMBASE (via EMBASE.com), In-Process MEDLINE and PubMed unique content (via PubMed.gov), and the Cochrane Database of Systematic Reviews for systematic reviews or randomized controlled trials published through August 27, 2019. We updated these searches on December 23, 2019. We also searched ClinicalTrials.gov for active studies through April 7, 2020. We also searched the websites of the following organizations for systematic reviews, technology assessments, and guidelines: Agency for Healthcare Research and Quality (AHRQ), U.S. Food and Drug Administration (FDA), American Health Information Management Association (AHIMA), American Medical Informatics Association (AMIA), Healthcare Information and Management Systems Society (HIMSS).

Study Selection

Using DistillerSR (Evidence Partners, Ottawa, Ontario, Canada), three reviewers screened titles, and six screened abstracts and full articles. For titles, only one reviewer was used to assess general relevance to the topic. For abstract screening, two reviewers were necessary to exclude an article from further consideration, but only one was necessary to order the full text. At full text, two reviewers assessed the study against the inclusion criteria and disagreements were resolved by a (senior-level) third reviewer. Full-text screening also involved determining which publications were associated with other included publications (e.g., subsequent publications of the same trial).

To confirm that the included PGHD devices were consumer products, we enlisted the assistance of device engineers. Three engineers examined the list of PGHD devices (manufacturer and model names) and determined whether any of the devices were not available for purchase directly by consumers. Any trials that had only included nonconsumer devices (e.g., devices requiring a prescription) were excluded.

Study-Level Data Extraction

For each included trial, one reviewer extracted general trial information (design, country, enrollment dates, statistical power methods, N at baseline, study duration), patient characteristics (key inclusion criteria, mean age, % female, baseline disease severity, and whether the population was rural), treatment details (including which specific PGHD device[s] were given to which treatment group[s]), risk-of-bias items (see next section), and outcome data (outcome category, specific outcome, results, and statistical tests). We extracted up to six categories of outcomes for each trial:

- Health outcomes (e.g., symptoms, quality of life [QOL], major health events such as strokes/heart attacks, exacerbations, mortality, survival, ER visits, hospital admissions, disease severity, disease progression)
- Surrogate outcomes (e.g., body mass index [(BMI)] for obesity, blood pressure (BP) for hypertension)
- Outcomes pertaining to Guiding Question 2 (interoperability, functions, acceptability/usability, sustainability, feasibility, fidelity, or integration into electronic health records) (only if the trial reported either health outcomes and/or surrogate outcomes)
- Cost-effectiveness (quantifying the trade-off between health effects and cost) (only if the trial reported health outcomes)
• Process outcomes (e.g., medication changes, diagnostic tests ordered, other health-care-centered outcomes) (only if the trial reported health outcomes)
• Whether the article reported cost-only data

Risk of Bias
For each trial, one reviewer assessed the risk of bias using nine items:

1. Random sequence generation?
2. Allocation concealment?
3. Groups similar at baseline or were differences controlled for?
4. Isolated effect?
5. Were outcomes pre-specified and reported?
6. Were participants analyzed based on originally assigned groups across time points?
7. Was attrition low and adherence high?
8. Were outcome assessors and data analysts masked?
9. Were reliable measures of outcomes used consistently across all participants?

We categorized each item as Low, High, or Unclear. We draw the reader’s attention to item 4, “Isolated effect?” With this item, we determined whether the treatments in the study permitted an estimate of a PGHD intervention’s isolated effect. This would be true, for example, if the control group received educational materials only, and the intervention group received the same educational materials along with a single PGHD device and no other interventions. However, in the majority of trials, the intervention group received additional intervention(s) beyond the PGHD intervention, such as an additional PGHD device, or a PGHD device combined with physical activity goals (e.g., 10,000 steps a day) in obesity trials. In these “multicomponent” situations, no isolated estimate is possible, so we assigned High risk of bias to item 4.

After completing the nine items, we examined them together to categorize the trial as Low, Moderate, or High risk of bias.

Results Classification
For isolated effects on health outcomes, we classified the results in one of four categories: 1) likely no effect, 2) unclear, 3) possible positive effect, or 4) likely positive effect. If the results consistently demonstrate the lack of an effect (via narrow confidence intervals around a null effect), we coded it as “likely no effect.” If the results have inconsistency in direction of effect and/or study authors could not reach a conclusion, the findings were coded as “unclear” for that outcome. If one or more outcomes have minor inconsistency in findings, but at least one study with moderate or low risk of bias showed a positive effect, the findings were coded as “possible positive effect.” If the results had a consistent positive effect, we coded it as “likely positive effect.”

When we categorized the data as “unclear” based solely on health outcomes, we then examined surrogate outcomes for that particular clinical condition. We then used these surrogate data, along with the direct effects on health outcomes, to determine which of the four categories was most appropriate.
Evaluation of Economic Evaluations

Economic evaluations were assessed using a modified version of the Consensus Health Economic Criteria (CHEC) tool previously described in the AHRQ report titled “Telehealth for Acute and Chronic Care Consultations.” These criteria related specifically to concerns of internal validity and the potential introduction of bias. All studies, regardless of economic evaluation design, were categorized as “low risk of bias,” “moderate risk of bias,” or “high risk of bias” after reviewing individual components of the modified CHEC tool.

Device Similarity and Usability Testing

For each PGHD device, device engineers assessed similarity to device(s) currently on the market by that manufacturer. They used the following scale:

1. This model is similar to one currently available from this manufacturer
2. This model is SOMEWHAT different than any that are currently available from this manufacturer
3. This model is VERY different than any that are currently available from this manufacturer
4. We could not reliably determine the similarity of this model to the ones currently available from this manufacturer.

As a supplement to the literature review, in February 2020, a usability assessment of the Weight Watchers (WW) app was conducted by five undergraduate students who are affiliated with the Interaction Design and Engineering for Advanced Systems (IDEAS) lab at Rowan University. These students downloaded the WW app onto their mobile phones, all of which were Apple devices, and used $25 iTunes gift cards to pay for a single month membership, which cost $24. (Since the version of the WW app for Android devices available on Google Play did not allow users to use a Google Play gift card to pay the monthly membership fee, and the app appeared to require users to sign up for a counseling session to start, the IDEAS lab was unable to test the app’s Android version.) The student testers used the app for 7-10 days before conducting a heuristic evaluation to identify both positive features and potential usability issues. WW Online was also the intervention investigated by an included obesity study.

Once all students had reviewed the app, a superset of all issues and positive features was created. Then the students they met as a group, facilitated by a human factors psychology professor. During the meeting all feedback was reviewed and, if necessary, explained, and students were asked to assign severity ratings to the issues and priority ratings. Issues were rated on a 0 to 3 scale with 0 = not an issue, 1 = minor, 2 = moderate, 3 = major issue. The suggestions for improvement were rated on a 3 point scale with 1 = consider only if time/resources allow, 2 = worth discussing, 3 = strongly recommended. None of the issues or benefits had an average rating greater than 2. Finally, the professor worked with the students to create two prioritized lists, one listing suggestions for changes that could increase usability and the other listing ideas for future enhancements, new features.

In addition, device engineers provided analyses they had recently completed on two types of consumer PGHD devices that were deemed to be of high interest to users: handheld and wearable blood pressure (BP) devices (for hypertension management) and Smartphone-enabled ECG monitors (for cardiac arrhythmia detection). For both device types, HD engineers performed a range of physical tests, reviewed product literature/specifications, and asked users...
about their experience with the device. They rated the devices in the categories of performance, safety, workflow, interoperability, maintenance, patient experience, and user experience. The rating ranged from Poor to Excellent, and each category rating contributed to an overall rating for the device.

**Data Presentation**

We created six sets of evidence tables—one set per clinical condition. The first listed ongoing clinical trials identified in ClinicalTrials.gov, the second listed general study information, the third listed patient characteristics, the fourth described the treatment groups, the fifth listed the individual risk-of-bias items for each trial included our overall judgment, and the sixth listed the included outcomes and statistical results. See the above section Study-Level Data Extraction for the specific information we extracted.

**Peer Review and Public Commentary**

We invited telehealth experts to provide external peer review of this technical brief; AHRQ and an associate editor also provided comments. The draft report will be posted on the AHRQ website for 3 weeks to elicit public comment. We will address all reviewer comments, revising the text as appropriate, and document everything in a disposition of comments report that will be made available 3 months after AHRQ posts the final technical brief to its website.

**Findings**

**Results of Literature Searches**

Our searches identified 8,521 potentially relevant articles, of which we excluded 5,718 at the title level (not relevant), 1,735 at the abstract level, 158 were secondary publications of other articles, and 363 were systematic reviews. We dual-screened the full text for the remaining 547. The review team included 123 of these, but upon further review of study devices by device engineers, 12 studies had used only non-consumer devices and were therefore excluded (see the list of 436 full-text exclusions in Appendix Table B-1). The remaining 111 unique studies were described in 157 publications.

**Overview of Evidence**

Among the 11 chronic conditions, hypertension (50 studies) and obesity (43 studies) had the number of included studies most (Table 2). The most common device categories were pedometers, accelerometers, and BP monitors.

<table>
<thead>
<tr>
<th>Chronic Condition</th>
<th># of Included Studies</th>
<th>Typical Device Categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obesity</td>
<td>43</td>
<td>Pedometers, accelerometers, scales</td>
</tr>
<tr>
<td>Diabetes prevention</td>
<td>3</td>
<td>Pedometers</td>
</tr>
<tr>
<td>Sleep apnea</td>
<td>3</td>
<td>Pedometers</td>
</tr>
<tr>
<td>Hypertension</td>
<td>50</td>
<td>BP monitors</td>
</tr>
<tr>
<td>Coronary artery disease</td>
<td>6</td>
<td>BP monitors, heart rate monitors</td>
</tr>
<tr>
<td>Heart failure</td>
<td>6</td>
<td>BP monitors, scales, heart rate monitors</td>
</tr>
<tr>
<td>Cardiac arrhythmias or conduction abnormalities</td>
<td>4</td>
<td>ECG monitors</td>
</tr>
<tr>
<td>Stroke</td>
<td>1</td>
<td>BP monitors</td>
</tr>
<tr>
<td>Parkinson’s Disease</td>
<td>0</td>
<td>NA</td>
</tr>
<tr>
<td>Chronic Condition</td>
<td># of Included Studies</td>
<td>Typical Device Categories</td>
</tr>
<tr>
<td>-------------------</td>
<td>-----------------------</td>
<td>--------------------------</td>
</tr>
<tr>
<td>COPD</td>
<td>9</td>
<td>Accelerometers, pedometers</td>
</tr>
<tr>
<td>Asthma</td>
<td>1</td>
<td>Spirometers, BP monitors</td>
</tr>
</tbody>
</table>

BP – Blood pressure; ECG – Electrocardiogram; COPD – Chronic obstructive pulmonary disease; N.A. Not applicable. The numbers in the table add to more than 111 because some studies were included for multiple chronic conditions.

Eighty-three of the 111 trials used a single PGHD device, 22 used 2 PGHD devices, and the other 6 used 3 or more PGHD devices. When the same study used two PGHD devices, it was most commonly either pedometer and scale or BP monitor and scale. Regarding device similarity, BP monitors were often similar to those currently available, whereas other devices were mostly rated either as similar or of unknown similarity. A graphic showing the six most common device categories appears in Figure 1 below.
Figure 1. Counts of Unique Devices in Categories

<table>
<thead>
<tr>
<th>Blood pressure monitors</th>
<th>Accelerometers</th>
<th>Pedometers</th>
</tr>
</thead>
<tbody>
<tr>
<td>48 devices</td>
<td>17 devices</td>
<td>19 devices</td>
</tr>
<tr>
<td>43 devices</td>
<td>8 devices</td>
<td>5 devices</td>
</tr>
</tbody>
</table>

- **Blood pressure monitors**: 48 devices, rated as “Similar” to current devices on the market by the same manufacturer.
- **Accelerometers**: 17 devices, rated as “Somewhat different” to current devices on the market by the same manufacturer.
- **Pedometers**: 19 devices, rated as “Very different” to current devices on the market by the same manufacturer.

**Note:** All specific devices are listed in Appendix Table C-1, along with similarity ratings, clinical condition(s), and specific studies in which the devices were used. Single devices not listed in the graphic above include an ECG monitor (rated as similar), an energy expenditure and physical activity monitor (unknown similarity), forehead thermometer (rated as similar), pedal machine (rated as similar), pulse oximeter (unknown similarity), and a spirometer (rated as similar), a bite counter (rated as similar), a body composition monitor (rated as somewhat different), and a combined accelerometer, heart rate monitor, galvanic skin response monitor, and cutaneous temperature monitor (rated as similar).

Current costs of PGHD devices are best obtained searching the internet, or by visiting specific manufacturer’s websites. Manufacturers whose device(s) were used in at least three studies include Omron (see www.omronhealthcare.com for blood pressure monitors and pedometers), A&D (see https://medical.andonline.com/home for blood pressure monitors and scales), Yamax (see https://www.yamaxx.com/digi/ for pedometers), Fitbit (see www.fitbit.com for accelerometers), Tanita (see www.tanita.com for scales), Microlife (see www.microlife.com for blood pressure monitors), and AliveCor (see www.alivecor.com for ECG monitors).

In the sections below, for each of the 11 clinical conditions, we address the 5 Guiding Questions. Appendix C contains all evidence tables.
Obesity

Guiding Question 1. Which specific consumer automated-entry PGHD applications/technologies/devices have been studied for measurement of health outcomes? What studies are in progress of consumer PGHD devices?

   a. What study designs have been used?
   b. What were the inclusion/exclusion criteria?
   c. What statistical analysis and data were used to determine study size and power?
   d. How long were patients followed?
   e. How was adherence measured?
   f. What was the comparator?
   g. Which outcomes were measured?

Devices

The 43 obesity trials used 50 different PGHD devices: 17 pedometers, 13 accelerometers, 8 scales, 5 BP monitors, 2 glucose meters, 1 pedal machine, 1 bite counter, 1 energy expenditure and physical activity monitor, 1 body composition monitor, and 1 combined accelerometer/heart rate monitor/galvanic skin response monitor/cutaneous temperature monitor. All these devices (along with devices included in trials of other clinical conditions) appear in Appendix Table C-1. The similarity judgments (how similar each device is to those currently on the market by this manufacturer) were:

- Pedometers:     8 similar, 5 somewhat different, 0 very different, 4 unknown
- Accelerometers: 5 similar, 1 somewhat different, 0 very different, 7 unknown
- Scales:          3 similar, 1 somewhat different, 0 very different, 4 unknown
- BP monitors:     3 similar, 0 somewhat different, 0 very different, 2 unknown
- Glucose meters: 1 similar, 0 somewhat different, 0 very different, 1 unknown
- Pedal machine:   1 similar, 0 somewhat different, 0 very different, 0 unknown
- Bite counter:    1 similar, 0 somewhat different, 0 very different, 0 unknown
- Energy expenditure and physical activity monitor: Unknown
- Body composition monitor: Somewhat different
- Combined accelerometer, heart rate monitor, galvanic skin response monitor, and cutaneous temperature monitor: Similar

Studies in progress

We identified 43 records in ClinicalTrials.gov that potentially involved PGHD interventions to prevent or treat obesity. As of April 7, 2020, 20206 were not yet recruiting patients, 26 were recruiting, 2 were enrolling by invitation, and 9 were active but not recruiting. More details about these records (including hyperlinks) appear in Appendix Table C-2.

Study designs

All 43 included obesity studies were randomized trials. Twenty-seven trials were conducted in the United States, and other prominent countries included South Korea (5 trials), Australia
(2 trials), and the United Kingdom (2 trials). The median number of patients at baseline was 96 (interquartile range [IQR] 44 to 177). Patient enrollment dates (reported by 29 trials) ranged from September 2001 to October 2017, the median study enrollment period was 1 year (IQR 6 months to 24 months), and the median month of patient enrollment was March 2012. The next sections describe additional aspects of these trials, which are tabled in detail in Appendix Table C-3, C-4, and C-5.

**Inclusion/exclusion criteria**

Studies generally included obese or overweight adults who did not have other major medical conditions and were willing and able to start a new exercise regimen. Seven trials enrolled only older adults (e.g., age ≥60), and another seven trials enrolled only children, adolescents, or young adults. Seventeen trials enrolled only patients with a specific comorbidity (e.g., 3 trials of overweight patients with sleep apnea). Twenty-one trials required comfort with technology, such as already owning a smartphone and/or the willingness and ability to receive ongoing text messages as part of a weight-loss intervention.

**Statistical power analyses**

Twenty-seven of the 43 trials conducted *a priori* power analyses. Eight had based effect-size estimates on prior work and had also accounted for anticipated attrition. Another 12 accounted for anticipated attrition, but did not state the bases for anticipated effect sizes. Another five stated the bases for anticipated effect sizes but did not account for anticipated attrition. The other two mentioned neither prior work nor attrition.

**Follow-up length**

The median followup length in the 43 obesity trials was 26 weeks (6 months), with an IQR from 3 to 12 months.

**Adherence measurement**

Device adherence was typically measured by the number of days during followup that patients used the consumer device or app or provided data to study staff.

**Comparators**

The 43 obesity trials were notable for the complexity of their comparisons. Twenty-seven of 43 trials had 2 treatment groups, 14 had 3 groups, 1 had 4 groups, and 1 had 5 groups. Four trials had a waitlist control group (i.e., not seen by study staff during the trial), and 30 trials had a control group seen by study staff during the trial but not receiving the PGHD device. In five trials, the control group did use at least one PGHD device (e.g., just a scale), and their outcomes were compared to an intervention group that received multiple PGHD devices (e.g., both a scale and an accelerometer). In 11 trials, 2 groups received the same device(s), but one group’s treatment was enhanced in some way (e.g., Facebook group to share performance). In five trials, a control group received a PGHD device in a different device category (e.g., a scale instead of an accelerometer), or the control group received a different PGHD device in the same category (e.g., a Fitbit Aria scale vs. a Tanita digital scale).
Outcomes reported

Only 16 of the 43 obesity trials reported a health outcome, which was invariably either QOL or AEs adverse effects. Almost all (41 of 43) reported a surrogate outcome (BMI or weight), and most (30 of 43) reported data informing device usage or ease of use or technical problems (which are the subject of Guiding Question 2, which appears next). None reported process outcomes, cost-only data, or cost-effectiveness data.

Guiding Question 2. What are the characteristics (e.g., interoperability, functions, acceptability/usability, sustainability, feasibility, fidelity, or integration into electronic health records) of these specific consumer automated-entry PGHD technologies?

Below, we discuss the pertinent obesity data for this Guiding Question separately for accelerometers/pedometers, scales, and other devices. All these data appear in the top section of Appendix Table C-7. After these sections, we discuss our usability testing of the WW App.

Accelerometers/pedometers

Patient usage of these devices was highly variable across the trials, ranging from 37% to 95% of trial days (median 64%). Among those who consistently used their devices, they wore it for most of their waking hours (range 9 to 14 hours a day). Satisfaction and ease of use were rated highly by patients in almost all the studies reporting such data. For example, in the study by Cadmus-Bertram et al. (2015), all 26 patients in the intervention group “liked” the accelerometer, and 76% would recommend it to a friend (56% preferred a clip-on, 29% preferred wrist-worn, and the other 24% had no preference). One exception to the general positive reaction was in the study by Smith et al., (2019) in which only two-thirds of patients reported a “positive impact of the device,” whereas the other 33% “negatively viewed the device.” Regarding technical problems, six studies reported some issues involving “equipment malfunction” (not further described by the authors), lost or broken device, difficulty downloading or logging in or saving step counts, and incompatibility with certain smartphones.

Scales

Use of body weight scales was also highly variable (18% to 79% of study days). One study reported ease of use of body weight scales and had found that 66% of patients (19/29) rated the scale as “helpful.” Two studies reported technical problems, with one reporting issues with Internet connectivity preventing scale initialization, and the other did not report what the technical problems with the scale were.

Other devices

The study by Yoo et al. (2009) reported that patients sent their BP readings (using the Omron T5M BP monitor) an average of 1.72 times/day (86% adherence). Green et al. (2014) provided patients with a BP monitor (Omron 711DLX), and they found it “extremely helpful” to monitor their BP at home and share readings with their provider.

The Bite counter, used by Turner-McGrievy et al. (2017) as a device that is now available for either iPhone or Android as an app, detects hand movement when eating in order to count bites. The study found that patients used it an average of 1.9 meals/day and 3.6 days/week.
Oh et al. (2015)\textsuperscript{11,12} reported ease-of-use data on their “SmartCare” service, which included the body composition monitor by Biospace (InBody IH-U070B). Patients rated the convenience of using this device at a mean of 3.5 out of 5 possible points.

Mameli et al. (2018)\textsuperscript{13} gave patients the Empatica E3 wristband from Italy, which is not only an accelerometer, but also a heart rate monitor, galvanic skin response monitor (sweat gland activity), and cutaneous temperature monitor. Patients used the wristband only 37% of days.

