Evidence Generation Guide for Apps and Wearables Developers: Study Designs Including Applied Examples

Final Report

Produced by Newcastle and York External Assessment Centre

JOYCE CRAIG, Associate Project Director
JUDITH SHORE, Research Assistant

September 2016
Contents

Abbreviations

Section 1: Introduction
1.1 Mobile apps and digital wearables 1
1.2 Growth in apps and wearables 1
1.3 aim and structure of document 2

Section 2: What are Quantitative and Qualitative Studies?
2.1 Quantitative Intervention Research 1
2.2 Qualitative Research and Comparison to Quantitative Research 2
2.3 Mixed Methods Research 3
2.4 All Studies 3

Section 3: Randomised Controlled Trials
3.1 Randomised Controlled Trials 5
3.2 RCT Study Designs 7
3.3 Parallel Group RCT 8
3.4 Crossover RCTs 10
3.5 Cluster RCTs 12
3.6 Factorial RCTs 15
3.7 Stepped Wedge Cluster RCT 18
3.8 Delayed-Start RCTs 19
3.9 Adaptive Designs 22

Section 4: Observational Studies of Interventions
4.1 Cohort Studies 24
4.2 Case Series or Case Reports 26
4.3 Cross-Sectional Studies 28
4.4 New-User Designs 30

Section 5: Safety Studies 32

Section 6: Accuracy and Diagnostic Studies 34

Section 7: Economic Studies alongside RCTs
7.1 Conducting an Economic Evaluation Within a Clinical Trial 37
7.2 Challenges with Apps Interventions 38
7.3 Forms of Economic Analyses 38
7.4 Cost Consequences Analysis 39

Section 8: Qualitative Studies 42
8.1 Focus Groups 42
8.2 Qualitative Research Interviews and Surveys 44
8.3 Observational Methods 49

Section 9: Mixed Methods 52

Section 10: Engineering Informed Study Designs 56
10.1 Ecological Momentary Assessment Studies 56
10.2 Small Studies and N-1 Studies 59
10.3 Interrupted Time Series 61
10.4 Multiphase Optimisation Strategy (MOST) 64
10.5 Sequential, Multiple Assignment, Randomised Trials (SMART) 66
10.6 Real World Studies 68
10.7 A/B Testing 70

Section 11: Patient-Reported Outcomes and Experiences Studies 72

Section 12: Feasibility and Pilot Studies 77

Section 13: Innovative Study Designs 79
13.1 Instrumental Variables 79
13.2 The Use of Big Data in Healthcare 79

Section 14: APP Development and Evaluation 82
14.1 Evidence on the Cycle of Development 82
14.2 Implications of the Evidence on the Cycle of Development 83
14.3 A Research Strategy for Care Apps 84

Section 15: Wearable Devices 87

Section 16: Systematic Reviews and Meta-Analyses 91
16.1 Systematic Reviews 91
16.2 Meta-Analyses 92

Section 17: Critical Appraisal and Associated Checklists to Assess the Quality of Studies 94

Section 18: Hierarchy of Evidence 98

Section 19: Health and Social Care Information Centre 101
19.1 Role of Health and Social Care Information Centre 101
19.2 HSCIC’s Role in Supporting Clinical Safety 101
19.3 HSCIC’s Role in Supporting IT Infrastructure 102
19.4 HSCIC’s Role in Supporting Information Systems and Information Governance 103
19.5 Information Standards and Collections 104

Section 20: Ethical and Other Approvals 105
20.1 Ethical Principles 105
20.2 Research Governance in NHS and Social Care 105
20.3 Approvals for Studies to be Conducted in the NHS in England 106
20.4 Approvals for Studies to be Conducted in Universities 107
20.5 Publication and Ethics 107

**Section 21: General Sources of Information** 109

21.1 Conduct and Reporting of Research 109
21.2 Health and Social Care Organisations 110
21.3 Independent Organisations 114
21.4 Regulatory Bodies 115
21.5 Standards Development Agencies 117
21.6 Platform Developers 119