Jakicic et al. (2016)\textsuperscript{14} reported usage data for a combined energy expenditure and physical activity monitor called the FIT Core by Body Media. Patients wore the device a median of 31% of days (170 days in the 18-month period), for a median of 4 hours/day.

Yoo et al. (2009)\textsuperscript{7} provided patients the Anycheck glucose meter by Insung, and they sent blood glucose recordings an average of 1.84 times per day (adherence rate 92%).

Carr et al. (2013)\textsuperscript{15} provided workers access to a pedal machine (MagneTrainer by 3D Innovations) at their workplace. They used it an average of 38% of possible days, for about 30 minutes per day. Workers rated the pedal machine biofeedback as extremely helpful (median of 5 on scale from 1 to 5).

**Usability Testing of the Weight Watchers Online App**

Regarding the WW Online app, all usability comments and recommendations from the five students at Rowan University appear in Appendix Table C-8. Note that this app was also tested in the included study by Thomas et al. (2017),\textsuperscript{16} which enrolled patients from May 2013 to March 2014, and the app was likely altered in the intervening 6-year period.

Overall, the students found the WW app to be well-designed and relatively easy to use. Some of the features that were highlighted as particularly beneficial include 24/7 chat to get immediate answers; the recipes feature; the ability to enter foods through different methods, including voice-recognition, bar code scanning, and manual entry supported by searching an extensive database (e.g., by brand names or restaurant names); and the activities and training videos that help users integrate exercise into their daily schedules. They also found the initial landing/home page and the success checklist to be well-organized and easy to navigate and the information in the FAQs section to be helpful.

Suggestions for improving usability include adding a + sign to the food recording screen rather than just showing a search option, using arrows (conventional for mobile apps) rather than buttons (unconventional) to navigate between weeks, adding a button or link to allow users to skip the tutorial that is shown on the Success Checklist page rather than requiring users to view the whole video before interacting further with the app, automatically launching a tutorial or presenting instructions or a guide upon first opening the app, and organizing the FAQs into topic areas. Suggestions for feature enhancements/upgrades include adding personalization to the recipes feature for users with specialized diets (e.g., low salt, vegan, gluten free), allowing users to access their total nutritional intake per day and providing tips connected to specific aspects of their diets (watch sodium, need vitamin C), and providing pointers to new articles or reminders of how healthy foods can help improve aspects of users’ health. After using the app for at least a week, students were also unclear about a few things (e.g., why is goal weight accessed through the settings section rather than entered during initial setup? and what would happen if they indicated they wanted “a little extra support” at a particular time of day?)
Guiding Question 3. What is the influence of specific consumer automated-entry PGHD technologies on health outcomes? Does this vary across different patient populations, different settings, or other modifiers of effectiveness?

Below, we first provide an overall summary of the most relevant data: studies that investigated whether there is an isolated effect of PGHD technologies on health outcomes. Because these data did not lead to a clear statement, we then discuss surrogate outcomes (i.e., BMI or weight) reported by isolated-effect studies, to determine whether we could use the surrogate data to make a statement about health outcomes. We then discuss studies investigating the multicomponent effects of PGHD technologies on either health outcomes or surrogate outcomes. All risk-of-bias assessments are in Appendix Table C-6, and the results are in Appendix Table C-7.

Isolated Effects on Health Outcomes

For preventing or treating obesity, three trials have examined whether PGHD technologies have isolated effects on health outcomes. In all three trials, the specific health outcome was QOL, which was measured using the SF-12 in two trials, and the SF-36 in one trial. We next discuss each trial in detail.

Fukuoka et al. (2019)\textsuperscript{17-22} (Low risk of bias) compared three groups: control group (Omron active style Pro HJA-350IT, which showed daily steps/intensity), regular (same accelerometer, plus 3-month access to the “mPED” app, which provided daily messages and physical activity diary, and “plus” (same as regular group except that they had 9-month app access). Researchers measured QOL at both 3 months and 9 months. Physical QOL for the two groups with app access was better than that for the control group at 3 months (means not reported, but $p=0.04$), but was not better than that for the control group at 9 months. For mental QOL, authors found no difference between groups at either time point.

Smith et al. (2019)\textsuperscript{5} (Moderate risk of bias) compared two groups: Exercise (16-week home-based program, with weekly calls to the study physiologist) and Exercise plus fitness tracker (same program and either the Fitbit Flex or the Fitbit One, which provided feedback on steps, stairs, and calories expended). At 16 weeks, authors found no statistical significant differences in either mental or physical QOL between the 2 groups.

Richardson et al. (2016)\textsuperscript{23} (High risk of bias) compared three groups: time-based walking goals (no PGHD device, but they set walking goals), simple pedometer (Yamax Digiwalker SW 200 and instructed to wear it daily, and step counts were reviewed over the phone with study staff, and step count goals modified accordingly), and “enhanced” pedometer (SportBrain iStep X), with the same step count review and goal modification, and on a website they received motivational messages, viewed graphs of their step counts over time, and engaged with an online community. For the SF-12 physical scores, the study found no difference in QOL between groups from baseline to 26 weeks. However, for the SF-12 mental scores, the two pedometer groups had larger improvements in mental QOL. Baseline scores were about 37-38 in all 3 groups (using a scale range 0-100 where higher scores are better; these baselines are considerably lower than published population averages of 70-80).\textsuperscript{24} Scores had improved to about 40 in the pedometer groups, but had decreased to 36 in the no-device group. This is approximately a 4% difference in mental QOL. We rated the study at high risk of bias due to...
several factors, including unclear randomization and allocation concealment, relatively high attrition, and unmasked outcome assessors.

In summary, we categorized this evidence as “unclear”. One low risk of bias study found a positive effect at 3 months that vanished by 9 months, and the single moderate-quality study found no statistically significant effects. The one study that did report a persistent positive effect (about a 4% difference in mental QOL, but no effect on physical QOL) we rated at high risk of bias.

Isolated Effects on Surrogate Outcomes

Due to the lack of clarity about health outcomes, we turned to surrogate outcomes. Sixteen obesity trials reported isolated effects on BMI or weight. We prioritized the BMI data because it controls for height, and only extracted weight data authors did not report BMI. To help interpret BMI reductions, note that for a person of average height (5’7” or 1.72 meters), a BMI reduction of 1 unit corresponds to 3 kilograms, which is 6.6 pounds. Due the complexity of treatment groups and comparisons (the 16 trials had 44 treatment groups), we separately discuss the results for 3 types of comparisons:

- Device presence: PGHD device vs no PGHD device, or adding another PGHD device
- Across-category. One PGHD device vs a PGHD device in a different category (e.g., accelerometer vs scale)
- Within-category. One PGHD device vs a different PGHD device in the same category (e.g., 1 pedometer vs another)

Fourteen of the 16 trials reported the effect of device presence, and the point estimates for their surrogate outcome data appear in Figure 2 (BMI) and Figure 3 (weight). Each point is a study result reflecting the between-group difference in change scores. One of the 14 studies appears in neither graph because authors only reported that the between-group difference was not statistically significant (no numerical data reported).
Figure 2. BMI differences in studies of isolated effects of device presence/absence

***statistically significant between-group difference

Each filled circle is a point estimate of the difference in BMI change scores. Points to the left of the middle line suggest that the PGHD technology resulted in greater weight loss than without the PGHD technology, whereas points to the right of the middle line suggest the reverse. The two dashed lines represent the minimal important difference of 5% body weight (differences less than this are considered not clinically important). For BMI, assuming patients were of average height, this is about 1.67 BMI units (or 5 kilograms), based on the overall average baseline weights.

Figure 3. Weight differences in studies of isolated effects of device presence/absence

***statistically significant between-group difference

Each filled circle is a point estimate of the difference in weight change scores. Points to the left of the middle line suggest that the PGHD technology resulted in greater weight loss than without the PGHD technology, whereas points to the right of the middle line suggest the reverse. The two dashed lines represent the minimal important difference of 5% body weight (differences less than this are considered not clinically important). For BMI, assuming patients were of average height, this is about 1.67 BMI units (or 5 kilograms), based on the overall average baseline weights.

Only 4 of the 14 trials found any statistically significant differences:
- Shin et al. (2017)\(^26,27\) (low risk of bias) had given two of three groups an accelerometer (Suwon Fitlife), and one of them had greater BMI reduction at 12 weeks than the
nondevice group (1 BMI unit lost vs 0.2 BMI units lost), but the other device group did not (0.5 BMI units lost).

- Richardson et al. (2016)\textsuperscript{23} (high risk of bias) found larger weight reductions in their enhanced pedometer group (SportBrain iStep X) then in their time-based walking goals group. They had lost an average of about 2 kg over 6 months, whereas the nondevice group had only lost 0.7 kg. Their simple pedometer group was not statistically significantly different in weight loss compared with the nondevice group.

- In Thomas et al. (2017)\textsuperscript{16} (low risk of bias), one of the two app groups (WW app alone) had statistically significantly greater weight reduction at 13 weeks (mean reduction of 2.7 kg, representing a 2.9\% reduction from baseline) than the nondevice group (mean reduction of 1.3 kg, representing a 1.5\% reduction from baseline), but there was no difference at 1 year. The other app group (WW app plus ActiveLink) had no different weight reduction from the nondevice group at either time point.

- Nicklas et al. (2014)\textsuperscript{28} (moderate risk of bias) found that their accelerometer group (Suzuken Lifecorder Plus) had great weight reduction at 44 weeks (average 8.5 kg lost) than their nondevice group (average 5 kg lost).

Overall, there is \textit{likely no effect of PGHD technologies on health outcomes} when used to prevent or treat obesity. We base this statement on four observations:

- Direct evidence on health outcomes is mostly unreported, but when reported, it is inconsistent.

- Evidence on the surrogate outcomes (BMI and weight) consistently suggest no effect. All 25 point estimates in the above figures indicate less than a 5\% body weight difference between groups. All four statistically significant results had point estimates less than 5\% body weight.

- Regarding studies’ \textit{a priori} power analyses, 5 of the 14 studies reported the outcome on which power analyses were based. This was weight/BMI for 4 studies, \textsuperscript{14,16,23,29} and steps per day in one study.\textsuperscript{30} Therefore, taken together with the small point estimates, low power was likely not the reason for the general lack of statistical significance.

- If there were an effect on health outcomes, it likely would have been seen in the BMI/weight data.

All of the above data (isolated effects of device presence on either helth outcomes or surrogate outcomes) are summarized in Table 3 below.
Table 3. Obesity: Categorization of Isolated Effects Due to Device Presence/Absence

<table>
<thead>
<tr>
<th>Health Outcomes</th>
<th>All results were p&lt;0.05 in favor of PGHD</th>
<th>Mixed results,* but most p&lt;0.05 results favored PGHD</th>
<th>No results were statistically significant</th>
<th>Mixed results,* but most p&lt;0.05 results against PGHD</th>
<th>All results p&lt;0.05 against PGHD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality of life</td>
<td></td>
<td>● ○</td>
<td>○</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Health outcomes not reported</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>????????????????</td>
</tr>
</tbody>
</table>

Statement about health outcomes based only on health outcomes data: **Unclear effect**

<table>
<thead>
<tr>
<th>Surrogate Outcomes</th>
<th>BMI</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>●</td>
<td>○</td>
</tr>
<tr>
<td></td>
<td>●</td>
<td>○○</td>
</tr>
<tr>
<td>Surrogate outcomes not reported</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Statement about health outcomes based on both health outcomes and surrogate outcomes: **Likely no effect on health outcomes**

Note: This table only displays studies that used an isolated-effect design to compare the presence vs absence of a PGHD device.

- ● Low risk of bias study of an isolated effect
- ○ Moderate risk of bias study of an isolated effect
- ○ High risk of bias study of an isolated effect
- ? Study did not report this category of outcome (the number of question marks indicates the number of studies)

* “Mixed results” means that a study investigated either multiple time points (e.g., both 13 weeks and 26 weeks) or multiple comparisons (e.g., PGHD device A vs no device, and PGHD device B vs no device), and data showed statistical significance for some but not all time points/comparisons.

BMI – Body mass index
PGHD – Patient generated health data
There were 6 across-category comparisons among the 16 trials; 2 were statistically significant:

- Turner-McGrievy et al. (2017)\textsuperscript{9,10} (moderate risk of bias) found that those who used the Calorie Counter app had greater BMI reduction at 26 weeks (average 6.8 BMI units) than those who received the Bite Counter (average reduction 3 BMI units).
- Ross et al. (2016)\textsuperscript{31} (moderate risk of bias) had given both an accelerometer (Fitbit Zip) and a smart scale (Fitbit Aria) to two of their three treatment groups, with one of these two groups also receiving reminder phone calls from study staff. They compared both groups to a third group that received only a pedometer (specific device not reported). The study found greater weight reduction among the Tech+Phone group at 26 weeks (average 6.4 kg lost) than the pedometer group (average 1.3 kg lost). However, the Tech-only group’s reduction (average 4 kg lost) was not statistically different than that for the pedometer group.

Finally, there was one within-category comparison among the 17 trials. Richardson et al. (2016)\textsuperscript{23} (high risk of bias) found that an enhanced pedometer group (SportBrain iStep X) had lost more weight (1.9 kg at 26 weeks) than the simple pedometer group (Yamax Digiwalker SW 200, 0.6 kg at 26 weeks).

**Multicomponent Effects on Health Outcomes**

Ten obesity trials reported multicomponent effects on health outcomes (“multicomponent” means that patients received other intervention[s] in addition to PGHD device[s], making it impossible to attribute improvements to the devices themselves). Therefore, we do not discuss each study in detail, but rather describe general aspects of the studies and results and refer the reader to Appendix Table C-7.

Six\textsuperscript{8,32-36} of the 10 studies were at high risk of bias, and the other four\textsuperscript{37-42} were at moderate risk of bias. Typical issues involved nonisolated effects, unclear randomization and allocation concealment, and unblinded outcome assessors. The reported health outcome was QOL in all 10 trials, and they used 11 different QOL instruments.

Eight of the 10 trials reported that QOL between groups was statistically nonsignificant at followup, regardless of the specific instrument used. By contrast, Haggerty et al. (2017)\textsuperscript{34} (high risk of bias) found that SF-12 scores were better in their telemicine group (the group received a Withings WiFi scale and recorded diet and physical activity on a website) than in a group receiving feedback text messages along with a different scale (Eat Smart Precision Digital Scale). Interestingly, however, a third group that received no PGHD device (“enhanced usual care”) had better mean QOL than the telemicine group, though authors did not report whether the difference was statistically significant. The other study reporting a statistically significant effect was Vorrink et al. (2016)\textsuperscript{35,43} (high risk of bias). The study’s two groups were usual care (no intervention) and intervention (which included an accelerometer embedded in a smartphone, physical activity goals, and physiotherapist monitoring via a website for step viewing and goal adjustment). Authors used the chronic respiratory questionnaire to measure QOL. For the dyspnea subscale, authors found no statistically significant difference between groups at follow-up. For the fatigue subscale, outcome data showed a statistically significant between-group difference, and authors stated “this was probably caused by great variability in the data rather
than the intervention.” This comment reveals a misunderstanding of statistical testing, which accounts for data variability along with effect sizes.

**Multicomponent Effects on Surrogate Outcomes**

Due to this evidence’s highly indirect nature, we do not discuss it; instead, we refer the reader to the tabulated data in Appendix Table C-7.

Of the 41 obesity-related records in clinicaltrials.gov, only 3 made PGHD-related comparisons and stated that they were collecting data on a health outcomes (quality of life).

**Guiding Question 4. What are the harms or adverse events associated with these specific consumer automated-entry PGHD technologies? Which patients in which settings are most at risk of harms?**

Only 6 of the 43 obesity trials reported data on adverse events (AEs) (see the Health Outcomes section of Appendix Table C-7). The reported types of events were:

- Overall number# of events
- Number of emergency room or urgent care visits
- Number# of significant medical/mental problems that restricted walking at least 7 consecutive days
- AEs requiring hospitalization
- Serious AEs
- Serious event alerts (device alert triggered when the patient had an overnight hospitalization or surgery)
- Depression alerts (device alert triggered when the patient had a score of 13 or greater on the CES-D questionnaire)
- Nonserious event alerts (study did not report how this alert was triggered)
- Rapid weight loss event alerts (device alert triggered when the patient experienced 6% or greater weight loss during a 4-week period)
- Resting BP alerts (device alert triggered when resting systolic blood pressure [SBP] ≥140 or resting diastolic blood pressure [DBP] ≥90

Only two of the six trials reported whether the between-groups comparison was statistically significant. Arbillaga-Etxarri et al. (2018)\(^3\) reported a statistically nonsignificant difference (p=0.363) in the rate of any AEs between their usual care group that received no PGHD device (73%) and their Urban Training group (77%) that had received six-components, which had included a pedometer (OnStep 50 Geonaute). The other study was Fukuoka et al., (2019)\(^1\) which compared a control group given an accelerometer (Omron active style Pro HJA-350IT) to an intervention group that also received an app, and authors reported three different types of AEs:

- For any AEs, there were nonsignificant findings for both the first 3 months of the study (p=0.23) and the next 6 months (p=0.46). However, the bases for the study power calculations were not harms, but rather efficacy.
For AEs requiring hospitalization, authors reported no statistical between-group comparison, but 5 events occurred among 69 control group participants and 2 events among 140 intervention group participants.

For AEs of significant medical/mental problems that restricted walking at least 7 consecutive days, at 3 months authors found a statistically significantly higher rate among the intervention group (20%) than among the control group (9%) (p=0.05). However, for the subsequent 6 months, the between-group difference was not statistically significant (intervention 34%, control 27%, p=0.71).

Overall, few obesity trials reported AEs (6/43 or 14%), they only occasionally made statistical comparisons (2 of 6 trials), and there were inconsistent results from those 2 trials (all tests statistically nonsignificant, except for 1 test favoring a control group at 1 of 2 time points).

Guiding Question 5. For the technologies demonstrating associations with outcomes of interest, what full economic evaluations provide information on the relative value for consumers?

None of the 43 included obesity trials reported cost-effectiveness data or cost-only data.

Diabetes Prevention

Guiding Question 1. Which specific consumer automated-entry PGHD applications/technologies/devices have been studied for measurement of health outcomes? What studies are in progress of consumer PGHD devices?

- a. What study designs have been used?
- b. What were the inclusion/exclusion criteria?
- c. What statistical analysis and data were used to determine study size and power?
- d. How long were patients followed?
- e. How was adherence measured?
- f. What was the comparator?
- g. Which outcomes were measured?

Devices

The three diabetes prevention trials used six different PGHD devices. The similarity judgments (how similar each device is to those currently on the market by this manufacturer) were:

- Two pedometers: Gruve pedometer (judged as Similar) and the Omron HJ-150 (judged as somewhat different)
- One accelerometer: Jawbone SenseWear armband: Unknown similarity
- One BP monitor: A&D UA-767 Plus BT: Similar
- One glucose meter: Lifescan OneTouch Ultra2: Similar
- One heart rate monitor: Suunto Memory Belt: Very different
All these devices (along with devices included in trials of other clinical conditions) appear in Appendix Table C-1.

**Studies in progress**
We identified four potentially relevant records in ClinicalTrials.gov. More details about these records (including hyperlinks) appear in Appendix Table C-9.

**Study designs**
Two\(^{42,44}\) of the three trials were randomized, and the other\(^{45}\) was nonrandomized. In the nonrandomized study, authors stated that “all participants attending appointments during week 1 were randomized to the intervention group, those attending appointments during week 2 were randomized to the control group, and so on.” Therefore, assignment to groups was alternating, not random.

One randomized controlled trial (RCT)\(^{44}\) was from the United States (89 patients enrolled, 26-week followup, median patient enrollment month January 2011), another RCT\(^{42}\) was from the United Kingdom (177 patients enrolled, 1-year followup, median patient enrollment month June 2011), and the nonrandomized trial\(^{45}\) was from Canada (149 patients enrolled, 1-year followup, median patient enrollment month June 2010). The next sections describe additional aspects of these trials, which are tabulated in detail in Appendix Table C-10, C-11, and C-12.

**Inclusion/exclusion criteria**
The U.S. RCT included obese adults in central Iowa (95% Caucasian), the U.K. RCT included young obese adults at risk of type 2 diabetes, and the Canadian study included rural adults who had at least two risk factors of metabolic syndrome.

**Statistical power analyses**
The U.S. RCT did not mention power analysis. The U.K. RCT’s power analysis had used prior data and had anticipated attrition. The Canadian study did not refer to prior data but did plan for attrition.

**Follow-up length**
One RCT\(^{44}\) followed patients for 6 months, and the other two studies followed patients for 1 year.

**Adherence measurement**
Only the Canadian study reported device adherence, which was measured as the percentage of measurements that patients completed, separately for BP measurements, fasting plasma glucose (FPG) measurements, pedometer measurements, and body weight measurements.

**Comparators**
The U.S. trial enrolled three groups: Guided weight loss (no device, but had weekly meetings with a health coach), physical activity monitor (SenseWear armband), and combined (received both of the other groups’ interventions). The U.K. trial enrolled two groups: control (educational leaflet) and intervention (Gruve accelerometer). The Canadian study enrolled two groups: active control (tailored exercise program) and intervention (same exercise program plus a smartphone
(Blackberry Curve 8300 or 8530), app (Healthanywhere), BP monitor (A&D UA-767PBT), glucose meter (Lifescan One Touch Ultra2), and heart rate monitor (Suunto Memory Belt).

**Outcomes reported**

The U.K. trial (QOL) and the Canadian study (adverse effects) reported health outcomes. All three trials reported surrogate outcomes: fasting glucose and HbA1c by both the U.K. trial and the Canadian trial; 2-hour glucose test by the U.K. trial; and metabolic syndrome score by the U.S. trial. Only the Canadian study (discussed next) reported outcomes pertaining to Guiding Question 2. None reported process outcomes, cost-only data, or cost-effectiveness data.

**Guiding Question 2. What are the characteristics (e.g., interoperability, functions, acceptability/usability, sustainability, feasibility, fidelity, or integration into electronic health records) of these specific consumer automated-entry PGHD technologies?**

The U.S. trial and the U.K. trial did not report these data. The Canadian study reported the percentage of measurements that patients provided at various time points (see data in Appendix Table C-14):

- For BP measurements, the overall percentage was 83%, which had decreased from 91% in weeks 1-12 to 78% in weeks 25-52.
- For fasting plasma glucose measurements, the overall percentage was 82%, which had decreased from 90% in weeks 1-12 to 77% in weeks 25-52.
- For pedometer measurements, the overall percentage was 71%, which had decreased from 84% in weeks 1-12 to 63% in weeks 25-52.
- For body weight measurements, the overall percentage was 42%, which had decreased from 64% in weeks 1-12 to 28% in weeks 25-52.

**Guiding Question 3. What is the influence of specific consumer automated-entry PGHD technologies on health outcomes? Does this vary across different patient populations, different settings, or other modifiers of effectiveness?**

The risk-of-bias assessments are in Appendix Table C-13, and the results are in Appendix Table C-14.

**Isolated Effects on Health Outcomes**

None of the three diabetes prevention studies reported whether there were isolated effects on health outcomes. As a result, we rated this evidence as unclear.

**Isolated Effects on Surrogate Outcomes**

The U.S. study by Peyer et al. (2017)\(^44\) (high risk of bias) found that the metabolic syndrome score (the sum of the z scores for 5 metabolic syndrome risk factors; 0 represents average risk, and high positive values represent higher metabolic syndrome risk) had improved greater in the group receiving both guided weight loss treatment and a SenseWear armband (average 2-point improvement from a baseline of +0.7) than in the group that received only guided weight loss
treatment (average 0.8-point improvement from a baseline of -0.4). The armband-only group (average 1.1-point improvement from a baseline of -0.2) was not statistically different from the other two groups. The other two studies had multicomponent effects.

Overall, there is an *unclear effect of PGHD technologies on health outcomes* when used to prevent diabetes (Table 4). Only one of three trials used an isolated-effect design, and even though the trial did find lower risk of metabolic syndrome among those using an armband, the trial did not report health outcomes, and was at high risk of bias.
<table>
<thead>
<tr>
<th>Health Outcomes</th>
<th>All results were p&lt;0.05 in favor of PGHD</th>
<th>Mixed results,* but most p&lt;0.05 results favored PGHD</th>
<th>No results were statistically significant</th>
<th>Mixed results,* but most p&lt;0.05 results against PGHD</th>
<th>All results p&lt;0.05 against PGHD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Health outcomes not reported</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Statement about health outcomes based only on health outcomes data: **Unclear effect**

<table>
<thead>
<tr>
<th>Surrogate Outcomes</th>
<th>Metabolic syndrome score</th>
<th>Surrogate outcomes not reported</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Statement about health outcomes based on both health outcomes and surrogate outcomes: **Unclear effect**

Note: This table only displays studies that used an *isolated-effect design* to compare the *presence vs absence* of a PGHD device.

- Low risk of bias study of an isolated effect; Moderate risk of bias study of an isolated effect; High risk of bias study of an isolated effect

Study did not report this category of outcome (the number of question marks indicates the number of studies)

* "Mixed results" means that a study investigated either multiple time points (e.g., both 13 weeks and 26 weeks) or multiple comparisons (e.g., PGHD device A vs no device, and PGHD device B vs no device), and data showed statistical significance for some but not all time points/comparisons.

PGHD – Patient generated health data
Multicomponent Effects on Health Outcomes
The U.K. study by Biddle et al. (2015) (moderate risk of bias) reported that QOL (as measured by the EuroQol 5D) was not statistically different between their groups (see Appendix Table C-14).

Multicomponent Effects on Surrogate Outcomes
Due to this evidence’s highly indirect nature, we do not discuss it; instead, we refer the reader to the tabulated data in Appendix Table C-14.

Of the four diabetes-prevention-related records in clinicaltrials.gov, none made PGHD-related comparisons and stated that they were collecting data on a health outcomes.

Guiding Question 4. What are the harms or AEs associated with these specific consumer automated-entry PGHD technologies? Which patients in which settings are most at risk of harms?
The U.S. trial and U.K. trial did not report these data. The Canadian study reported four AEs among three control group patients and none in the intervention group. The intervention group, also reported device-initial alarms for high SBP (no alarms during the 1-year followup), high DBP (7 alarms), and high glucose (12 alarms, but 11 were from a single patient). See more details in Appendix Table C-14.

Guiding Question 5. For the technologies demonstrating associations with outcomes of interest, what full economic evaluations provide information on the relative value for consumers?
None of the three diabetes prevention trials reported cost-effectiveness data or cost-only data.

Sleep Apnea

Guiding Question 1. Which specific consumer automated-entry PGHD applications/technologies/devices have been studied for measurement of health outcomes? What studies are in progress of consumer PGHD devices?
   a. What study designs have been used?
   b. What were the inclusion/exclusion criteria?
   c. What statistical analysis and data were used to determine study size and power?
   d. How long were patients followed?
   e. How was adherence measured?
   f. What was the comparator?
   g. Which outcomes were measured?
Devices
The three sleep apnea trials used three different PGHD devices. The similarity judgments (how similar each device is to those currently on the market by this manufacturer) were:

- Two pedometers: Misfit Shine (judged as somewhat different) and the Samsung Charm (judged as similar)
- One BP monitor: Omron 705CP (judged as similar)

All three devices (along with devices included in trials of other clinical conditions) appear in Appendix Table C-1.

Studies in progress
We identified six potentially relevant records in ClinicalTrials.gov. More details about these records (including hyperlinks) appear in Appendix Table C-15.

Study designs
All three included sleep apnea trials were randomized. Kim et al. (2019)\textsuperscript{46} was conducted in South Korea (43 patients enrolled, 4-week followup, median patient enrollment month July 2017), Mendelson et al. (2014)\textsuperscript{32} was conducted in France (107 patients enrolled, 17-week followup, median patient enrollment month October 2010), and Cho et al. (2018)\textsuperscript{47} was conducted in South Korea (47 patients enrolled, 4-week followup, median patient enrollment month September 2016).

The next sections describe additional aspects of these trials, which are tabled in detail in Appendix Table C-16, C-17, and C-18.

Inclusion/exclusion criteria
Kim et al. (2019)\textsuperscript{46} included obese/overweight adults with sleep apnea who could use a mobile app and a wearable device. Mendelson et al. (2014)\textsuperscript{32} included adults with sleep apnea and BMI <40 kg/m\textsuperscript{2} with a high cardiovascular risk score (cardiovascular risk SCORE > 5%) or history of cardiovascular disease. Cho et al. (2018)\textsuperscript{47} included obese/overweight adults with habitual snoring or witnessed sleep apnea but not using a continuous positive airway pressure or similar device.

Statistical power analyses
Only Mendelson et al. (2014)\textsuperscript{32} mentioned power analyses; these had used prior data and had also accounted for possible attrition.

Follow-up length
The two trials from South Korea each had a 4-week followup, whereas Mendelson et al. (2014)\textsuperscript{32} had a 17-week followup.

Adherence measurement
Only Kim et al. (2019)\textsuperscript{46} reported device adherence, which was measured by the percentage of days on which patient data were collected regarding physical activity, sleep parameters, and weight.
Comparators

Kim et al. (2019)\textsuperscript{46} compared three groups: no device, App only (MyHealthKeeper) which used the phone’s accelerometer, and App+ pedometer (Samsung Charm). Mendelson et al. (2014)\textsuperscript{32} had two groups: Standard (SenseWear Pro2 armband) and Telemedicine (BP monitor Omron 705CP plus SenseWear Pro2 armband). Cho et al. (2018)\textsuperscript{47} had two groups: Control (no device) and App+pedometer (Misfit Shine).

Outcomes reported

Only Mendelson et al. (2014)\textsuperscript{32} reported health outcomes (QOL as measured by the “MCS” and “PCS,” which the authors did not define). All three trials reported a surrogate outcome (number of days/week apnea was witnessed, or the Epworth sleepiness scale, or the apnea-hypopnea index, lowest oxygen saturation, or the oxygen desaturation index). Only Kim et al. (2019)\textsuperscript{46} reported data for Guiding Question 2.

Guiding Question 2. What are the characteristics (e.g., interoperability, functions, acceptability/usability, sustainability, feasibility, fidelity, or integration into electronic health records) of these specific consumer automated-entry PGHD technologies?

Only Kim et al. (2019)\textsuperscript{46} reported data pertinent to this Guiding Question. The two intervention groups (app only, or app+pedometer) each had data collection rates of 50% (of study days) for physical activity, 32% for sleep data, and 32%-33% for weight data. Patient satisfaction was measured as whether patients responded to daily clinician feedback on the app, with a maximum of 7 if the patient responded “satisfied” to each daily clinician comment over the 7 days. No response was scored as “not satisfied.” These scores ranged from 3.1 to 4.2 over the 4 weeks for the app-only group and 3.7 to 4.7 for the app+wearable group. No statistical comparison was made, but the latter group had higher means for each of the 4 weeks, suggesting the pedometer increased app engagement.

Guiding Question 3. What is the influence of specific consumer automated-entry PGHD technologies on health outcomes? Does this vary across different patient populations, different settings, or other modifiers of effectiveness?

The risk-of-bias assessments are in Appendix Table C-19, and the results are in Appendix Table C-20.

Isolated Effects on Health Outcomes

None of the three sleep apnea trials reported whether there were isolated effects on health outcomes. As a result, we rated this evidence as unclear.

Isolated Effects on Surrogate Outcomes

Kim et al. (2019)\textsuperscript{46} (high risk of bias) reported statistically nonsignificant differences between their three groups at 4 weeks with respect to the number of days/week when apnea was witnessed.
Overall, there is an *unclear effect of PGHD technologies on health outcomes* when used to prevent diabetes (Table 5). Only one of three trials used an isolated-effect design, it found no statistically significant difference between group in the number of days/week when apnea was witnessed, and it was at high risk of bias.
Table 5. Sleep Apnea: Categorization of Isolated Effects Due to Device Presence/Absence

<table>
<thead>
<tr>
<th>Health Outcomes</th>
<th>All results were p&lt;0.05 in favor of PGHD</th>
<th>Mixed results,* but most p&lt;0.05 results favored PGHD</th>
<th>No results were statistically significant</th>
<th>Mixed results,* but most p&lt;0.05 results against PGHD</th>
<th>All results p&lt;0.05 against PGHD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Health outcomes not reported</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>?</td>
</tr>
</tbody>
</table>

Statement about health outcomes based only on health outcomes data: Unclear effect

<table>
<thead>
<tr>
<th>Surrogate Outcomes</th>
<th>Number of days/week when apnea was witnessed</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Surrogate outcomes not reported</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Statement about health outcomes based on both health outcomes and surrogate outcomes: Unclear effect

Note: This table only displays studies that used an isolated-effect design to compare the presence vs. absence of a PGHD device.

- Low risk of bias study of an isolated effect
- Moderate risk of bias study of an isolated effect
- High risk of bias study of an isolated effect
- ? Study did not report this category of outcome (the number of question marks indicates the number of studies)

* “Mixed results” means that a study investigated either multiple time points (e.g., both 13 weeks and 26 weeks) or multiple comparisons (e.g., PGHD device A vs no device, and PGHD device B vs no device), and data showed statistical significance for some but not all time points/comparisons.

PGHD – Patient generated health data
Multicomponent Effects on Health Outcomes
Mendelson et al. (2014)\textsuperscript{32} reported statistically nonsignificant differences between their standard group and their telemedicine group in QOL at 17 weeks.

Multicomponent Effects on Surrogate Outcomes
Due to this evidence’s highly indirect nature, we do not discuss it; instead, we here refer the reader to the tabulated data in Appendix Table C-20.

Of the six sleep-apnea-related records in clinicaltrials.gov, two made PGHD-related comparisons and stated that they were collecting data on a health outcomes (sleep apnea symptoms).

Guiding Question 4. What are the harms or AEs associated with these specific consumer automated-entry PGHD technologies? Which patients in which settings are most at risk of harms?
None of the three sleep apnea trials reported data harms or AEs.

Guiding Question 5. For the technologies demonstrating associations with outcomes of interest, what full economic evaluations provide information on the relative value for consumers?
None of the three sleep apnea trials reported cost-effectiveness data or cost-only data.

Hypertension

Guiding Question 1. Which specific consumer automated-entry PGHD applications/technologies/devices have been studied for measurement of health outcomes? What studies are in progress of consumer PGHD devices?
\begin{itemize}
  \item a. What study designs have been used?
  \item b. What were the inclusion/exclusion criteria?
  \item c. What statistical analysis and data were used to determine study size and power?
  \item d. How long were patients followed?
  \item e. How was adherence measured?
  \item f. What was the comparator?
  \item g. Which outcomes were measured?
\end{itemize}

Devices
The 50 hypertension trials used 53 different PGHD devices: 41 BP monitors, 4 pedometers, 3 glucose meters, 3 scales, and 2 heart rate monitors. All these devices (along with devices included in trials of other clinical conditions) appear in Appendix Table C-1. The similarity
judgments (how similar each device is to those currently on the market by this manufacturer) were:

- BP monitors: 38 similar, 1 somewhat different, 0 very different, 2 unknown
- Pedometers: 3 similar, 1 somewhat different, 0 very different, 0 unknown
- Glucose meters: 2 similar, 0 somewhat different, 0 very different, 1 unknown
- Scales: 1 similar, 0 somewhat different, 0 very different, 2 unknown
- Heart rate monitors: 1 similar, 0 somewhat different, 1 very different, 0 unknown

**Studies in progress**

We identified 16 records in ClinicalTrials.gov that potentially involved PGHD interventions to prevent or treat hypertension. As of April 7, 2020, seven were active but not recruiting, seven were recruiting, and two were not yet recruiting. More details about these records (including hyperlinks) appear in Appendix Table C-21.

**Study designs**

Forty-eight of the 50 hypertension studies were randomized trials, and two were nonrandomized comparative studies. Twenty-two studies were conducted in the United States, and other notable countries included the United Kingdom (6 studies), Canada (3 studies), Denmark (2 studies), Finland (2 studies), and South Korea (2 studies). The median number of patients per study at baseline was 223 (IQR 101 to 416). Patient enrollment dates were reported in 28 studies and ranged from May 1999 to June 2017. The next sections describe additional aspects of these trials, which are tabled in detail in Appendix Table C-22, C-23, and C-24.

**Inclusion/exclusion criteria**

Studies generally included adult patients with a diagnosis of primary hypertension or uncontrolled BP with antihypertensive medication. Patients with comorbidities, such as diabetes or chronic kidney disease, were often included. Common exclusion criteria included diagnosis of secondary hypertension, severe cognitive impairment, and pregnancy. Multiple studies also considered the comfort with technology as inclusion or exclusion criteria, such as requiring patients to own a smartphone or excluding patients who were unable to use the telehealth device.

**Statistical power analyses**

Forty of the 50 studies reported information about power analyses. Of these, 20 accounted for anticipated attrition and 5 reported using data from previous work to estimate the potential effect size. One study explicitly stated the analysis was retrospective, and two studies stated that no power analyses were performed.

**Follow-up length**

The median length of follow-up in the 50 hypertension studies was 6 months, with an IQR from 4 to 12 months.

**Adherence measurement**

Device adherence was typically measured by the number of BP readings as determined from automatic transmission from the devices or as recorded by patients.
Comparators

Forty-three of 50 studies had 2 treatment groups, 4 studies had 3 groups, and 3 studies had 4 groups. Forty-three studies included a usual care control group. Three of the studies used a PGHD device in the control group. The remaining four studies used active comparators without a PGHD device, such as diet alone, counseling alone, or exercise program alone.

Outcomes reported

Only 21 of the 50 hypertension studies reported on health outcomes, which were generally QOL outcomes or AEs. All studies reported surrogate outcomes, including SBP, DBP, or BP control. Twenty-seven reported on device usage, ease of use, or technical problems. Of studies including health outcomes, 12 reported on process outcomes (e.g., number of primary care consultations or antihypertensive medication changes) and 3 reported on economic outcomes. Seven reported on costs.

Guiding Question 2. What are the characteristics (e.g., interoperability, functions, acceptability/usability, sustainability, feasibility, fidelity, or integration into electronic health records) of these specific consumer automated-entry PGHD technologies?

Below, we discuss the characteristics of PGHD devices for hypertension. More details are provided in Appendix Table C-24. We also discuss the findings by device engineers on wearable and handheld BP monitors.

Twenty-six of the 27 studies reported device characteristic outcomes for BP monitors. Adherence to use of BP monitors ranged from 38% to 89% of patients. Some studies reported that adherence declined over time. For example, Bosworth et al. (2009)\(^{48}\) reported that during the first 2 months, 91% of those using the PGHD device were adherent, while 64% were adherent during the last 2 months. Studies also measured BP monitor use in various ways, including total number of transmissions over the study and average number of transmission per week. Studies measuring ease of use or satisfaction of BP devices found that patients tended to view them favorably. For example, Magid et al. (2013)\(^{49}\) reported that 68% of patients using the monitor found it very or extremely easy to use, and Rifkin et al. (2013)\(^{50}\) reported that 96% of patients reported they would continue to use the BP device. Two of the studies reported problems with the devices. Bosworth et al. (2011)\(^{51-53}\) found that 35 alerts were triggered by the monitoring system due to device problems, which represented 5% of the total alerts that occurred during the study. Lakshminarayan et al. (2018)\(^{54}\) found that some patients experienced issues with the BP device and smartphone provided to transmit BP data, including an inability to hold charge and difficulty using the phone app to see BP data.

Two studies reported use of glucometers. Earle et al. (2010)\(^{55,56}\) found that 4,099 readings were taken over the course of the study, with an average transmission rate of 1.8 readings per person per month. Yoo et al. (2009)\(^{7}\) found that patients transmitted blood glucose recordings an average of 1.84 times per day (adherence rate 92.2%).

Two studies examined use of weight scales. Bennett et al. (2018)\(^{57-59}\) reported that 36% of patients weighed themselves at least 5 days a week on average. Yoo et al. (2009)\(^{7}\) found that patients transmitted body weight measurements 0.87 times per day (adherence rate 87.4%).

Technical report on wearable and handheld BP devices
Device engineers performed an independent evaluation of one handheld and two wearable BP devices. The BodiMetrics Performance Monitor is a mobile handheld device that consumers can use to measure their SBP without applying a cuff. This consumer-marketed device has not undergone FDA review for its BP-measuring technology, and it does not have the same accuracy as BP monitors regulated by FDA. Device engineers performed a range of physical tests, reviewed product literature/specifications, and asked users about their experience with the device. Device engineers rated device performance as poor mainly because it does not measure DBP (a major disadvantage), and assessing BP as a cardiovascular risk factor requires measurement of SBP and DBP. Also, SBP measurements were less accurate than required for clinical applications and did not meet guidelines recommended by the American Medical Association or by the U.S. Centers for Disease Control and Prevention for selecting a device for home monitoring. Workflow was rated as excellent due to retention of BP trend information, data storage when device is not linked with the smartphone app, and forwarding data to the physician or caregiver (all considered major advantages). Device engineers rated interoperability and cybersecurity as good. Maintenance was rated as excellent due to the major advantages of relatively infrequent intervals between battery charging (every 3 to 5 weeks) and relatively short battery recharging time (1 hour). Despite the device being intuitive to use, user experience was rated as poor due to the major disadvantage of lack of adequate user support (calls and e-mails not returned). Safety was rated good, with the minor advantage that the smartphone app plots graphs that allow users to see whether they have elevated SBP. Cost of ownership was estimated as $300 over a 3-year period. Overall, device engineers rated the device as poor due to its inability to measure DBP and the lack of user support.