**Section 22: Academic Health Science Networks** 120

**Section 23: Glossary** 122
Acknowledgements

York Health Economics Consortium would like to thank all experts advising NICE who gave their time to shape this document and commented on earlier drafts.
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGATE</td>
<td>Adaptive, Goal-directed Adherence Tracking and Enhancement</td>
</tr>
<tr>
<td>AHRQ</td>
<td>Agency for Healthcare Research and Quality</td>
</tr>
<tr>
<td>AI</td>
<td>Adaptive interventions</td>
</tr>
<tr>
<td>CASP</td>
<td>Critical Appraisal Skills Programmes</td>
</tr>
<tr>
<td>CBER</td>
<td>Center for Biologics Evaluation and Research</td>
</tr>
<tr>
<td>CBT</td>
<td>Cognitive behavioural therapy</td>
</tr>
<tr>
<td>CCA</td>
<td>Cost consequences analysis</td>
</tr>
<tr>
<td>CDER</td>
<td>Center for Drug Evaluation and Research</td>
</tr>
<tr>
<td>CHEERS</td>
<td>Consolidated Health Economic Evaluation Reporting Standards</td>
</tr>
<tr>
<td>CHW</td>
<td>Community health workers</td>
</tr>
<tr>
<td>CONSORT</td>
<td>Consolidated Standards of Reporting Trials</td>
</tr>
<tr>
<td>COSMIN</td>
<td>Consensus-based Standards for the selection of health Measurement Instruments</td>
</tr>
<tr>
<td>CVD</td>
<td>Cardiovascular disease</td>
</tr>
<tr>
<td>EAC</td>
<td>External Assessment Centre</td>
</tr>
<tr>
<td>EMA</td>
<td>Ecological momentary assessment</td>
</tr>
<tr>
<td>EMI</td>
<td>Ecological momentary intervention</td>
</tr>
<tr>
<td>EQ 5D</td>
<td>EuroQol-5D</td>
</tr>
<tr>
<td>FDA</td>
<td>US Food and Drug Administration</td>
</tr>
<tr>
<td>FXS</td>
<td>Fragile X-syndrome</td>
</tr>
<tr>
<td>GD</td>
<td>Gestational Diabetes</td>
</tr>
<tr>
<td>GP</td>
<td>General Practitioner</td>
</tr>
<tr>
<td>HRT</td>
<td>Hormone replacement therapy</td>
</tr>
<tr>
<td>HSCIC</td>
<td>Health and Social Care Information Centre</td>
</tr>
<tr>
<td>ICH</td>
<td>International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use</td>
</tr>
<tr>
<td>ISPOR</td>
<td>International Society for Pharmacoeconomics and Outcomes Research</td>
</tr>
<tr>
<td>IT</td>
<td>Information Technology</td>
</tr>
<tr>
<td>ITS</td>
<td>Interrupted time series</td>
</tr>
<tr>
<td>IV</td>
<td>Instrumental variables</td>
</tr>
<tr>
<td>MARS</td>
<td>Mobile App Rating Scale</td>
</tr>
<tr>
<td>MHRA</td>
<td>Medicines &amp; Healthcare products Regulatory Agency</td>
</tr>
<tr>
<td>MOST</td>
<td>Multistage optimisation strategy</td>
</tr>
<tr>
<td>MRC</td>
<td>Medical Research Council</td>
</tr>
<tr>
<td>NHS</td>
<td>National Health Service</td>
</tr>
<tr>
<td>NICE</td>
<td>National Institute for Health and Care Excellence</td>
</tr>
<tr>
<td>NIHR</td>
<td>National Institute for Health Research</td>
</tr>
<tr>
<td>OHS</td>
<td>Oxford Hip Scale</td>
</tr>
<tr>
<td>PREM</td>
<td>Patient-reported experience measure</td>
</tr>
<tr>
<td>PRISMA</td>
<td>Preferred Reporting Items for Systematic Reviews and Meta-Analyses</td>
</tr>
<tr>
<td>PRO</td>
<td>Patient-reported outcome</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>PROM</td>
<td>Patient-reported outcome measures</td>
</tr>
<tr>
<td>PROQOLID</td>
<td>Patient-Reported Outcome and Quality of Life Instrument Database</td>
</tr>
<tr>
<td>QALY</td>
<td>Quality-adjusted life-years</td>
</tr>
<tr>
<td>QUADAS</td>
<td>Quality Assessment of Diagnostic Accuracy Studies</td>
</tr>
<tr>
<td>RCT</td>
<td>Randomised controlled trial</td>
</tr>
<tr>
<td>ROC</td>
<td>Receiver-Operating Characteristic</td>
</tr>
<tr>
<td>SASED</td>
<td>Smartphone Alcohol and Side Effects Diary</td>
</tr>
<tr>
<td>SIGN</td>
<td>Scottish Intercollegiate Guidelines Network</td>
</tr>
<tr>
<td>SMART</td>
<td>Sequential, Multiple Assignment, Randomised Trials</td>
</tr>
<tr>
<td>STARD</td>
<td>Standards for the Reporting of Diagnostic accuracy studies</td>
</tr>
<tr>
<td>STROBE</td>
<td>Strengthening the Reporting of Observational studies in Epidemiology</td>
</tr>
<tr>
<td>STTP</td>
<td>Structured treatment and teaching programme</td>
</tr>
<tr>
<td>TREND</td>
<td>Transparent Reporting of Evaluations with Nonrandomized Designs</td>
</tr>
<tr>
<td>UK</td>
<td>United Kingdom</td>
</tr>
<tr>
<td>USA</td>
<td>United States of America</td>
</tr>
<tr>
<td>VAS</td>
<td>Visual analogue scale</td>
</tr>
</tbody>
</table>
Section 1: Introduction