The Everlast TR10 Heart Rate Activity Tracker is one of two wrist-wearable BP devices that device engineers evaluated. Consumers can use the device to measure their SBP and DBP without applying a cuff. This consumer-marketed device has not undergone FDA review for its BP-measuring technology, and it does not have the same accuracy as BP monitors regulated by FDA. Device engineers rated device performance as fair. Minor disadvantages of the TR10 include a lack of a recommended interval for calibration, no recommendation on frequency of calibration in the user manual, and both SBP and DBP measurements were less accurate than required for clinical applications. Workflow was rated as poor due to major disadvantages that include inability to store data when the wearable device is offline and disconnected from the smartphone app and inability to forward BP data to a physician or caregiver digitally. A minor advantage was that the smartphone app allows users to set a generic alarm to remind them when to measure their BP. Device engineers rated interoperability and cybersecurity as good. Maintenance was rated as excellent due to the major advantages of relatively infrequent intervals between battery charging (once per week) and relatively short battery recharging time (1 to 2 hours). Despite the major advantage of being comfortable to wear, user experience was rated as poor due to the major disadvantage of lack of adequate user support (lack of device information on the manufacturer website, calls and e-mails not returned) and a weak wristband securement system (e.g. if the wristband catches on something it can easily fall off). Safety was rated good, with the minor advantage that the smartphone app provides notification that allows users to see whether they have high BP. Cost of ownership was estimated as $70 over a 3-year period. Overall, device engineers rated the device as poor due to the inability to store BP measurements in the absence of Internet connectivity between the smartphone app and the TR10 and the lack of user support.
The Heartisans Blood Pressure Watch is the second wrist-wearable device that device engineers evaluated. Consumers can use it to measure their SBP and DBP without applying a cuff. This consumer-marketed device has not undergone FDA review for its BP-measuring technology, and it shows significantly less accuracy than BP monitors regulated by FDA. Device engineers rated device performance as unacceptable because during testing they found that, for both SBP and DBP, the mean absolute difference between measurements taken with the device and a conventional, hospital-grade cuff-type BP monitor exceeded 10 mm Hg. The mean differences far exceed the widely accepted AAMI SP10 criteria for a medical BP monitor, which require a mean difference no greater than 5 mm Hg with a standard deviation of \( \leq 8 \) mm Hg. This could lead to missed instances of elevated BP or false readings of elevated BP. Workflow was rated as excellent because the smartphone app for the Heartisans Watch retains multiple weeks of daily measurements that are both time- and date-stamped, allowing the user to review measurements. Also, data can be stored even when the device is not connected to the smartphone app, and data can be forwarded to a clinician or caregiver as a PDF attachment in an e-mail. Device engineers rated interoperability and cybersecurity as good. Maintenance was rated as fair despite a short battery charging time (1.5 hours) due to the major disadvantage of short battery life (only 8 or 9 hours during testing). Despite the major advantage of being a water-resistant device (can be completely submerged for 30 minutes), user experience was rated as poor due to the major disadvantage of lack of user support (no response to e-mails and no phone number listed on the product website). Safety was rated good, with the minor advantage that the smartphone app provides notification that allows users to see whether they have high BP. Cost of ownership was estimated as $149 over a 3-year period. Overall, device engineers rated the device as unacceptable due to its very low accuracy compared to that of FDA-approved hospital-grade, cuff-type BP monitors. They recommend against using the device to track BP.

Guiding Question 3. What is the influence of specific consumer automated-entry PGHD technologies on health outcomes? Does this vary across different patient populations, different settings, or other modifiers of effectiveness?

Below, we discuss the efficacy results in four categories: 1) isolated effects on health outcomes, 2) isolated effects on surrogate outcomes, 3) multicomponent effects on health outcomes, and 4) multicomponent effects on surrogate outcomes. Risk-of-bias assessments are in Appendix Table C-25, and the results are in Appendix Table C-27.

Isolated Effects on Health Outcomes

Six studies evaluated the isolated effects of a PGHD intervention on health outcomes. The PGHD devices that these studies examined were the following BP monitors: iHealth BP7 Wireless Wrist Monitor, Omron 637, Omron 773AC, Omron HEM-705 CP, Omron HEM-712C, and Omron M10-IT.

Bosworth et al. (2009)\(^{48}\) (moderate risk of bias) compared the Omron 773AC or 637 (depending on patient arm circumference) to usual care in terms of hospitalizations. A third group received behavioral management alone, and a fourth received a combination of PGHD and behavioral management. The proportion of hospitalized patients did not differ across groups (range 19.5% to 22.6%, \( p=0.91 \)).
Broege et al. (2001)\textsuperscript{60} (high risk of bias) evaluated the impact of using Omron HEM-702 compared to a clinic-only group on Short Form (SF)-36 QOL. After 3 months, overall QOL scores decreased in both groups but were not significantly different between groups (PGHD -6; clinic -4).

Green et al. (2008)\textsuperscript{61,62} (low risk of bias) evaluated the use of Omron HEM-705 CP compared to usual care for SF-12 QOL. A third group received a combination of PGHD and pharmacist care. At 12-month followup, there were no significant differences between PGHD alone and usual care groups for the SF-12 general health (PGHD 66.6, SD 20.9; usual care 67.1, SD 20.4; between group difference -0.1, 95\% confidence interval [CI]: -4.0 to 3.7), physical health (PGHD 77.7, SD 30.3; usual care 78.1, SD]27.7; between group difference -0.4, 95\% CI: -5.6 to 4.7), or emotional health (PGHD 72.1, SD 16.8; usual care 71.5, SD 17.7; between-group difference 0.5, 95\% CI: -2.7 to 3.8) subscales.

Hebert et al. (2012)\textsuperscript{63} (low risk of bias) evaluated use of the Omron HEM-712C BP monitor compared to usual care and reported on mortality. A third group received a combination of PGHD and nurse management. Over 18 months, 8 deaths occurred in the study. There was no significant difference across the groups (p=0.453).

McManus et al. (2018)\textsuperscript{64-67} (low risk bias) compared the impact of using the Omron M10-IT BP monitor to usual care on the EuroQol 5-dimension (EQ-5D)-5L score. A third group received a combination of the PGHD device and telemonitoring. At 12 months, there was no significant difference between the PGHD alone and usual care groups (mean difference [MD] -0.01; 95\% CI: -0.04 to 0.02; p-value: 0.4862).

Zha et al. (2019)\textsuperscript{68} (moderate risk of bias) compared the iHealth BP7 Wireless Wrist Monitor to usual care for the SF-36 QOL score. At 6 months, there was a significant difference between the groups for the mental health subscale, favoring usual care (PGHD 54.49, SD 20.11; usual care 80.73, SD 18.73; p-value: 0.003); however, it should be noted that at baseline, the usual care group also had significantly higher scores than the PGHD group. There were no significant differences between the groups for the physical function (PGHD 57.21, SD 23.97; usual care 56.82, SD 35.81; p-value: 0.51), role function (PGHD 60.00, SD 20.11; usual care 67.54, SD 29.3; p-value: 0.53), bodily pain (PGHD 43.69, SD 25.31; usual care 55.46, SD 33.84; p-value: 0.84), general health (PGHD 42.38, SD 21.22; usual care 57.71, SD 20.49; p-value: 0.08), vitality (PGHD 45.71, SD 17.43; usual care 50.46, SD 15.95; p-value: 0.08), social foundation (PGHD 50.29, SD 31.78; usual care 60.50, SD 36.71; p-value: 0.31), and role emotional (PGHD 48.62, SD 25.64; usual care 67.99, SD 27.35; p-value: 0.07) subscales, as well as for the physical (PGHD 42.63, SD 11.21; usual care 44.25, SD 11.17; p-value: 0.72) and mental (PGHD 43.11, SD 10.23; usual care 49.65, SD 11.87; p-value: 0.18) component summary scores.

Overall, based on the health outcomes alone, PGHD interventions have unclear effects on health outcomes. Only a small number of the studies reported on health outcomes, and those that did were generally reported being powered based on surrogate outcomes. See Table 6 for a summary of the study findings.

### Isolated Effects on Surrogate Outcomes

Eighteen studies\textsuperscript{48,54,60-81} examined the isolated effects of a PGHD intervention on surrogate outcomes. All evaluated the effect of device presence except for one study,\textsuperscript{76} which compared two BP monitors.

All 18 studies reported on the effects of PGHD interventions on SBP. Seven\textsuperscript{61,62,64-67,70-72,74,76,77,80} found that those receiving a PGHD device had significantly improved SBP compared
to those in the control group, though in some cases, results for only specific subgroups or time points were significant. For example, Bosworth et al. (2011)\textsuperscript{70-72} found significant improvement only for nonwhite patients at 12 months; differences were not significant for white patients at any timepoint or at 24 months for any subgroup. Based on a minimal important difference of 2 mmHg,\textsuperscript{82} 18/31 point estimates from the 17 device presence studies suggest PGHD interventions may have meaningful effects on SBP reduction (see Figure 4).

All but one of the studies\textsuperscript{54} reported on PGHD’s impact on DBP. The overall findings for DBP were similar to those for SBP. Five\textsuperscript{64-67,70,74,79,80} studies found that PGHD interventions significantly reduced DBP compared to control, but like for SBP, in some cases significance was seen only for particular subgroups or time points in a study. Based on a minimal important difference of 2 mmHg,\textsuperscript{82} 11/30 point estimates from 16 device presence studies suggest PGHD interventions may have meaningful effects on DBP reduction (see Figure 5).

Eleven studies\textsuperscript{48,61-63,69,73-75,78-81,83} examined BP control. Studies tended to define BP control as $<140/90$ mmHg. Two of the studies\textsuperscript{74,80} reported significant improvements in BP control among those using the PGHD intervention. The remaining studies generally found nonsignificant improvements in BP control.

Overall, taking into account the data on surrogate outcomes, PGHD interventions have a possible positive effect on health outcomes. See Table 6 for a summary of findings (both health outcomes and surrogate outcomes) from the device-presence studies reporting isolated effects.
Figure 4. SBP differences in studies of isolated effects of device presence/absence

<table>
<thead>
<tr>
<th>Study Details</th>
<th>Difference in SBP [mmHg]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marquez-Contreras 2006; Omron M4 automatic monitor vs. Usual care; 3 months</td>
<td>-2 mmHg</td>
</tr>
<tr>
<td>***P &lt; 0.001; Omron HEM-7121 vs. Control Group; 6 months</td>
<td></td>
</tr>
<tr>
<td>***Kaurik-Kelle 2007; Omron IC BP monitor vs. Usual care; 3 months</td>
<td></td>
</tr>
<tr>
<td>Kims 2006; Withings BP monitor vs. Usual care; 6 months</td>
<td></td>
</tr>
<tr>
<td>**Boosworth 2011; Omron 733MC or 637 vs. Usual care; 12 months; Non-white patients</td>
<td>-2 mmHg</td>
</tr>
<tr>
<td>**Suda 2012; Omron HEM-705CP vs. Usual care; 60 days</td>
<td></td>
</tr>
<tr>
<td>***Zalebski 2005; Omron 705 CPN vs. Usual care; 12 weeks</td>
<td></td>
</tr>
<tr>
<td>Bove 2013; MicroElec 3AC1-4P, Digi-Walker SW-206; Taylor Digital LCD Scale vs. Usual care; 6 months</td>
<td></td>
</tr>
<tr>
<td>Q: 2017; Omron HEM-7121 vs. Control Group; 2 years</td>
<td></td>
</tr>
<tr>
<td>Lakshminarayan 2018; Withings BP monitor vs. Usual care; 13 weeks</td>
<td></td>
</tr>
<tr>
<td>***Boosworth 2009; Omron 773MC or 637 vs. Usual care; 12 months</td>
<td></td>
</tr>
<tr>
<td>***McManus 2018; Omron M-7121 vs. Usual care; 12 months</td>
<td></td>
</tr>
<tr>
<td>***Zhu 2019; Health BP7 Wireless Blood Pressure Wrist Monitor vs. Usual care; 6 months</td>
<td></td>
</tr>
<tr>
<td>Aekplakone 2016; Omron HEM-7117 vs. Usual care; 6 months</td>
<td></td>
</tr>
<tr>
<td>***Grahn 2008; Omron HEM-705 CP vs. Usual care; 12 months</td>
<td></td>
</tr>
<tr>
<td>Aekplakone 2006; Omron HEM-7117 vs. Usual care; 12 months</td>
<td></td>
</tr>
<tr>
<td>McManus 2018; Omron M30-F vs. Usual care; 6 months</td>
<td></td>
</tr>
<tr>
<td>***Boosworth 2011; Health BP7 Wireless Blood Pressure Wrist Monitor vs. Usual care; 3 months</td>
<td></td>
</tr>
<tr>
<td>Boosworth 2013; Omron 773MC or 637 vs. Usual care; 12 months; White patients</td>
<td></td>
</tr>
<tr>
<td>Hubert 2012; Omron HEM-7121C vs. Usual care; 9 months</td>
<td></td>
</tr>
<tr>
<td>Boosworth 2013; Omron 773MC or 637 vs. Usual care; 24 months; White patients</td>
<td></td>
</tr>
<tr>
<td>Boosworth 2009; Omron 773MC or 637 vs. Usual care; 24 months</td>
<td></td>
</tr>
<tr>
<td>Midlman-Petersen 2017; ALO 7678 vs. Omron 705CP vs. Usual care; 3 months</td>
<td></td>
</tr>
<tr>
<td>Q: 2017; Omron HEM-7131 vs. Control Group; 1 year</td>
<td></td>
</tr>
<tr>
<td>Boosworth 2011; Omron 773MC or 637 vs. Usual care; 24 months; Non-white patients</td>
<td></td>
</tr>
<tr>
<td>Marquez-Contreras 2006; Omron M4 automatic monitor vs. Usual care; 3 months</td>
<td></td>
</tr>
<tr>
<td>MBiogas 2003; Omron TA-302 vs. Clinic group; 3 months</td>
<td></td>
</tr>
<tr>
<td>Q: 2017; Omron HEM-7121 vs. Control Group; 4 years</td>
<td></td>
</tr>
<tr>
<td>Marquez-Contreras 2006; Omron M4 automatic monitor vs. Usual care; 1 month</td>
<td></td>
</tr>
<tr>
<td>Hubert 2012; Omron HEM-7121C vs. Usual care; 18 months</td>
<td></td>
</tr>
<tr>
<td>Q: 2017; Omron HEM-7121 vs. Control Group; 3 years</td>
<td></td>
</tr>
</tbody>
</table>

Change from baseline prioritized over follow-up values if both reported; follow-up values used otherwise

***statistically significant between-group difference
† statistical significance not reported
PGHD – Patient generated health device
SBP – Systolic blood pressure
Figure 5. DBP differences in studies of isolated effects of device presence/absence

Change from baseline prioritized over follow-up values if both reported; follow-up values used otherwise

***statistically significant between-group difference
† statistical significance not reported
DBP – Diastolic blood pressure
PGHD – Patient generated health device
<table>
<thead>
<tr>
<th>Health Outcomes</th>
<th>Mixed results, * but most p&lt;0.05 results favored PGHD</th>
<th>No results were statistically significant †</th>
<th>Mixed results, * but most p&lt;0.05 results against PGHD</th>
<th>All results p&lt;0.05 against PGHD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality of life</td>
<td>● ○ ○ ○</td>
<td></td>
<td>● ○ ○ ○</td>
<td>● ○ ○ ○</td>
</tr>
<tr>
<td>Mortality</td>
<td>○</td>
<td></td>
<td>● ○ ○ ○</td>
<td>● ○ ○ ○</td>
</tr>
<tr>
<td>Hospitalizations</td>
<td>○</td>
<td></td>
<td>● ○ ○ ○</td>
<td>● ○ ○ ○</td>
</tr>
</tbody>
</table>

Statement about health outcomes based only on health outcomes data: **Unclear effect**

<table>
<thead>
<tr>
<th>Surrogate Outcomes</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>SBP</td>
<td>● ○ ○ ○</td>
</tr>
<tr>
<td></td>
<td>● ● ● ○ ○ ○</td>
</tr>
<tr>
<td></td>
<td>● ● ● ○ ○ ○</td>
</tr>
<tr>
<td></td>
<td>● ● ● ○ ○ ○</td>
</tr>
<tr>
<td></td>
<td>● ● ● ○ ○ ○</td>
</tr>
<tr>
<td></td>
<td>● ● ● ○ ○ ○</td>
</tr>
<tr>
<td></td>
<td>● ● ● ○ ○ ○</td>
</tr>
<tr>
<td></td>
<td>● ● ● ○ ○ ○</td>
</tr>
<tr>
<td></td>
<td>● ● ● ○ ○ ○</td>
</tr>
</tbody>
</table>

Statement about health outcomes based on both health outcomes and surrogate outcomes: **Possible positive effect**

Note: This table only displays studies that used an isolated-effect design to compare the presence vs absence of a PGHD device.

- ● Low risk of bias study of an isolated effect;
- ○ Moderate risk of bias study of an isolated effect;
- ○ High risk of bias study of an isolated effect;
- ? Study did not report this category of outcome (the number of question marks indicates the number of studies).
* “Mixed results” means that a study investigated either multiple time points (e.g., both 13 weeks and 26 weeks), multiple patient subgroups (without providing results for overall study population), or multiple outcomes within a category (e.g., multiple quality of life scales), and data showed statistical significance for some but not all time points/subgroups/outcomes.

† Two studies did not report if differences in surrogate outcomes were statistically significant; they have conservatively been grouped with the non-significant studies.

BP – Blood pressure; DBP – Diastolic blood pressure; PGHD – Patient generated health data; SBP – Systolic blood pressure
**Multicomponent Effects on Health Outcomes**

Sixteen studies\(^{32,45,48,49,61-67,84-106}\) examined the multicomponent effect of PGHD on health outcomes.

Eleven studies\(^{32,61,62,64-67,84-104,106}\) reported the effect of PGHD multicomponent with other interventions on QOL measures. One study\(^{84}\) found that multicomponent PGHD significantly improved QOL compared to usual care on the EQ-5D-3L scale, while another study\(^{64-67}\) reported that multicomponent PGHD significantly decreased QOL on the EQ-5D-5L scale. The remaining studies generally found modest nonsignificant differences between multicomponent PGHD and usual care, though the direction of the effects generally favored PGHD.

Three studies\(^{63,95-97,105}\) evaluated multicomponent PGHD’s impact on mortality. Mortality was low, and there were no significant differences between the multicomponent PGHD and control groups.

Three studies\(^{48,49,95-97}\) examined multicomponent PGHD’s effect on hospitalizations. None of these studies reported significant differences compared to control groups.

Two studies\(^{49,95-97}\) examined multicomponent PGHD’s effects on emergency room visits. Neither reported significant differences compared to control groups.

**Multicomponent Effects on Surrogate Outcomes**

Thirty-eight studies\(^{7,8,32,45,48-53,55-59,61-67,69-72,84-122}\) evaluated multicomponent PGHD for surrogate outcomes, including SBP, DBP, and BP control. See Appendix Table C-27 for further details.

Three of the 16 ongoing studies including hypertension patients identified on ClinicalTrials.gov reported that they were planning to examine health outcomes (quality of life).
Guiding Question 4. What are the harms or AEs associated with these specific consumer automated-entry PGHD technologies? Which patients in which settings are most at risk of harms?

Only 12 of the 50 hypertension studies\(^4^5,4^8,5^1-5^3,5^7-5^9,6^4-6^7,8^1,8^5-1^0^3,1^0^5\) reported data on AEs (see the Health Outcomes section of Appendix Table C-27). The reported types of events were:

- Overall number of events
- Individual serious AEs (including deaths, strokes, and myocardial infarction)
- Individual AEs (including pain, fatigue, and stiffness)
- Alerts for high or low BP (device alert triggered when SBP >220 or <60 or DBP >110 or <40)

Two of the 12 studies\(^6^4-6^7,9^8-1^0^2\) reported statistically significant differences between PGHD and control groups in terms of AEs. McManus et al. (2010)\(^9^8-1^0^2\) found that swelling of legs was significantly higher for those in the PGHD group (combination of Omron 705IT BP monitor and medication management) than for those in the usual care group, although for other AEs reported, including stiff joints, pain, and fatigue, there were no significant differences between the groups. McManus et al. (2018)\(^6^4-6^7\) found that occurrence of dry mouth was significantly higher in the PGHD group (Omron M10-IT) than in the usual care group, although for the other reported AEs, including pain, stiff joints, and sleep difficulties, there were no significant differences. In the remaining studies, AEs were not significantly or substantially different between the PGHD and control groups.

Guiding Question 5. For the technologies demonstrating associations with outcomes of interest, what full economic evaluations provide information on the relative value for consumers?

Four of the 50 hypertension studies\(^6^4-6^7,8^8-1^0^2\) reported PGHD’s impact on economic outcomes. Risk-of-bias evaluations are in Appendix Table C-26, and the results are in Appendix Table C-27.

Margolis et al. (2013)\(^8^8-9^4\) examined PGHD’s cost-effectiveness (A&D Medical 767PC BP monitor plus pharmacist management) compared to usual care. They estimated that use of the PGHD intervention would increase cost by $139 (95% CI: $46 to $347) per mmHg of reduced SBP, by $265 (95% CI: $83 to $743) per mmHg of reduced DBP, and by $7,337 (95% CI: $2278 to $26 329) per person achieving BP control (defined as <140/90 mmHg).

McKinstry et al. (2013)\(^9^5-9^7\) evaluated PGHD (Stabil-O-graph mobile BP monitor) compared to usual care and found that the costs per patient were higher and mean ambulatory SBP per patient was lower in the PGHD group, with an incremental cost-effectiveness ratio (ICER) of £25.60 per mmHg ambulatory SBP reduction (95% CI: £16.05 to £46.69).

McManus et al. (2010)\(^9^8-1^0^2\) evaluated PGHD (combination Omron 705IT BP plus medication management) compared to usual care. For men, using PGHD modeled over a lifetime (defined by the study as 35 years) results in a 0.24 increase in quality-adjusted life-years (QALYs) (PGHD 9.16 gained; usual care 8.92 gained), with an ICER of £1624. For women, using PGHD modeled over a lifetime (35 years) results in a 0.12 increase in QALYs (PGHD 10.57 gained; usual care 10.46 gained), with an ICER of £4923/QALY.
McManus et al. (2018)\textsuperscript{64-67} evaluated PGHD (Omron M10-IT BP monitor) compared to usual care and estimated that use of the PGHD device would result in 11.0447 QALYs compared to 11.0040 with usual care. The ICER was £3035/QALY.

**Coronary Artery Disease**

**Guiding Question 1.** Which specific consumer automated-entry PGHD applications/technologies/devices have been studied for measurement of health outcomes? What studies are in progress of consumer PGHD devices?

- a. What study designs have been used?
- b. What were the inclusion/exclusion criteria?
- c. What statistical analysis and data were used to determine study size and power?
- d. How long were patients followed?
- e. How was adherence measured?
- f. What was the comparator?
- g. Which outcomes were measured?

**Devices**

The six included coronary artery disease (CAD) trials used seven different PGHD devices: three heart rate monitors, three BP monitors, and one accelerometer. The similarity judgment (how similar each device is to those currently on the market by this manufacturer) was “similar” for all seven devices. All these devices (along with devices included in trials of other clinical conditions) appear in Appendix Table C-1.

**Studies in progress**

We identified 16 records in ClinicalTrials.gov that potentially involved PGHD interventions to prevent or treat CAD. As of April 7, 2020, four were not yet recruiting patients, eight were recruiting, and four were active but not recruiting. More details about these records (including hyperlinks) appear in Appendix Table C-28.

**Study designs**

All six included CAD studies were randomized trials. Two trials were conducted in Belgium, one in Canada, one in France, one in Spain, and one in The Netherlands. Patient enrollment dates (reported by 5 trials) ranged from December 2007 to August 2016, and the mean study enrollment period was 104 weeks. The mean number of patients per study at baseline was 108 (range 78 to 203). The next sections describe additional aspects of these trials, which are tabled in detail in Appendix Table C-29, C-30, and C-31.
Inclusion/exclusion criteria
Five studies included patients with documented CAD/acute coronary syndrome either with or without prior CAD events (myocardial infarction or revascularization procedures). One trial included patients with cardiovascular risk factors or history of CAD. Four trials required patients to have Internet access with or without a computer.

Statistical power analyses
Three studies conducted a priori power analyses (a fourth study may have conducted power analysis a priori, but this was unclear). The same three studies based effect-size estimates on prior work and accounted for anticipated attrition. One study stated the anticipated effect size but did not provide the basis for this effect size. The other two studies provided no information on power analysis.

Follow-up length
The mean followup length in the six CAD trials was 36.5 weeks (range 12 to 68 weeks).

Adherence measurement
None of the studies measured device adherence. One study measured adherence to training sessions.

Comparators
Three studies had control groups consisting of usual/standard care. Two studies had control groups that received center-based cardiac rehabilitation (CR). One study had a control group that used a modified motion sensor and did not receive physician feedback. The remaining study had a control group that received home CPAP.

Outcomes reported
Regarding health outcomes, four studies reported QOL, while mortality, rehospitalization, emergency room visits, and exercise capacity (function) was reported in one study each. One study reported economic outcomes (QALYs and societal costs). In addition, two studies reported acceptability.

Guiding Question 2. What are the characteristics (e.g., interoperability, functions, acceptability/usability, sustainability, feasibility, fidelity, or integration into electronic health records) of these specific consumer automated-entry PGHD technologies?
Below, we discuss the pertinent CAD data for this Guiding Question separately for heart rate monitors, BP monitors, and accelerometers. The risk-of-bias assessments are in Appendix Table C-32, and the results are in Appendix Table C-34. None of the studies reported on interoperability, functions, sustainability, or integration into electronic health records.

Heart rate monitors
Three trials included heart rate monitors as part of the intervention. Only one trial (Kraal et al.123) reported on acceptability and fidelity (adherence) to the overall home training program (which involved heart rate monitoring). Patients in the home-based group were more satisfied
with their cardiac rehabilitation program than patients in the center-based group (home-based: 8.7/10, center-based: 8.1/10, \( p = 0.02 \)).

**BP monitors**

Three trials included BP monitors as part of the intervention. Only one trial (Blasco et al.\textsuperscript{124}) reported on acceptability of the wireless application protocol (WAP) used to transmit data from BP monitors and other devices through their mobile phones. Almost all patients (98\%) completed more than 50\% of WAP sessions, and more than 83\% completed more than 75\% of them. Only 0.5 messages per patient were missed due to the mobile phone being turned off.

**Accelerometers/pedometers**

One study\textsuperscript{125} used an accelerometer but did not report any data related to Guiding Question 2.

**Guiding Question 3. What is the influence of specific consumer automated-entry PGHD technologies on health outcomes? Does this vary across different patient populations, different settings, or other modifiers of effectiveness?**

Below, we discuss the efficacy results in four categories: 1) isolated effects on health outcomes, 2) isolated effects on surrogate outcomes, 3) multicomponent effects on health outcomes, and 4) multicomponent effects on surrogate outcomes. For specific outcome data, see Appendix Table C-34.

**Health Outcomes**

Of the four studies that reported QOL outcomes, none found a statistically significant difference between the PGHD arm and the control arm (scores were almost identical in each group).\textsuperscript{32,123,124,126} The findings were similar in studies with isolated effects and studies with multicomponent effects.

The only study that reported mortality found a significant reduction in mortality in the PGHD arm compared to usual care (0 vs. 5 deaths, \( p = 0.029 \)).\textsuperscript{124} Because three PGHD devices were used in the PGHD arm and none in the control arm, the effects of the individual devices could not be isolated.

One study each\textsuperscript{125,127} reported hospitalization or emergency room visit outcomes, and neither found a statistically significant between-group difference in either outcome. Event rates were on average two-fold higher in the control arms, but the studies lacked adequate statistical power to detect a difference because the overall event rates were low. Only one of these studies\textsuperscript{125} isolated the PGHD intervention’s effect.

One study\textsuperscript{127} reported a significant increase from baseline in total time on exercise stress test (seconds) in the PGHD arm compared to the control arm. However, the PGHD intervention’s effect was not isolated because the PGHD group also received scheduled chat sessions and education sessions with medical personnel that the usual care group did not receive.

Overall, the evidence suggests a possible positive effect of PGHD devices on health outcomes for patients with CAD. See Table 7 for a visual display of the findings of studies that isolated effects of PGHD devices.
**Surrogate Outcomes**

None of the studies reported surrogate outcomes.

Out of 16 ongoing trials of PGHD in patients with CAD identified in ClinicalTrials.gov, 7 planned to measure a health outcome or outcomes related to CAD.
## Table 7. Coronary Artery Disease: Categorization of Isolated Effects Due to Device Presence/Absence

<table>
<thead>
<tr>
<th>Health Outcomes</th>
<th>All results were p&lt;0.05 in favor of PGHD</th>
<th>Mixed results,* but most p&lt;0.05 results favored PGHD</th>
<th>No results were statistically significant</th>
<th>Mixed results,* but most p&lt;0.05 results against PGHD</th>
<th>All results p&lt;0.05 against PGHD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality of life</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Mortality</td>
<td>☐</td>
<td></td>
<td>☐</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hospitalization</td>
<td>☐</td>
<td></td>
<td>☐</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Health outcomes not reported</td>
<td>☐</td>
<td></td>
<td>☐</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Statement about health outcomes based only on health outcomes data: **Possible positive effect**

Note: This table only displays studies that used an *isolated-effect design* to compare the *presence vs absence* of a PGHD device.

- ☐ Low risk of bias study of an isolated effect; ☐ Moderate risk of bias study of an isolated effect; ☐ High risk of bias study of an isolated effect

? Study did not report this category of outcome (the number of question marks indicates the number of studies)

* "Mixed results" means that a study investigated either multiple time points (e.g., both 13 weeks and 26 weeks) or multiple comparisons (e.g., PGHD device A vs no device, and PGHD device B vs no device), and data showed statistical significance for some but not all time points/comparisons.

PGHD – Patient generated health data
Guiding Question 4. What are the harms or AEs associated with these specific consumer automated-entry PGHD technologies? Which patients in which settings are most at risk of harms?

None of the included studies identified or reported any AEs related to the PGHD technologies.

Guiding Question 5. For the technologies demonstrating associations with outcomes of interest, what full economic evaluations provide information on the relative value for consumers?

One study by Kraal et al.\textsuperscript{123} reported QALYs and societal costs associated with a home-based cardiac rehabilitation (CR) training program that used heart rate monitoring with data transmission for patient and physician review (see risk-of-bias evaluation in Appendix Table C-33). The QALYs calculated for the center-based group (0.78 +/- 0.08) were similar to the QALYs for the home-based CR group (0.77 +/- 0.13, p=0.73). From a societal perspective (i.e. the sum of healthcare and non-health-care costs), costs per patient were lower for patients in the home-based group, although the difference did not reach statistical significance (p=0.09).

Heart Failure

Guiding Question 1. Which specific consumer automated-entry PGHD applications/technologies/devices have been studied for measurement of health outcomes? What studies are in progress of consumer PGHD devices?

a. What study designs have been used?
b. What were the inclusion/exclusion criteria?
c. What statistical analysis and data were used to determine study size and power?
d. How long were patients followed?
e. How was adherence measured?
f. What was the comparator?
g. Which outcomes were measured?

Devices

The 6 heart failure (HF) trials used 15 different PGHD devices: 6 BP monitors, 5 scales, 3 heart rate monitors, and 1 pulse oximeter. All these devices (along with devices included in trials of other clinical conditions) appear in Appendix Table C-1. The similarity judgments (how similar each device is to those currently on the market by this manufacturer) were:

- BP monitors: 5 similar, 0 somewhat different, 0 very different, 1 unknown
• Scales: 3 similar, 1 somewhat different, 0 very different, 1 unknown
• Heart rate monitors: 2 similar, 0 somewhat different, 0 very different, 1 unknown
• Pulse oximeter: 0 similar, 0 somewhat different, 0 very different, 1 unknown

Studies in progress
We identified three records in ClinicalTrials.gov that potentially involved PGHD interventions to management of heart failure. As of April 7, 2020, two were not yet recruiting patients and one was recruiting. More details about these records (including hyperlinks) appear in Appendix Table C-41.

Study designs
All six included heart failure studies were randomized trials. Two trials were conducted in the United States and one each in Canada, Austria, Denmark, and Germany. Patient enrollment dates (reported by 5 trials) ranged from October 2003 to May 2017, and the mean study enrollment period was 123 weeks. The mean number of patients per study at baseline was 613 (range 100 to 1571). The next sections describe additional aspects of these trials, which are tabled in detail in Appendix Table C-36, C-37, and C-38.

Inclusion/exclusion criteria
All studies enrolled patients with heart failure, but each had slightly different criteria. One study included patients with heart failure rated as New York Heart Association (NYHA) class 2, 3 or 4. Another study included patients with heart failure rated as NYHA Class 2 or 3 and had a left ventricular ejection fraction of 45% or lower (or if more than 45%, were being treated with oral diuretics). Another study required current hospital admission or recent discharge (within prior 2 weeks) with a primary diagnosis of heart failure, considered high risk for readmission (history of hospital readmissions for cardiac-related reasons or ejection fraction ≤20%). Two studies enrolled patients with decompensated heart failure (defined as heart failure with the initiation of or an increase in diuretic treatment). The remaining study included ambulatory patients given a diagnosis of heart failure with a left ventricular ejection fraction less than 40%.

Statistical power analyses
Four studies conducted a priori power analyses. All four studies described the basis for their effect-size estimates, but only one cited a prior study as a basis for the estimate. Only one study mentioned accounting for anticipated attrition. The other two studies provided no information on power analysis.

Follow-up length
The mean followup length in the 6 trials was 34.7 weeks (range 26 to 52 weeks).

Adherence measurement
Two studies measured patient adherence based on daily transmission of data to the monitoring center.
Comparators

Five studies had control groups consisting of usual/standard care. The remaining study specified pharmacologic treatment (a component of usual care) as the comparator.

Outcomes reported

All studies reported health outcomes. The most commonly reported health outcomes were QOL (4 studies), hospitalization or readmission (3 studies), mortality (3 studies), and emergency room visits (2 studies). Four studies reported outcomes related to Guiding Question 2 (3 studies reported adherence, and 1 reported acceptability). One study reported a process outcome (physician adjustment of medication).

Guiding Question 2. What are the characteristics (e.g., interoperability, functions, acceptability/usability, sustainability, feasibility, fidelity, or integration into electronic health records) of these specific consumer automated-entry PGHD technologies?

Below, we discuss the pertinent HF data for this Guiding Question separately for heart rate monitors, BP monitors, and accelerometers. All these data appear in Appendix Table C-40. None of the studies reported on interoperability, functions, sustainability, or integration into electronic health records.

Heart rate monitors

All trials included heart rate monitors as part of the intervention. Three trials\textsuperscript{128-130} reported data on adherence, but this was not specific to heart rate monitors. The studies measured adherence to data transmission to the monitoring centers; these data included information from other devices (BP monitors, scales) in addition to heart rate monitors. Studies reported adherence differently, with one study reporting overall 95% adherence, another study reporting that 97% of patients were 70% adherent with daily data transmission, and another study reporting a range of adherence rates for patients within the study. One study\textsuperscript{131} reported acceptability; of 20/42 patients (48%) who filled out a questionnaire, 100% reported that the equipment was simple and easy to use and the program made them feel more in control of their health. The equipment included multiple devices, not just a heart rate monitor.

BP monitors

All trials included BP monitors as part of the intervention. See information under heart rate monitors above, as all trials included both devices and did not measure adherence or acceptability separately for each device.

Scales

All studies included scales. See information under heart rate monitors above, as all trials included both devices and did not measure adherence or acceptability separately for each device.

Pulse oximeters

One trial\textsuperscript{131} included a pulse oximeter as part of the intervention. Although this study reported data on acceptability (see under heart rate monitors above), it included all devices in the measurement of acceptability.
Guiding Question 3. What is the influence of specific consumer automated-entry PGHD technologies on health outcomes? Does this vary across different patient populations, different settings, or other modifiers of effectiveness?

Below, we discuss the efficacy results in four categories: 1) isolated effects on health outcomes, 2) isolated effects on surrogate outcomes, 3) multicomponent effects on health outcomes, and 4) multicomponent effects on surrogate outcomes. Risk-of-bias assessments are in Appendix Table C-39, and the results are in Appendix Table C-40.

Health Outcomes

There was some inconsistency in findings regarding QOL among HF studies. Two studies isolated the effect of the combined PGHD interventions but could not isolate the effect of the individual PGHD devices used as part of the intervention. These studies used different QOL instruments: one used the SF-36 mental (MCS) and physical component scores (PCS) and a disease-specific instrument. This study found no statistically significant between-group difference in the SF-36 PCS or the disease-specific instrument score, but did find a difference in the SF-36 mental component score favoring the PGHD intervention. The other study used the disease-specific Minnesota Living with Heart Failure Questionnaire (MLHFQ) and found a statistically significant between-group difference favoring the PGHD intervention.

Two studies that combined PGHD and other interventions used the MLHFQ and reported mixed findings, with one study reporting a significant between-group difference favoring the PGHD intervention while the other study found no significant between-group difference. In the latter study, the direction of effect still favored the PGHD arm.

Of three studies that reported mortality, one reported a statistically significant between-group difference in mortality (favoring lower mortality in the PGHD arm). This study combined multiple PGHD devices with another intervention, so the PGHD intervention’s effect could not be isolated. The direction of effect was not consistent among the three studies.

Three studies that reported hospitalizations or hospital readmissions did not find a statistically significant between-group difference and did not show consistency in the direction of effect. These studies all used combinations of PGHD devices and did not isolate the effect of individual devices.

Two studies reported emergency room visits and neither found a statistically significant between-group difference, with the direction of effect in both studies slightly favoring the usual care arm.

Overall, the evidence suggests a possible positive effect of combined PGHD devices in patients with heart failure, but because all studies combined multiple PGHD devices together, the possible effect of any individual PGHD device remains unclear. See Table 8 for a visual display of the findings of studies that isolated effects for PGHD interventions.

Surrogate Outcomes

No studies reported surrogate outcomes that met inclusion criteria.

All three ongoing trials identified in ClinicalTrials.gov addressing PGHD for heart failure planned to measure a health outcome or outcomes.
Table 8. Heart Failure: Categorization of Isolated Effects Due to Device Presence/Absence

<table>
<thead>
<tr>
<th>Health Outcomes</th>
<th>All results were p&lt;0.05 in favor of PGHD</th>
<th>Mixed results,* but most p&lt;0.05 results favored PGHD</th>
<th>No results were statistically significant</th>
<th>Mixed results,* but most p&lt;0.05 results against PGHD</th>
<th>All results p&lt;0.05 against PGHD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality of life</td>
<td>☀</td>
<td>☄</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mortality</td>
<td>☄</td>
<td>☀</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hospitalizations</td>
<td>☄</td>
<td>☄</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Emergency room visits</td>
<td>☄</td>
<td>☄</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Health outcomes not reported</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Statement about health outcomes based only on health outcomes data: **Possible positive effect**

Note: This table only displays studies that used an isolated-effect design to compare the presence vs absence of a PGHD device.

- ☀ Low risk of bias study of an isolated effect; ☄ Moderate risk of bias study of an isolated effect; ☉ High risk of bias study of an isolated effect

* “Mixed results” means that a study investigated either multiple time points (e.g., both 13 weeks and 26 weeks) or multiple comparisons (e.g., PGHD device A vs no device, and PGHD device B vs no device), and data showed statistical significance for some but not all time points/comparisons.

PGHD – Patient generated health data
Guiding Question 4. What are the harms or AEs associated with these specific consumer automated-entry PGHD technologies? Which patients in which settings are most at risk of harms?

None of the included studies identified or reported any AEs related to the PGHD technologies.

Guiding Question 5. For the technologies demonstrating associations with outcomes of interest, what full economic evaluations provide information on the relative value for consumers?

None of the included studies performed an economic evaluation.

Cardiac Arrhythmias

Guiding Question 1. Which specific consumer automated-entry PGHD applications/technologies/devices have been studied for measurement of health outcomes? What studies are in progress of consumer PGHD devices?

a. What study designs have been used?

b. What were the inclusion/exclusion criteria?

c. What statistical analysis and data were used to determine study size and power?

d. How long were patients followed?

e. How was adherence measured?

f. What was the comparator?

g. Which outcomes were measured?

Devices

The four included cardiac arrhythmia trials all used the same electrocardiogram (ECG) monitor: the AliveCor Kardia. We rated the model as similar to other model(s) currently available from the manufacturer. This device (along with devices included in trials of other clinical conditions) appears in Appendix Table C-1.

Studies in progress

We identified eight records in ClinicalTrials.gov that potentially involved PGHD interventions to prevent or treat cardiac arrhythmias. As of April 7, 2020, one was not yet
recruiting patients, five were recruiting, and two were active but not recruiting. More details about these records (including hyperlinks) appear in Appendix Table C-41.

**Study designs**

All four included studies were randomized trials. Two trials were conducted in the United States and two in the United Kingdom. Patient enrollment dates (reported by 2 trials) ranged from November 2013 to January 2018, and the mean study enrollment period was 78 weeks. The mean number of patients per study at baseline was 386 (range 58 to 1004). The next sections describe additional aspects of these trials, which are tabled in detail in Appendix Table C-42, C-43, and C-44.

**Inclusion/exclusion criteria**

One study included patients documented with at least one atrial fibrillation (AF) risk factor, another study included patients with paroxysmal AF, a third study included patients presenting with an episode of palpitations or pre-syncope with no obvious cause, and the remaining study included individuals >65 years of age with a CHADS-VASc score ≥2 not in receipt of oral anticoagulation therapy without a known diagnosis of AF currently, a known contraindication to anticoagulation, or permanent cardiac pacing implantation. Three studies required patients have the ability to use the AliveCor heart monitor and have a smartphone or Internet access.