1.1 MOBILE APPS AND DIGITAL WEARABLES

The broad scope of digital health includes categories such as mobile health (mHealth), health information technology (IT), wearable devices, telehealth and telemedicine, and personalized medicine. This report focuses on the need to generate evidence to support the case for adopting a subset of these categories, being mobile apps and digital wearables within the NHS. These have the potential to:

- Improve peoples’ ability to better manage their health and well-being and thereby reduce their risks of contracting disorders such as cancer, heart disease and diabetes.
- Enhancing an individual’s ability to self-manage such conditions and support adherence to treatment strategies.
- Alert healthcare professionals when there is a change in their condition and enable clinicians to personalise treatments.
- Reduce system inefficiencies and costs.

Despite these potential benefits, the healthcare system lags behind other industries in adopting these technologies. One of the biggest concerns identified by doctors is the limited evidence on outcomes, including cost savings [1]. This document focus on providing such evidence to promote faster uptake of these technologies in the NHS.

1.2 GROWTH IN APPS AND WEARABLES

Over the past decade, there has been an exponential growth in the use of mobile apps in healthcare for a variety of purposes ranging from capturing health data directly from patients to presenting healthcare information and delivering or shaping interventions [2]. A systematic search of the Apple iOS iTunes and Android Google Play app stores targeting high-need, high-cost, “patient-facing” apps for use by individuals with chronic illnesses and in English, estimated that there are currently around 1,500 healthcare apps which fit this classification available for download [3]. However, one of the negative concomitants to this growth in mobile apps is the large number of poor quality apps that are available on the market [4]. There have been recent steps in developing frameworks for evaluating the user engagement [3] and the quality of user experience [4] to assist app developers in product design, however, to date there has been little focus on the quality of evidence generated to support the adoption of mobile apps [5].

In healthcare, wearables have been used for a long time, for example in hearing aids and in detecting health disorders such as sleep apnoea or irregular heart rhythms. However, recently there has been a rapid increase in sales of healthcare and fitness related devices. For example, in 2014, estimated global wearable unit sales were 35 million, rising to 78 million in 2015 and forecast to be 121 million in 2016. Unit sales of wearable technology in the UK from
January to September 2014 were 420,000 units, with health and fitness trackers forming almost 40% of sales, followed by wrist sport computers (26%), cameras, headsets and glasses (24%) and smart watches forming just over 10% of sales [6].

In the UK, currently around 1 in 7 (14%) own any wearable technology, with 7% owning a fitness band and 3% owning a smart watch. Consumers aged 16-34 show the strongest interest in wearables, being about 3 times more likely than those aged 35+ to own a fitness band [7]. However those born between 1946 and 1964 are now catching up being the fast growing segment of the market.

Whilst the growth in apps, smart phone users and adoption of a wide range of biosensors in wearables offers opportunities to provide clinically and cost-effective solutions to meeting patients’ needs and address some of the problems facing the NHS there are still concerns about the quality, reliability and security of the data and information from them [1].

1.3 AIM AND STRUCTURE OF DOCUMENT

This document does not prescribe methodologies for specific research questions to address each concern, but is rather part of a larger effort to develop a systematic approach to determining which study designs are best able to address given research questions. Its aim is to provide information for app and wearable developers and users of evidence, including commissioners, evaluators and the public, to understand the relative strengths and limitations of various research designs and how their use may affect study results and interpretation.

Some users, e.g. commissioners, may have a particular focus on safety, effectiveness and value for money. However, underpinning all of these is acceptability to potential users, hence any testing with users should involve not just measures of effect, but also of engagement [8]. Data collection to inform how well an app or wearable is meeting its goals will also inform optimisation, through further development and evaluation. Hence data collection should be viewed as an on-going process, recognising that not all data variables can be collected all of the time and some prioritisation will be necessary. Such long-term follow-up also enables the developer to determine whether short-term changes persist and potentially identify unintended consequences and rare adverse events.