**Statistical power analyses**

Three studies conducted *a priori* power analyses. One study had based effect-size estimates on prior work, and two studies had accounted for anticipated attrition. One study stated the anticipated effect size but did not provide the basis for this effect size. One study provided no information on power analysis.

**Follow-up length**

The mean followup length up in the 4 trials was 42 weeks (range 12 to 80 weeks).

**Adherence measurement**

One study measured device adherence by the number of times patients recorded a daily ECG. No other studies measured adherence.

**Comparators**

Three studies had control groups consisting of usual/standard care. The remaining study had a control group of patients who received continuous anticoagulation.

**Outcomes reported**

All studies reported health outcomes, with mortality the most commonly reported outcome (3 studies), followed by stroke, major/clinically significant bleeding, and emergency room visits (2 studies each). Other health outcomes included hospitalization, deep vein thrombosis/pulmonary embolism, and major adverse cardiac events (1 study each). Three studies reported outcomes related to Guiding Question 2 (adherence/fidelity to protocol in 2 studies, acceptability in 1 study). Two studies reported a process outcome (medication initiation or medication change).
Guiding Question 2. What are the characteristics (e.g., interoperability, functions, acceptability/usability, sustainability, feasibility, fidelity, or integration into electronic health records) of these specific consumer automated-entry PGHD technologies?

Below, we discuss the pertinent data for this Guiding Question separately for the AliveCor ECG monitor. The data appear in Appendix Table C-46. None of the studies reported on interoperability, functions, sustainability, or integration into electronic health records.

Heart rate monitors

All trials included the AliveCor ECG monitors as part of the intervention. Two trials reported adherence/fidelity to protocol. Goldenthal et al. reported that 41 patients (36%) recorded greater than 180 times, on average once per day, and 77 (67%) used the device in the last month of their study period. Ninety-three (81%) averaged transmission at least once per week, and 86 (75%) used the device in the second half of the study. Stavrakis et al. reported that 4/29 patients (14%) in the intermittent anticoagulation arm crossed over to the continuous anticoagulation arm due to failure to submit rhythm strips. The 29 patients in the intermittent arm had a median of 3 failed submissions of rhythm strips (IQR 0 to 5). Reed et al. was the only study that reported acceptability; 80/92 (87.0%) patients considered the AliveCor monitor easy to use.

Device engineers independently evaluated the AliveCor KardiaMobile and KardiaBand Smartphone-enabled ECG monitors (KardiaBand is a watchband for the Apple Watch). Both use the Kardia app, which can be used on a smartphone or Apple Watch. Device engineers performed a range of physical tests, reviewed product literature/specifications, and asked users about their experience with the device. They rated device performance as excellent because the detection algorithm works within the Kardia app even when the smartphone is not connected to the Internet or (for KardiaBand) when the paired iPhone is out of range (considered a major advantage). It also has a guest mode that allows users to share the device with others while keeping the primary user’s own data separate from other users’ data. It also includes orientation correction in case a user holds the device in the wrong orientation. Workflow was rated as good, as the Kardia app has a voice notes feature that allows users to record a message while recording their ECG; messages are automatically transcribed. Interoperability was rated as good because the Kardia app generates PDF reports, which can be imported into a patient’s electronic health record (EHR), allowing clinicians to track the patient’s condition within their EHR system. Device engineers rated maintenance as good, with no significant issues identified. They rated the user experience as good based on human factors testing; volunteers generally liked the Kardia system and found it easy to use despite having no previous experience with it. Device engineers further noted that the long battery life (1 year for KardiaMobile and 2 years for KardiaBand) is a major advantage in providing a good patient experience. A major disadvantage was that neither device provides a battery-replacement warning, so users may be confused if and when the device stops working. Safety was rated as good, as no serious safety concerns were identified. The estimated cost of ownership for a 2-year period was $310 for KardiaMobile and $400 for KardiaBand. The overall rating for both of the AliveCor devices across all categories was good.

Device engineers also evaluated another smartphone-enabled ECG monitor (Cardiac Designs ECG Check). Similar to the AliveCor devices, the ECG Check was rated as good in the categories of interoperability, maintenance, patient/user experience, and safety. However, it was
rated as inferior to the AliveCor devices in the categories of performance and workflow. Performance was rated as fair due to the major disadvantage of inability to record ECGs without a network connection. Workflow was rated as fair because user-entered symptoms are not included in the report that users can e-mail to their clinicians. The device’s overall rating was fair because of these disadvantages. Cost of ownership was estimated at $140 over a 2-year period.

Although their original analysis was performed in 2018, device engineers looked at current models and did not note significant changes that would affect the overall device rating.

**Guiding Question 3. What is the influence of specific consumer automated-entry PGHD technologies on health outcomes? Does this vary across different patient populations, different settings, or other modifiers of effectiveness?**

Below, we discuss the efficacy results in four categories: 1) isolated effects on health outcomes, 2) isolated effects on surrogate outcomes, 3) multicomponent effects on health outcomes, and 4) multicomponent effects on surrogate outcomes. The risk-of-bias assessments are in Appendix Table C-45, and the data are in Appendix Table C-46.

**Health Outcomes**

Three studies\textsuperscript{135-137} that reported mortality and isolated the effect of the AliveCor monitor found no statistically significant between-group difference in mortality rates. However, these studies lacked adequate statistical power to detect a small difference due to the low number of deaths.

One study that reported stroke alone\textsuperscript{135} and another study that reported combined stroke/transient ischemic attack\textsuperscript{137} did not find a statistically significant between-group difference in these outcomes. Both studies isolated the effect of the AliveCor monitor. However, these studies lacked adequate statistical power to detect a small difference due to the low number of events.

Two studies\textsuperscript{135,137} found no significant between-group difference in major bleeding. Both isolated the effect of the AliveCor monitor but lacked adequate statistical power due to the low number of events.

One study\textsuperscript{136} that isolated the effect of AliveCor reported a statistically significant higher rate of emergency room visits for palpitation or pre-syncope in the AliveCor group. The authors hypothesized that some patients may have needed more reassurance than remote ECG transmission provided. Another study\textsuperscript{134} that did not isolate the effect of AliveCor from other interventions (text messaging) did not find a significant between-group difference in emergency room visits, although the number of visits was higher in the usual care group.

No outcomes reported by a single study (hospitalization, deep vein thrombosis/pulmonary embolism, or major adverse cardiac events) reported a statistically significant between-group difference, but the studies lacked adequate statistical power to detect a difference due to the low event rates.

**Surrogate Outcomes**

The only potential surrogate outcome reported in studies of isolated effects was time to arrhythmia detection. The single study\textsuperscript{136} that reported this outcome found a statistically
significant between-group difference favoring the AliveCor arm for reducing the time to arrhythmia detection.

Although no statistically significant between-group differences in health outcomes favored the PGHD intervention, the studies lacked sufficient statistical power to detect differences in those outcomes. However, one low risk study found a higher rate of emergency room visits in the AliveCor group, and the reason for this remains uncertain. Therefore, despite the significant positive effect of AliveCor observed for time to arrhythmia detection in this same study, the effect of AliveCor on health outcomes in patients with cardiac arrhythmias is unclear.

See Table 9 for a visual representation of PGHD effects on outcomes in studies of isolated effects. Overall, considering both health outcomes and surrogate outcomes, we deemed the evidence unclear on whether health outcomes are improved by PGHD interventions.

Out of eight ongoing trials of PGHD for cardiac arrhythmias, four planned to measure a health outcome or outcomes.
Table 9. Cardiac Arrhythmias: Categorization of Isolated Effects Due to Device Presence/Absence

<table>
<thead>
<tr>
<th>Health Outcomes</th>
<th>All results were p&lt;0.05 in favor of PGHD</th>
<th>Mixed results,* but most p&lt;0.05 results favored PGHD</th>
<th>No results were statistically significant</th>
<th>Mixed results,* but most p&lt;0.05 results against PGHD</th>
<th>All results p&lt;0.05 against PGHD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mortality</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stroke/ transient ischemic attack</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Major bleeding</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Emergency room visits for palpitation or pre-syncpe</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Major adverse cardiac events</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deep vein thrombosis/ pulmonary embolism</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Health outcomes not reported</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Statement about health outcomes based only on health outcomes data: **Unclear effect**

<table>
<thead>
<tr>
<th>Surrogate Outcomes</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Time to arrhythmia detection</td>
<td>●</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surrogate outcomes not reported</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Statement about health outcomes based on both health outcomes and surrogate outcomes: **Unclear effect**

Note: This table only displays studies that used an isolated-effect design to compare the presence vs absence of a PGHD device.

● Low risk of bias study of an isolated effect; ○ Moderate risk of bias study of an isolated effect; ⬤ High risk of bias study of an isolated effect

? Study did not report this category of outcome (the number of question marks indicates the number of studies)
Guiding Question 4. What are the harms or AEs associated with these specific consumer automated-entry PGHD technologies? Which patients in which settings are most at risk of harms?

None of the included studies identified or reported any AEs related to the PGHD technologies, although one study reported a higher rate of emergency room visits for palpitation or pre-syncope in the AliveCor group.

Guiding Question 5. For the technologies demonstrating associations with outcomes of interest, what full economic evaluations provide information on the relative value for consumers?

None of the included studies performed economic evaluations.

Stroke

Guiding Question 1. Which specific consumer automated-entry PGHD applications/technologies/devices have been studied for measurement of health outcomes? What studies are in progress of consumer PGHD devices?

a. What study designs have been used?
b. What were the inclusion/exclusion criteria?
c. What statistical analysis and data were used to determine study size and power?
d. How long were patients followed?
e. How was adherence measured?
f. What was the comparator?
g. Which outcomes were measured?

Devices

We included one trial on stroke. It used a BP monitor (Omron M6), and we judged it as similar to those currently on the market by this manufacturer.

Studies in progress

We identified four potentially relevant records in ClinicalTrials.gov. More details about these records (including hyperlinks) appear in Appendix Table C-47.

Study designs

Kerry et al. (2013)\textsuperscript{85-87} was a randomized trial in the United Kingdom that enrolled 381 patients and followed them for 1 year, and had a median enrollment month of June 2008. The next sections describe additional aspects of this trial, which are tabled in detail in Appendix Table C-48, C-49, and C-50.
Inclusion/exclusion criteria
Kerry et al. (2013)\textsuperscript{85-87} included patients with hypertension with a history of stroke or transient ischemic attack.

Statistical power analyses
Kerry et al. (2013)\textsuperscript{85-87} performed a power analysis that accounted for possible attrition, but did not mention use of prior data to inform the power analysis.

Follow-up length
One year.

Adherence measurement
Kerry et al. (2013)\textsuperscript{85-87} reported the percentage of patients who required the assistance of a care provider to take their BP, the percentage who recorded a full set of BP readings in the preceding 4 weeks, the percentage who said they still used the BP monitor 1.5 years after the cessation of nursing support, and the percentage who said they used the BP monitor at least once a month.

Comparators
Kerry et al. (2013)\textsuperscript{85-87} compared a group receiving only usual care to another group that received a BP monitor (Omron M6) as well as monthly calls from the nurse to check technique and review BP readings.

Outcomes reported
Kerry et al. (2013)\textsuperscript{85-87} reported health outcomes (QOL as measured by the EuroQol 5D), AEs (falls, recurrent stroke), device usage, and a process outcome (number of primary care consultations).

Guiding Question 2. What are the characteristics (e.g., interoperability, functions, acceptability/usability, sustainability, feasibility, fidelity, or integration into electronic health records) of these specific consumer automated-entry PGHD technologies?
Kerry et al. (2013)\textsuperscript{85-87} reported that 30% of intervention patients (51/168) required the assistance of a care provider to take their BP, 48% (80/168) recorded a full set of BP readings in the previous 4 weeks, 80 of 84 (95%) intervention patients answering questions at 18 months (after cessation of nurse support) said they still used the monitor, and 57 said they used it at least once a month. The data appear in Appendix Table C-52.
Guiding Question 3. What is the influence of specific consumer automated-entry PGHD technologies on health outcomes? Does this vary across different patient populations, different settings, or other modifiers of effectiveness?

The risk-of-bias assessments are in Appendix Table C-51, and the results are in Appendix Table C-52.

Isolated Effects on Health Outcomes

The trial did not report whether there were isolated effects on health outcomes. As a result, we rated this evidence as unclear.

Isolated Effects on Surrogate Outcomes

The trial did not use an isolated-effect design.
Table 10. Stroke: Categorization of Isolated Effects Due to Device Presence/Absence

<table>
<thead>
<tr>
<th>Health Outcomes</th>
<th>All results were p&lt;0.05 in favor of PGHD</th>
<th>Mixed results, * but most p&lt;0.05 results favored PGHD</th>
<th>No results were statistically significant</th>
<th>Mixed results, * but most p&lt;0.05 results against PGHD</th>
<th>All results p&lt;0.05 against PGHD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Health outcomes not reported</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Statement about health outcomes based only on health outcomes data: **Unclear effect**

<table>
<thead>
<tr>
<th>Surrogate Outcomes</th>
<th>Surrogate outcomes not reported</th>
<th></th>
</tr>
</thead>
</table>

Statement about health outcomes based on both health outcomes and surrogate outcomes: **Unclear effect**

Note: This table only displays studies that used an isolated-effect design to compare the presence vs absence of a PGHD device.
- 🔴 Low risk of bias study of an isolated effect;
- 🔴 Moderate risk of bias study of an isolated effect;
- 🔴 High risk of bias study of an isolated effect;
- 🔴 Study did not report this category of outcome (the number of question marks indicates the number of studies)

* “Mixed results” means that a study investigated either multiple time points (e.g., both 13 weeks and 26 weeks) or multiple comparisons (e.g., PGHD device A vs no device, and PGHD device B vs no device), and data showed statistical significance for some but not all time points/comparisons.

PGHD – Patient generated health data
Multicomponent Effects on Health Outcomes

Kerry et al. (2013)\textsuperscript{85-87} reported data on the 1-year rates of falls as well as recurrent stroke. The rate of falls was 17\% in the usual care group and 19\% in the BP monitoring group (difference not statistically significant). The rate of recurrent stroke was 8\% in the usual care group and 6\% in the BP monitoring group (difference not statistically significant). They also reported QOL as measured by the EuroQol 5D, and the groups did not have statistically significant differences at 1 year.

Multicomponent Effects on Surrogate Outcomes

Due to this evidence’s highly indirect nature, we do not discuss it; instead, we here refer the reader to the tabulated data in Appendix Table C-52.

Process Outcomes

Kerry et al. (2013)\textsuperscript{85-87} reported that over the 1-year period, the usual care group had an average of 5.4 primary care consultations compared to 5.2 in the BP monitoring group (difference not statistically significant).

Of the four stroke-related records in clinicaltrials.gov, one made a PGHD-related comparison and stated that they were collecting data on a health outcomes (quality of life).

Guiding Question 4. What are the harms or AEs associated with these specific consumer automated-entry PGHD technologies? Which patients in which settings are most at risk of harms?

The trial reported data on both falls and recurrent stroke, which we summarized in Guiding Question 3 above.

Guiding Question 5. For the technologies demonstrating associations with outcomes of interest, what full economic evaluations provide information on the relative value for consumers?

The stroke trial did not report either economic outcomes or cost-only data.
Parkinson’s Disease

No included studies addressed Parkinson’s disease. We identified four records in ClinicalTrials.gov that potentially involved PGHD interventions to prevent or treat Parkinson’s disease. As of April 7, 2020, one was enrolling by invitation, two were recruiting, and one was not yet recruiting. More details about these records (including hyperlinks) appear in Appendix Table C-53.

Chronic Obstructive Pulmonary Disease

Guiding Question 1. Which specific consumer automated-entry PGHD applications/technologies/devices have been studied for measurement of health outcomes? What studies are in progress of consumer PGHD devices?

a. What study designs have been used?
b. What were the inclusion/exclusion criteria?
c. What statistical analysis and data were used to determine study size and power?
d. How long were patients followed?
e. How was adherence measured?
f. What was the comparator?
g. Which outcomes were measured?

Devices

The nine chronic obstructive pulmonary disease (COPD) trials used nine different PGHD devices: four accelerometers, three pedometers, one BP monitor, and one forehead thermometer. All these devices (along with devices included in trials of other clinical conditions) appear in Appendix Table C-1. The similarity judgments (how similar each device is to those currently on the market by this manufacturer) were:

- Accelerometers: 2 similar, 1 somewhat different, 0 very different, 1 unknown
- Pedometers: 1 similar, 0 somewhat different, 0 very different, 2 unknown
- BP monitors: 1 similar, 0 somewhat different, 0 very different, 0 unknown
- Forehead thermometer: 1 similar

Studies in progress

We identified seven records in ClinicalTrials.gov that potentially involved PGHD interventions to prevent or treat COPD. As of April 7, 2020, two were active but not recruiting, three were recruiting, and one was not yet recruiting. More details about these records (including hyperlinks) appear in Appendix Table C-54.

Study designs

All nine studies of PGHD devices for COPD were RCTs. The RCTs were conducted in The Netherlands (3), Spain (2), Belgium (1), Germany (1), Japan (1), and the United Kingdom (1). Patient enrollment dates were not reported in two RCTs and ranged from September 2010-December 2010 to June 2015-July 2016 in the remaining seven RCTs. There were between 27
and 407 patients at the baseline visit per study. The next sections describe additional aspects of these trials, which are tabled in detail in Appendix Table C-55, C-56, and C-57.

**Inclusion/exclusion criteria**

Patients with a diagnosis of COPD were enrolled in all nine RCTs. Eight out of nine RCTs excluded patients with significant comorbidities, including significant cardiovascular comorbidities, mobility problems, and sarcoidosis. Four RCTs enrolled only patients 40 years of age or older, and a fifth RCT enrolled only patients 35 years of age or older. A sixth RCT enrolled retirees but did not specify a minimum age requirement for study entry. Two RCTs recruited patients who were either current smokers or had a history of smoking. Additional details appear in Appendix Table C-56.

**Statistical power analyses**

Four RCTs did not perform sample size calculations. The remaining five RCTs were calculated to detect a clinically meaningful difference in steps per day (2), exacerbation-free weeks (1), moderate-intensity physical activity (1), and a measure not clearly specified (1).

**Follow-up length**

The followup length was 1 to 12 months.

**Adherence measurement**

Two RCTs did not report PGHD device adherence.

**Pedometer**

Arbillaga-Etxarri et al. (2018) measured device adherence in two ways. Physiotherapists administering both interventions noted patients’ spontaneous reports of unwillingness to follow the instructions at the baseline visit, as well as spontaneous reports of nonadherence (i.e., not having followed the instructions) at the 12-month visit. At the 12-month followup visit, patients also answered a questionnaire about satisfaction with the study components and any potential AEs experienced during or after walks in the previous 12 months.

Demeyer et al. (2017) measured Fitbug adherence by chart review.

**Accelerometer**

Vorrink et al. (2016) measured adherence to the intervention as the percentage of days the intervention was used, and as percentage of days the physical activity goal was obtained.

Kawagoshi et al. (2015) reported that adherence was measured when the patient comes for a consultation every 2 to 4 weeks. The accelerometer stores data for up to 36 days, which investigators retrieved at patient visits.

Tabak et al. (2014) measured adherence in the intervention group by number of visits to the web portal and the time the activity sensor was worn. Only those days were included where at least 50% of the day was measured. Adherence was calculated by dividing the number of days the activity sensor was worn by the minimal number of days that was prescribed (i.e. ≥four days/week).

**Blood pressure**

Jodar-Sanchez et al. (2013) measured adherence automatically. After taking these measurements, patients’ data were sent via a hub (Tele-Modem, Aerotel Medical Systems)
connected to the patient’s home telephone line. Once measurements had been recorded by each connected instrument, the user pressed a button to activate data transmission.

**Forehead thermometer**

Boer et al. (2019)\(^{143}\) measured adherence based on the web-based interface, although this measure is not specific to the forehead thermometer of interest.

**Comparators**

The comparator in six of nine RCTs was usual care, which included education on COPD. Pulmonary rehabilitation (PR) was the comparator in two RCTs. In one study, PR entailed a multidisciplinary home-based program, including breathing training, exercise training, and an intensive educational program. In the other RCT, PR was an outpatient program lasting 8 weeks, with two supervised sessions per week. In the final RCT, physiotherapy in the form of weekly group training sessions at the local physiotherapy practice was the comparator.

**Guiding Question 2. What are the characteristics (e.g., interoperability, functions, acceptability/usability, sustainability, feasibility, fidelity, or integration into electronic health records) of these specific consumer automated-entry PGHD technologies?**

Below, we discuss the pertinent COPD data for this Guiding Question (see Appendix Table C-60).

Two RCTs did not report PGHD device adherence.

**Pedometer**

Arbillaga-Etxarri et al. (2018)\(^{37}\) reported that of the 132 Urban Training patients participating in the followup visit, 90\% used the pedometer and satisfaction with the pedometer was high or very high (9.0±1.8).

Demeyer et al. (2017)\(^{138}\) reported patients wore the Fitbug step counter for a median (Q1-Q3) of 91\% (84-98\%) of the days they were included in the coaching program, representing 6.3 (5.8-6.8) days/week.

**Accelerometer**

Vorrink et al. (2016)\(^{35,43}\) reported the intervention was used on 89±18.5\% of the study days. The personal physical activity goal was obtained on 34±16\% of these days.

Kawagoshi et al. (2015)\(^{139}\) reported that patients wore their pedometers for 293 (SD 49) days of a year (80.4\%, SD 13.3\%).

Tabak et al. (2014)\(^{140}\) reported that the activity coach was worn more than prescribed over the course of the one month study: for 17.5±2.2 days on average, which is 109\% of prescribed use. Only two patients used the system for fewer than the prescribed 16 days (13 and 14 days). In other words, 86\% of the patients complied with the activity coach. The average duration per day was almost 10 hours (588±101 minutes).

**Blood pressure**

Jodar-Sanchez et al. (2013)\(^{141,142}\) reported that patients in the telehealth group took daily BP measurements on 75\% of days (average 90 days; SD 0.22).
Forehead thermometer

Boer et al. (2019)\textsuperscript{143} report that 38 of the 43 patients (88\%) in the PGHD group used the app 727 times in total during followup. No usage data were available for five patients. The range in frequency of usage per participant was 1 to 250 times with a median of 7 (IQR 3 to 14). Results of the evaluation questionnaire showed that more patients reported to have used their mHealth tool often (scores 6 and 7 on the 7-point rating scale) compared with patients in the control group who used their paper action plan (44.4\% vs 17.2\%, respectively).

Boer et al. (2019)\textsuperscript{143} also had participants evaluate the supportive function of either the mHealth tool (as a whole unit) or the paper action plan by using a paper survey, including closed-ended questions regarding the use, difficulty in use, and intended future use of the mHealth tool or the paper action plan. Additionally, three questions were asked related to clarity, suitability, and followup of the advice given by the mHealth tool or the paper action plan. All questions included answers on a 7-point rating scale, from strongly disagree (score 1) to strongly agree (score 7). The survey also included one question about frequency of usage at times of symptom worsening, with answers on a 7-point rating scale varying from 1=never to 7=always. In addition, participants of the intervention group were asked to complete the System Usability Scale (SUS). The SUS contains 10 questions on system usability, which are calculated into 1 total score between 0 and 100. SUS scores less than 68 are considered as low, greater than or equal to 68 and less than or equal to 80.3 as good, and greater than 80.3 as excellent. A total of 58 (67\%) participants returned an evaluation form, of which 28 were in the intervention group. The mHealth tool was rated as a more useful support tool than the paper action plan (p=0.02). No differences were found between the mHealth tool and the paper action plan in the self-reported frequency of use; in difficulty and future use of the tool; or in clarity, suitability, and follow-up of the advice. Twenty-six participants in the PGHD group completed all 10 SUS questions, with a mean score of 78.5 (SD 14.4).

Guiding Question 3. What is the influence of specific consumer automated-entry PGHD technologies on health outcomes? Does this vary across different patient populations, different settings, or other modifiers of effectiveness?

The risk-of-bias assessments are in Appendix Table C-58, and the results are in Appendix Table C-60. Below, we discuss the efficacy results in four categories: 1) isolated effects on health outcomes, 2) isolated effects on surrogate outcomes, 3) multicomponent effects on health outcomes, and 4) multicomponent effects on surrogate outcomes

Isolated Effects on Health Outcomes

Kawagoshi et al. (2015),\textsuperscript{139} Nolan et al. (2017),\textsuperscript{144} and Vorrink et al. (2016)\textsuperscript{35,43} were the only RCTs that were designed in a way that allowed the effects of the PGHD of interest to be isolated. Both Kawagoshi and Nolan assessed PR plus pedometer use (Kenz Lifecorder EX, Nagoya, Japan and Yamax Digi-walker CW700; Yamax, Bridgnorth, United Kingdom, respectively) versus PR alone. Vorrink assessed usual care plus accelerometer (HTC Desire accelerometer) versus usual care alone.

Kawagoshi et al. (2015)\textsuperscript{139} reported that one patient in each treatment arm died and one patient in each treatment arm required hospitalization due to a COPD exacerbation. No tests of statistical significance were performed.
Kawagoshi et al. (2015)\textsuperscript{139} also reported Chronic Respiratory Questionnaire (CRQ) total scores. Both groups improved over time but study authors did not perform a between group test of statistical significance. The results were: PGHD baseline 98 (SD 20), 1-year follow-up 108 (SD 19) and UC 99 (SD 19) and 1-year follow-up 110 (SD 19). The authors reported on the subscales dyspnea (only PGHD group improved over time) and fatigue (neither group improved significantly over time). Again, no between group test of statistical significance was performed.

Nolan et al. (2017)\textsuperscript{144} also reported on deaths and hospitalizations. Two patients in each study arm died during the study. No test of statistical significance was performed. Nolan found a nonsignificant difference in rates of hospitalization between the PGHD and usual care groups. There were 56 total hospital admissions (PGHD 23; usual care 33; p=0.50). Thirty of these admissions were for COPD (PGHD 14; usual care [UC] 16; p=0.29).

Nolan et al. (2017)\textsuperscript{144} also reported on CRQ dyspnea scores (5 to 35, higher scores better health status). The authors found nonstatistically significant differences between the two study arms: change from baseline to immediately following PR, PGHD 3.7 (95% CI: 2.1 to 5.2) versus usual care 5.6 (95% CI: 4.2 to 7.0), p=0.07 and change from baseline to 6 months following PR: PGHD 1.8 (95% CI: -0.1 to 3.6) versus usual care 3.7 (95% CI: 2.1 to 5.3), p=0.10.

For CRQ total scores (range 20 to 140, with higher scores representing better health), change from baseline to immediately following PR, PGHD 11.0 (95% CI: 3.0 to 20.0) versus usual care 20.0 (95% CI: 8.0 to 27.0), p=0.008 and change from baseline to 6 months following PR: PGHD 3.0 (95% CI: -8.0 to 16.0) versus usual care 10.0 (95% CI: -2.0 to 19.0), p=0.07. Nolan et al. (2017)\textsuperscript{144} note that “unexpectedly, short-term improvements in CRQ scores following PR were significantly greater in the control group than in the intervention group for the total score (P 0.01)”. The authors also adjusted for baseline CRQ values, and the group effect for differences in the total scores remained significant. Between-group differences in CRQ did not persist at 6 months.

Vorrink et al. (2016)\textsuperscript{35,43} also reported CRQ results for dyspnea and fatigue (scores range from 1 to 7 for each subscale). Only fatigue showed a significant group by time interaction, however one group was not consistently favored over the other.

For dyspnea, the results were as follows: PGHD baseline 4.84 (SD 0.15), change at 3 months 0.17 (95% CI: -0.45 to 0.38), change at 6 months 0.11 (95% CI: -0.14 to 0.35), change at 12 months -0.17 (95% CI: -0.44 to 0.09) and UC baseline 4.79 (SD 0.15), change at 3 months 0.01 (95% CI: -0.21 to 0.23), change at 6 months -0.13 (95% CI: -0.33 to 0.08), and change at 12 months -0.08 (95% CI: -0.30 to 0.14), showing no differences between the groups (p=0.859). The group by time interaction was also nonsignificant (p=0.179).

For fatigue, the results were: PGHD baseline 4.35 (SD 0.1), change at 3 months 0.05 (95% CI: -0.15 to 0.26), change at 6 months -0.19 (95% CI: -0.39 to 0.01), change at 12 months -0.14 (95% CI: -0.35 to 0.07) and UC baseline 4.20 (SD 0.13), change at 3 months -0.06 (95% CI: -0.28 to 0.17), change at 6 months 0.13 (95% CI: -0.12 to 0.37), change at 12 months -0.12 (95% CI: -0.37 to 0.13).

**Isolated Effects on Surrogate Outcomes**

Vorrink et al. (2016)\textsuperscript{35,43} reported on lung function but did not find a statistically significant between-group difference for this outcome. FEV1/FVC showed no between group difference (p=0.34) or group by time interaction (p=0.908), meaning that the decline over time was not significantly different between the groups. FEV1 was significantly higher in the intervention group at the end of followup (PGHD group decreased at an average of 56 mL over the 1-year
followup period and 98 mL in the UC group [p=0.05]). However, the group by time interaction was non-significant (p=0.508), meaning that there was no effect of the intervention on FEV1.

Kawagoshi et al. (2015)\textsuperscript{139} reported 6-minute walk distance (6MWD) and found both groups significantly improved on this outcome. The results were as follows: PGHD baseline 369 (SD 119), 1-year follow-up 445 (SD 138) and UC baseline 404 (SD 148) and 1-year follow-up 467 (SD 151). No between groups comparison was performed.

Nolan et al. (2017)\textsuperscript{144} reported results from the Incremental Shuttle Walk Test (m). The authors did not find a significant between group difference for either the end of treatment time point or the 6-month followup visit. The results were as follows: change from baseline to end of treatment (PGHD 60 [95% CI 20 to 90] and UC 50 [95% CI 10 to 90], p=0.83) and change from baseline to 6-month followup (PGHD 30 [95% CI 0 to 70] and UC 10 [95% CI -30 to 70], p=0.25).

Vorrink et al. (2016)\textsuperscript{35,43} also reported 6MWD results and did not find a statistically significant difference for this outcome either. There was no significant decrease in 6MWD over time (p=0.53), and no differences between the groups (p=0.485). The group by time interaction was also non-significant (p=0.585). The results were as follows: PGHD baseline 456 (SD 14), change at 3 months 4.1 (95% CI −2.8 to 11.1), change at 6 months 4.8 (95% CI −3.9 to 13.5), change at 12 months 0.8 (95% CI −8.8 to 10.3) and UC baseline 461 (SD 8), change at 3 months 1.9 (95% CI −4.1 to 7.9), change at 6 months 3.3 (95% CI −2.9 to 9.6), and 4 (−2.4 to 10.3).

See Table 11 for a summary of isolated effects on health outcomes and surrogate outcomes.

Of the seven records related to COPD identified in clinicaltrials.gov, two made PGHD comparisons and also stated that they would measure health outcomes (https://ClinicalTrials.gov/show/NCT03238339 mentioned exacerbations and hospitalizations, and https://ClinicalTrials.gov/show/NCT04138173 mentioned health status, hospital admissions, and transplantation-free survival).
Table 11. COPD: Categorization of Isolated Effects Due to Device Presence/Absence

<table>
<thead>
<tr>
<th>Health Outcomes</th>
<th>All results were p&lt;0.05 in favor of PGHD</th>
<th>Mixed results,* but most p&lt;0.05 results favored PGHD</th>
<th>No results were statistically significant</th>
<th>Mixed results,* but most p&lt;0.05 results against PGHD</th>
<th>All results p&lt;0.05 against PGHD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Death</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hospitalization</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quality of life</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Health outcomes not reported</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Statement about health outcomes based only on health outcomes data: **Unclear effect**

<table>
<thead>
<tr>
<th>Surrogate Outcomes</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>FEV1 and FEV1/FVC</td>
<td>●</td>
</tr>
<tr>
<td>6-minute walk distance/Incremental shuttle walk test</td>
<td>○ ○ ●</td>
</tr>
<tr>
<td>Surrogate outcomes not reported</td>
<td></td>
</tr>
</tbody>
</table>

Statement about health outcomes based on both health outcomes and surrogate outcomes: **Unclear effect**

Note: This table only displays studies that used an isolated-effect design to compare the presence vs absence of a PGHD device.

- ● Low risk of bias study of an isolated effect; ○ Moderate risk of bias study of an isolated effect; ○ High risk of bias study of an isolated effect

Study did not report this category of outcome (the number of question marks indicates the number of studies)

* “Mixed results” means that a study investigated either multiple time points (e.g., both 13 weeks and 26 weeks) or multiple comparisons (e.g., PGHD device A vs no device, and PGHD device B vs no device), and data showed statistical significance for some but not all time points/comparisons.

COPD – Chronic obstructive pulmonary disease
FEV1 – Forced expiratory volume in one second
FEV1/FVC – Forced expiratory volume in one second/forced vital capacity
PGHD – Patient generated health data
**Multicomponent Effects on Health Outcomes**

Six RCTs assessed PGHD devices used in multicomponent telehealth interventions. On the whole, the interventions assessed did not improve health outcomes but may lead to more AEs. The one exception was the study by Jehn et al., (2013)\textsuperscript{145} which examined the AiperMotion 300 accelerometer combined with other at-home nonconsumer devices. See below for more details.

Arbillaga-Etxarri et al. (2018)\textsuperscript{37} reported that the Urban training group (including the pedometer Onstep-50, Geonaute) experienced more lower-extremity muscle pain than the control group. For all other outcomes, including other AEs, exacerbations (severe leading to hospital or ER admission) and QoL assessed by COPD assessment test (CAT) and Clinical COPD Questionnaire (CCQ) they did not find significant between-group differences.

Boer et al. (2019)\textsuperscript{143} assessed a forehead thermometer (FTN, Medisana AG) combined with other nonconsumer devices and reported COPD hospitalizations, exacerbation-free weeks, and QOL assessed by Nijmegen Clinical Screening Instrument. Boer did not find any statistically significant between-group differences for any of these outcomes.

Demeyer et al. (2017)\textsuperscript{138} studied the Fitbug Air combined with other interventions and reported on exacerbations, AEs, and QOL assessed by CAT and CCQ. Although Demeyer et al. found no between-group differences in exacerbations; they found that patients in the Fitbug Air arm had more musculoskeletal AEs than control group patients. For QOL, patients in the Fitbug Air arm had better scores on the functional state domain of the CCQ than control.

Jehn et al. (2013)\textsuperscript{145} examined the Aipermotion 300 accelerometer plus other at-home nonconsumer devices and reported exacerbations, hospital admissions due to exacerbations, and hospital length of stay for a COPD exacerbation. For all these outcomes, the PGHD group was favored. QOL measured by CAT scores was also reported. No test of significance was performed for differences in CAT scores between the groups at followup.

Jodar-Sanchez et al. (2013)\textsuperscript{141,142} assessed BP with the UA-767 BT, A&D Company as part of a multicomponent intervention. The authors reported death (1 from each group), COPD-related emergency room visits (favors PGHD), COPD-related hospitalizations and length of stay (no between group differences found) and QOL measured with the St George Respiratory Questionnaire (SGRQ)\textsuperscript{15} (no between-group differences found), EuroQol-5D (no between-group differences found), and a nonstandardized instrument (favors PGHD).

Tabak et al. (2014)\textsuperscript{140} reported only QOL as measured by the CCQ (no between-group difference found) and the Medical Research Council Dyspnea Scale (no between-group difference found).

**Multicomponent Effects on Surrogate Outcomes**

Three RCTs of multicomponent effects reported a surrogate outcome. Arbillaga-Etxarri et al. (2018)\textsuperscript{37} did not find a significant between group difference in 6MWD. Demeyer et al. (2017)\textsuperscript{138} found the change in 6MWD was significantly different (13.4, 95% CI (3.40 to 23.5) m, \(p<0.01\)), favoring the PGHD. Jehn et al. (2013)\textsuperscript{145} reported that the PGHD group showed a significant improvement in 6MWT distance between baseline and 9-month followup (Mean Difference in 6MWT: +87.0 (SD 65.7) meters, \(p=0.006\)) versus UC group, which showed no significant change (MD +23.9 (SD 70.3) meters, \(p=0.23\)). There was no between group difference found.
Jehn et al. (2013)\textsuperscript{145} also examined the Aipermotion 300 accelerometer plus other at-home nonconsumer devices and reported FEV1\% At 9-month followup, the PGHD group increased by 2.5\% (SD 5.2\%) while UC decreased by 0.07\% (SD 9.2\%). Both groups remained unchanged from baseline to 9-month followup, and the authors did not find statistically significant between group differences.

Guiding Question 4. What are the harms or AEs associated with these specific consumer automated-entry PGHD technologies? Which patients in which settings are most at risk of harms?

Only one of nine RCTs of COPD PGHD reported device-specific harms. Nolan et al. (2017)\textsuperscript{144} gathered data on AEs throughout the study (no further explanation provided). These authors report that one participant experienced an allergic reaction to the nickel baseplate of the accelerometer during baseline assessments and as a result was not randomly assigned.

Guiding Question 5. For the technologies demonstrating associations with outcomes of interest, what full economic evaluations provide information on the relative value for consumers?

A different version of the Kenz Lifecorder is currently unavailable from Amazon (https://www.amazon.in/Kenz-Lifecorder%C2%AE-Activity-Monitor-Pedometer/dp/B07RDKDVZC), so price information is not listed at this time. Yamax Digiwalker CW700; Yamax, Bridgnorth, United Kingdom, is available for $32.00 US. See https://www.yamax.co.uk/yamax-pedometers/cw700-cw701-digi-walker/ for more details. The HTC Desire accelerometer embedded in a smartphone A8181, HTC, Taoyuan, Taiwan, is available for $85.00 US at https://www.amazon.com/HTC-Smartphone-Touchscreen-Bluetooth-International-Warranty/dp/B0035ER8OY.

One study on COPD provided data from a cost-utility analysis. Jodar-Sanchez et al. (2013)\textsuperscript{141,142} reported costs in Euros at 2014 prices. Our risk-of-bias evaluation is in Appendix Table C-59. The analysis included: 1) accident and emergency department visits, specialized care consultations and hospital admissions evaluated according to public prices, 2) time employed by the CCC was estimated according to the alert type generated in the triage application: 25 min for clinical alerts, 15 min for alerts generated by nonadherence to the system, nonreceipt of data and technical reinforcement, and 10 min for alerts generated for other reasons. The clinical call center’s cost per hour was calculated in line with the salary rates of the Andalusian Health Service. 3) The time employed by the case manager was estimated depending on the exacerbation’s level of seriousness: 20 min for mild to moderate exacerbations, 25 min for severe exacerbations, and 30 min for very severe exacerbations. The case manager’s cost per hour was also calculated in line with the salary rates of the Andalusian Health Service. 4) The time employed by technical staff was estimated as 60 min for equipment installation and 30 min for technical incidents. This cost includes travel to the patient’s house, which was calculated by the company supplying the service. 5) The software and equipment cost was provided by the service-supplying company and was calculated using the equivalent annual cost, a method which accounts for both depreciation and the opportunity cost of the capital. The lifetime of the equipment and software was set at 5 years and the discount rate at 3\%. To calculate the
software’s cost, authors allowed for a maximum of 500 patients telemonitored with this infrastructure/software and used this information to estimate the cost associated for each patient. EuroQol-5D was used to estimate a utility score. The telehealth program’s effectiveness was estimated as a QALYs gain. For each patient, QALYs were calculated by using the area under the curve analysis, with linear interpolation of utility scores between baseline and 4 months of followup. Deceased patients were assigned a EuroQol-5D utility score of zero at 4 months. For each patient, the QALY (not taking into account the differences in the basal utility scores) and QALY gain (taking into account the differences in the basal utility scores) corresponding to the 4 months of monitoring was calculated. Results of cost-utility analysis were expressed in terms of the ICER, calculated as the difference in the average costs between the PGHD group and the control group divided by the difference in the average QALY gain between PGHD group and control group. Discounting of costs and QALYs was not necessary because the time horizon of the study, 4 months, did not extend beyond 12 months. To analyze uncertainty and verify the ICER’s robustness, the authors conducted an analysis using a nonparametric bootstrap with 5000 replications.

Average healthcare cost was €2064 for the PGHD group and €1103 for the UC group (a difference of €961; 95% CI: -809 to 2731). The average total cost was €2300 for the PGHD group and €1103 for the UC group (difference of €1197; 95% CI: -579 to 2973). The utility score’s average increase was 0.036 for the PGHD group and 0.003 for the UC group (difference of 0.032 score; 95% CI: -0.12 to 0.19). The UC group obtained a higher average QALY than the PGHD group, determined by the differences in the basal utility scores.

Cost-utility analysis based on 5000 bootstrap replications: The average total cost per patient was €2300 for the PGHD group and €1103 for the UC group, resulting in an incremental cost of €1197 (-498.97 to 2892.80). Corresponding figures for patients without and with comorbidities are as follows: PGHD 855.13 vs. UC 1353.85 ICER -498.72 (-2451.38 to 1453.94) and PGHD 2781.73 vs. UC 948.91 ICER 1832.83 (-223.00 to 3888.66). The average QALY gain for all patients combined was 0.0059 for the PGHD group and 0.0006 for the UC group, resulting in an incremental QALY gain of 0.0053 (-0.0193 to 0.0300). For patients without comorbidity, the figures are PGHD 0.0288 vs. UC 0.0082 ICER 0.0206 (-0.0259 to 0.0671) and for patients with comorbidity PGHD -0.0017 vs. UC -0.0041, ICER 0.0024 (-0.0251 to 0.0300).

Authors obtained an ICER of 223,726 E/QALY. The acceptability curve showed that for a willingness to pay of 30 000 €/QALY, the telehealth program’s probability of being cost-effective was 15%.
Asthma

Guiding Question 1. Which specific consumer automated-entry PGHD applications/technologies/devices have been studied for measurement of health outcomes? What studies are in progress of consumer PGHD devices?

- a. What study designs have been used?
- b. What were the inclusion/exclusion criteria?
- c. What statistical analysis and data were used to determine study size and power?
- d. How long were patients followed?
- e. How was adherence measured?
- f. What was the comparator?
- g. Which outcomes were measured?

Devices

The one identified asthma trial used a single PGHD device: a spirometer (Medical International Research, SmartOne). We rated it as “similar” to other devices on the market by this manufacturer. All devices (along with devices included in trials of other clinical conditions) appear in Appendix Table C-1.

Studies in progress

We identified five records in ClinicalTrials.gov that potentially involved PGHD interventions to prevent or treat asthma. As of April 7, 2020, four were recruiting and one was active but not recruiting. More details about these records (including hyperlinks) appear in Appendix Table C-61.

Study designs

The single study of a PGHD device for asthma used a randomized crossover design. It was conducted in Sweden. Patients were enrolled from May 2016 to September 2018. There were 77 patients at the baseline visit. The next sections describe additional aspects of this trial, which are tabled in detail in Appendix Table C-62, C-63, and C-64.

Inclusion/exclusion criteria

The study subjects were children aged ≥6 years and adults with a doctor’s diagnosis of asthma and Asthma Control Test/ (ACT)/Childhood Asthma Control Test (C-ACT) scores <20 points. A mean score ≤19 points indicated uncontrolled asthma in both tests. Exclusion criteria were presence of any comorbidity with significant impact on symptom control, participation in drug trials, and patient/caregiver difficulties in reading Swedish.

Statistical power analyses

The sample size was estimated assuming that AsthmaTuner would improve the average ACT/C-ACT score by 2 points compared with conventional treatment (mean SD 3.3). Assuming a dropout rate up to 10%, power calculations estimated that enrollment of 43 adults and
43 children would be clinically relevant and feasible to attain 80% power at a 5% significance level. The authors did not indicate that this estimate was based on prior research.

**Follow-up length**

The follow-up length was 2 months.

**Adherence measurement**

Device adherence to AsthmaTuner was captured by the cloud-based system’s backend data storage feature.

**Comparators**

The comparator in Ljungberg et al. (2019)\textsuperscript{146} was usual care with an individualized printed treatment plan. Specifically, usual care was defined as nondigital self-management using individual printed treatment plans, which contained treatment adjustments of prescribed medications according to symptoms of controlled, partly controlled, or uncontrolled asthma, along with instructions according to national guidelines.

**Guiding Question 2. What are the characteristics (e.g., interoperability, functions, acceptability/usability, sustainability, feasibility, fidelity, or integration into electronic health records) of these specific consumer automated-entry PGHD technologies?**

Below, we discuss the pertinent asthma data for this Guiding Question (see results in Appendix Table C-66.

**MIR SmartOne, part of the AsthmaTuner system**

Approximately 81% (62/77) of patients used AsthmaTuner (including the SmartOne spirometer, the PGHD of interest) on average once weekly or more. This figure included 73% (27/37) of adult patients with asthma and 87.5% (35/40) of pediatric patients with asthma assigned to the intervention. This was the only data provided in Ljungberg et al. (2019)\textsuperscript{146} that addressed Guiding Question 2.
Guiding Question 3. What is the influence of specific consumer automated-entry PGHD technologies on health outcomes? Does this vary across different patient populations, different settings, or other modifiers of effectiveness?

The risk-of-bias assessments are in Appendix Table C-65, and the results are in Appendix Table C-66.

Isolated Effects on Health Outcomes

Asthma symptom control was assessed at baseline and at the end visit of each treatment period with the ACT questionnaire in patients aged $\geq 12$ years and C-ACT in children aged 6-11 years. The PGHD group experienced better symptom control than patients receiving UC. However, when the patients were divided into subgroups by age, only pediatric patients using the PGHD had better symptom control than patients receiving UC.

The mean ACT/C-ACT score was 19.45 (95% CI: 18.70 to 20.21) for the PGHD group vs. 18.75 (17.97 to 19.53) for UC, difference 0.70 (95% CI: 0.06 to 1.34, $p=0.03$). These figures were: PGHD 19.14 (18.08 to 20.19) vs. UC 18.78 (17.63 to 19.94), difference 0.33 (95% CI: -0.68 to 1.35, $p=0.51$) for adult patients and PGHD 19.75 (95% CI: 18.65 to 20.85) vs. UC 18.73 (17.61 to 19.84), difference 0.97 (95% CI: 0.13 to 1.81, $p=0.02$) for pediatric patients.

A summary of the health outcomes appears in Table 12 below.
Table 12. Asthma: Categorization of Isolated Effects Due to Device Presence/Absence

<table>
<thead>
<tr>
<th>Health Outcomes</th>
<th>Symptom Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>All results were p&lt;0.05 in favor of PGHD</td>
<td>Mixed results,* but most p&lt;0.05 results favored PGHD</td>
</tr>
</tbody>
</table>

Statement about health outcomes based only on health outcomes data: **Possible positive effect**

Note: This table only displays studies that used an isolated-effect design to compare the presence vs absence of a PGHD device.

- Low risk of bias study of an isolated effect; ☣ Moderate risk of bias study of an isolated effect; ☎ High risk of bias study of an isolated effect
- ? Study did not report this category of outcome (the number of question marks indicates the number of studies)
- * "Mixed results" means that a study investigated either multiple time points (e.g., both 13 weeks and 26 weeks) or multiple comparisons (e.g., PGHD device A vs no device, and PGHD device B vs no device), and data showed statistical significance for some but not all time points/comparisons.

PGHD – Patient generated health data
Guiding Question 4. What are the harms or AEs associated with these specific consumer automated-entry PGHD technologies? Which patients in which settings are most at risk of harms?

Ljungberg et al. (2019)\footnote{146} reported that three patients (one each) terminated the study due to severe snake bite, pertussis, and another respiratory diagnosis. The study provided no other details.

Guiding Question 5. For the technologies demonstrating associations with outcomes of interest, what full economic evaluations provide information on the relative value for consumers?

Ljungberg et al. (2019)\footnote{146} did not report economic outcomes or cost-only data. The PGHD device is available through Amazon at a cost of $109.00 US. See \url{https://www.amazon.com/MIR-Smart-Personal-Pocket-Spirometer/dp/B07LGZ64KB} for more details.

Summary and Implications

Much research has investigated the use of automated-entry consumer devices to collect patient data to prevent or treat chronic conditions. Many of these devices, such as pedometers and BP monitors, have been on the market for many years, and therefore are in relatively mature phases of their product cycles. Others are relatively new (e.g., ECG monitors, body composition monitors), and there are far fewer manufacturers for certain device categories. Overall, PGHD devices are clearly providing a wealth of data to both patients and their providers.

But has this information actually improved health? Overall, we found mixed evidence; our primary results are summarized in Table 13 below and graphically in Figure 6 below. We found “possible positive effects” for four conditions: hypertension (mostly BP monitors), CAD (BP monitors, heart rate monitors), heart failure (BP monitors, scales), and asthma (spirometer). For obesity, we categorized the evidence as “likely no effect” on health outcomes, based on consistent lack of effect on BMI/weight, as well as unreported/inconsistent data on quality of life. The evidence was unclear for the other six conditions.

Table 13. Primary Findings

<table>
<thead>
<tr>
<th>Clinical Condition</th>
<th>Results Categorization for Isolated Health Outcomes</th>
<th>Comments</th>
</tr>
</thead>
</table>
| Obesity            | Likely no effect                                     | 3 of 43 included trials reported whether there were isolated effects on health outcomes (specifically, quality of life):
- 1 (low risk of bias) found a benefit of PGHD on physical quality of life at 3 months but not at 9 months, and no effect at either time point for mental quality of life.
- 1 (moderate risk of bias) found statistically nonsignificant differences
- 1 (high risk of bias) found a statistically nonsignificant effect on physical quality of life, but found a small benefit (4%) of PGHD on mental quality of life at 6 months.
14 trials reported whether there were isolated effects of device presence on surrogate outcomes (BMI or weight), and all 25 point estimates were less than the minimal important difference (5% body weight). |
<table>
<thead>
<tr>
<th>Clinical Condition</th>
<th>Results Categorization for Isolated Health Outcomes</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diabetes prevention</td>
<td>Unclear</td>
<td>None of the three trials reported whether there were isolated effects on health outcomes. One trial reported a surrogate outcome (metabolic syndrome risk) and it found an advantage of PGHD, however it was at high risk of bias.</td>
</tr>
<tr>
<td>Sleep apnea</td>
<td>Unclear</td>
<td>None of the three trials reported whether there were isolated effects on health outcomes. One trial reported a surrogate outcome (number of days on which apnea events were witnessed) and it found no statistically significant difference and was at high risk of bias.</td>
</tr>
<tr>
<td>Hypertension</td>
<td>Possible positive effect</td>
<td>Six of the 50 included studies reported whether there were isolated effects on health outcomes (including quality of life, mortality, and hospitalizations) - Four studies (one high risk of bias, one moderate, and two low) found overall no significant effects on quality of life, although one study found a significant effect favoring usual care for one subscale. - One study (moderate risk of bias) found no significant effects on hospitalizations. - One study (low risk of bias) found no significant effects on mortality. Seventeen studies reported whether there were isolated effects of device presence on surrogate outcomes (SBP, DBP, and BP control). - Seven of the 17 studies reporting on SBP found statistically significant findings favoring PGHD. Eighteen of 31 point estimates were greater than the minimal important difference of 2 mmHg. - Six of 16 studies reporting on DBP found statistically significant findings favoring PGHD. Eleven of 30 point estimate were greater than the minimal important difference of 2 mmHg. - Two of 11 studies reporting on BP control found statistically significant findings favoring PGHD.</td>
</tr>
<tr>
<td>Coronary artery disease</td>
<td>Possible positive effect</td>
<td>Mortality was significantly lower in the PGHD arm in the only study that reported it. Re-hospitalization was also lower but did not reach statistical significance.</td>
</tr>
<tr>
<td>Heart failure</td>
<td>Possible positive effect</td>
<td>Different quality of life measures favored the PGHD intervention group in two studies that isolated the effect of PGHD (although since multiple PGHD devices were used in each study the effect of any individual PGHD device could not be isolated).</td>
</tr>
<tr>
<td>Cardiac arrhythmias or conduction abnormalities</td>
<td>Unclear</td>
<td>There were no statistically significant between-group differences in health outcomes that favored the PGHD intervention. However, one low risk study found a higher rate of emergency room visits in the AliveCor group, and the reason for this remains uncertain. Therefore, despite the significant positive effect of PGHD observed for time to arrhythmia detection in this same study, the effect of PGHD on health outcomes in patients with cardiac arrhythmias is unclear.</td>
</tr>
<tr>
<td>Stroke</td>
<td>Unclear</td>
<td>The single trial did not report whether there were isolated effects on health outcomes</td>
</tr>
<tr>
<td>Parkinson’s Disease</td>
<td>Unclear</td>
<td>No studies met inclusion criteria</td>
</tr>
<tr>
<td>COPD</td>
<td>Unclear</td>
<td>3 of 9 RCTs reported isolated effects on health outcomes. 1 RCT each assessed Kens Lifecorder EX pedometer (High risk of bias, hospitalization, death and quality of life), Yamax Digi-walker CW700 pedometer (Low risk of bias, quality of life, hospitalization, death); and the HTC Desire accelerometer (Moderate risk of bias, quality of life).</td>
</tr>
<tr>
<td>Asthma</td>
<td>Possible positive effect</td>
<td>1 study met inclusion criteria (moderate risk of bias), and it found better symptom control in the PGHD group overall and in the pediatric population alone.</td>
</tr>
</tbody>
</table>
We have some concerns about the applicability of the patients and interventions in the included trials. Most of the patients enrolled in the included studies had typical characteristics of specific chronic conditions. However, females were underrepresented in studies of CAD and sleep apnea (<20% of enrolled patients in all studies for these conditions) and in some studies of COPD. Furthermore, few studies included rural populations for whom home-based PGHD interventions might be advantageous, since in-person visits with healthcare providers may be more of a burden for these populations (e.g., lack of nearby health care offices/clinics may mean longer travel time). Also, many studies required that participants have access to smartphones or the internet and have enough technical knowledge to operate remote monitoring technologies. Therefore, individuals without such access or knowledge are not well-represented in the literature.

Regarding the applicability of the interventions in the included studies, we note four issues:
- Some tested devices are available only in certain countries (e.g., body composition monitor InBody IU-U070B by Biospace in South Korea, or the Creta scale by Soehnle in Germany). Sixty-one of the 111 trials (55%) were conducted outside the United States, and their devices’ U.S. availability may be unclear.
• Some devices tested in these trials are no longer available, or their versions have been substantially updated since the trials completed. Across all clinical conditions, there were 116 unique devices, and device engineers rated 78 as similar to current devices, 11 as somewhat different, 1 as very different, and 26 as unknown.

• Some components of PGHD interventions are highly specialized or multicomponent with non-device related interventions and may not be feasible at other institutions. An example is the Arbillaga-Etxarri et al. (2018) study, where in addition to receiving a pedometer, patients spoke with a “respiratory physiotherapist” who was “adequately trained in behavioral strategies” and conducted “motivational interviews” that were “centered on empathy, reflective listening and affirmation, and addressed patients’ resistance (personal difficulties, barriers and limitations) to eliciting behavioral change.”

• Some technologies may require additional supportive technologies not available at all institutions. A key example is whether health care settings have the technology to receive electronically transmitted data and possibly include automatically into the electronic medical record. Websites and apps are often used to receive data, but it is unclear whether routine health care settings could easily access those data to improve patient management.

There has been a growing number of studies focused on recent PGHD technologies involving Apple products that were not reviewed in this report because they did not meet inclusion criteria. In particular, the Apple Heart study was a recent large observational study enrolling over 400,000 participants that tested a smartwatch (Apple Watch) app’s ability to identify atrial fibrillation during typical use. Because this was a single-arm study that did not include a control group of patients who received either usual care or another intervention, it did not meet inclusion criteria for our review. This study also focused on arrhythmia detection and did not report any health outcomes relevant to our review, except for AEs.

A recent evidence map for PCORI on mobile health technologies for self-management of several chronic diseases revealed some similarities and differences with the findings of this Technical Brief. The maps can be found at https://www.pcori.org/evidence-maps/results-strength-evidence-mHealth-systematic-reviews-2. We conducted a systematic literature search and identified 99 relevant systematic reviews published between 2010 and 2017. Some of the mobile health technologies (e.g. text messaging) included in that report did not meet the definition of consumer PGHD used in the Technical Brief. The only chronic diseases that overlapped with the ones examined in the current Technical Brief are cardiovascular disorders (encompassing CAD, chronic arrhythmias and heart failure), obesity, and respiratory disorders (encompassing asthma and COPD). The majority of systematic reviews in the PCORI report had unclear findings regarding health outcomes, which is consistent with the overall findings of the Technical Brief. However, for cardiovascular disorders the PCORI evidence map found no systematic reviews that reported a possible positive effect of mobile apps or wearable monitors, whereas the PGHD Technical Brief found a possible positive effect for health outcomes in studies of CAD and heart failure. Conversely, for obesity the PCORI evidence map found three systematic reviews that reported a possible positive effect and two that reported a positive effect on health outcomes (3 other systematic reviews were unclear and one showed no effect). The evidence base on obesity in this Technical Brief suggested there was likely no effect of PGHD interventions on health outcomes (although few studies reported health outcomes). The reason for this discrepancy may be that the PCORI report did not determine whether weight loss
in the systematic reviews was clinically significant, while our Technical Brief did make this
distinction. For respiratory disorders the majority of systematic reviews in the PCORI report
were unclear, but two reviews found a possible positive effect and one review found a positive
effect on clinical outcomes. Our Technical Brief found unclear evidence for the effect of PGHD
interventions on COPD but a possible positive effect on asthma. We note that many of the
systematic reviews in the PCORI report covered multicomponent interventions that did not
isolate the effect of mobile apps or devices, and some devices included in that report would not
have met the definition of consumer PGHD technologies used in the Technical Brief. These are
likely reasons for some of the differences in findings between the two reports.
Next Steps

For each clinical condition, we discuss the types of future research that are mostly likely to benefit patients with these conditions.

For obesity, a critical need is to measure health outcomes, rather than merely BMI or weight. Few of the obesity studies even chose to measure health outcomes.

Only one of the included studies on diabetes prevention had used a glucose meter, and this may be a promising target for future research. Such a meter could educate patients about the correspondence between eating/exercise behavior and their short-term glucose, which may prevent later rises in HbA1c and progression to diabetes.

None of the three included sleep apnea studies used a sleep-specific consumer device (CPAP devices require a prescription). Currently CPAP is the leading therapy for sleep apnea so it would make sense that this would require a prescription. However some newer devices, such as the Apple Watch could help accelerate the need to diagnose sleep apnea in select patients, by providing select information to signal that a patient has a high probability of sleep apnea. This would be helpful since the current method of checking for sleep apnea is based on a patient’s habitus and it is not routinely screened for.

Although current evidence suggests PGHD may improve surrogate outcomes for hypertension, most promisingly SBP, there is still some inconsistencies in the findings. Furthermore, few existing studies address important health outcomes, such as mortality and cardiovascular events, directly. Longer-term RCTs that isolate PGHD’s effects are needed to clarify impacts on both surrogate and health outcomes.

For patients with CAD, future RCTs evaluating PGHD interventions should attempt to enroll more female patients because they are underrepresented in currently available RCTs (<20% of enrolled patients). Also, more studies of CAD should evaluate health outcomes, including overall mortality, cardiovascular mortality, hospitalizations, and emergency room visits.

For patients with cardiac arrhythmias, more RCTs are needed to evaluate the effect of home-based smartphone-connected ECG monitors on health outcomes, including prevention of stroke, hospitalization, or death. Although not reported in the evidence base, the potential harms related to false positives (e.g. patients unnecessarily visiting emergency rooms) should be investigated in future studies. In particular, the increasing popularity of the Apple Watch and its app for cardiac arrhythmia detection means that RCTs are needed to evaluate this technology’s impact on the health outcomes noted above.

For patients with heart failure, current RCTs have generally combined different PGHD interventions (BP monitors, ECG monitors, and weighing scales) without attempting to isolate any individual device’s effect on health outcomes. Future RCTs should attempt to isolate the effect of individual PGHD devices to determine which devices are most useful to help patients to manage heart failure. These studies should examine upstream outcomes such as hospitalization, medication adherence and change in care.

For patients with stroke, we included only one trial that used BP measurements and showed no isolated effects on health outcomes. A future important PGHD study to evaluate stroke would involve home-based post-stroke rehabilitation using rehabilitation technologies. Regular rehabilitation is essential in stroke recovery and to improve long-term impact on patients’ QOL, especially in terms of mobility. Other stroke studies should focus on secondary prevention, since the recurrence rate is significant in this patient population.

Patients with Parkinson’s disease might benefit from PGHD devices that measure mobility. We found four relevant records in ClinicalTrials.gov on this condition, and one planned to use a
device (ParkinPal) that “utilizes a smartwatch to periodically record motion data,” and the data are “processed in a connected smartphone and translated into symptom scores for dyskinesia, slowness, and tremor.” The symptom scores can be used by providers to adjust medications and signal if the disease is getting worse or better.

For COPD, future researchers need to understand the importance of isolating the effect of individual monitoring strategies, so their unique contributions can be understood. Patients with severe comorbidities, including mobility limitations and cardiovascular issues, were generally excluded from the COPD evidence base. Future research should test some of these monitoring strategies on patients with COPD and comorbidities to determine whether they are more or less effective in higher risk COPD populations.

Regarding asthma, future research should focus on patients with adult asthma with comorbidities and patients with asthma living in other settings and countries and determine why up to 20% of patients did not adhere to device use.
References


97


168. BaronFaust R. Sticking with ART there may be an APP for that. AIDS Read. 2013 Jul 10;1-3.


242. Demment MM, Graham ML, Olson CM. How an online intervention to prevent excessive gestational weight gain is used and by whom: a randomized controlled process evaluation. J Med Internet Res. 2014;16(8):e194. Also available: http://dx.doi.org/10.2196/jmir.3483. PMID: 25143156.


106


329. Kamps AWA, Roorda RJ, Brand PLP. Peak flow diaries in childhood asthma are unreliable. Thorax. 2001;56(3):180-2. Also available: http://dx.doi.org/10.1136/thorax.56.3.180, PMID: 11182008.


127