This document provides a brief introduction to various study designs that can be used to provide quantitative and qualitative evidence. A research strategy to make a product market-ready for use in the NHS is likely to draw on several study designs, at different stages in its life cycle (see Section 14 for an example of such a research strategy). Hence developers will likely build a cumulative evidence base, informed by information gained at each key stage such as initial development, user and clinician testing and scaling, before its widespread adoption.

The rapidly evolving nature of these technologies adds further complexity to this picture. Developers may see a benefit in delaying undertaking a complex, and potentially expensive, clinical trial until the app or wearable and its associated components are stable and lessons have been learned from implementing it across several sites, making it possible to estimate
the potential impact on clinical and patient outcomes before commissioning the pivotal study. However, commissioners may not want to adopt the product which has the potential for harm or requires material service re-design or funding until such a definitive study is conducted.

Related to this is that a major trial is likely to identify new information that can be used for quality improvement but to improve the product may limit the value to be gained from the results of the expensive study – whilst refusing to improve may undermine the chances of success as weaknesses have been identified but not remedied [9].

Hence for developers, the research strategy should be an important part of development and commercialisation plans.

As a consequence of the probable need for several study designs over the lifetime of an app or wearable to accumulate evidence, study types are not mapped to specific research questions. Rather the designs provide a range of resources available to generate the evidence required to answer each research question. Moreover, each product is very different; providing a spectrum of potential harms, reach, and clinical, administrative and technical benefits, so the parameters to be measured will also vary. There is thus no single study design, which must be used to answer a specific criterion.

For each study design, this document provides:

- A brief overview of design features;
- A list of potential strengths and disadvantages;
- References;
- Where possible, an example of its use in practice with an app and a wearable product, presented as a short summary of a published study for a product using the relevant design.

After reviewing quantitative designs, safety, diagnostic and economic designs are presented before reviewing qualitative and mixed method study designs. The next sections consider engineering approaches to generate evidence, designs to inform patient-reported outcomes and experiences, feasibility and pilot studies, and innovative study designs such as instrumental variable studies and how big data can inform clinical effectiveness. The final study design sections address issues specific to wearables, systematic reviews, critical appraisal and the hierarchy of evidence.

Subsequent sections provide information for the website but do not relate to study design. These discuss:

**Section 2: Role of the**
Health and Social Care Information Centre (HSCIC) in ensuring the clinical safety, interoperability and information governance of apps;

Section 3:
• Ethical and Other Approvals;

Section 4:
• General Sources of Information which apps developers may find useful to consult when planning a new study;

Section 5: Examples of how
• Academic Health Science Networks can support developers in generating and evaluating evidence.

Section 6: A
Glossary is provided.

Where the words “app” or “apps” are used in this document these are shorthand for digital applications, which may include a mobile app, a web-based application or in certain cases a digital service. The term ‘wearable’ is used as shorthand for digital healthcare-related wearable device, including the sensors, digital processing applications and communication equipment. The evidence for either should measure the performance of the wearable or app software and sensors, related components such as face-to-face appointments, dedicated websites and the platform. Evaluation will consider the intervention package and not the app or wearable in isolation from these components.

References

Section 7: What are Quantitative and Qualitative Studies?

7.1 QUANTITATIVE INTERVENTION RESEARCH

Quantitative intervention research is designed to test well-specified hypotheses, for example, to determine whether an intervention did more harm than good compared to current practice. The aim of quantitative studies is to generate numerical data, which can be analysed to inform results to assist in answering the research problem. Such studies usually have a research protocol which, amongst other aspects, describes the planned methodology for data collection and the statistical methods which will be used (see ICH Good Clinical Practice guidelines). Typical quantitative research study designs are randomised controlled trials (RCT), cohort and case studies.

7.1.1 Experimental and Non-Experimental Study Designs

There are 2 main classifications of quantitative intervention studies. They can have experimental study designs and methods or non-experimental.

In experimental study designs, patients/users are often allocated to either an app/wearable or a comparator intervention such as standard care, for example, with the protocol defining the nature and timing of subsequent observations and actions. Randomised controlled trials (RCTs) provide the strongest level of confidence that observed outcomes are a result of the intervention.

In non-experimental studies, patients are not allocated to a treatment or intervention; rather patients and clinicians are observed in a setting, often a usual care setting, with outcomes observed for different interventions. These describe the impact on patients of real-world treatment decisions. A non-experimental study design may be used to inform on comparative effectiveness and safety but cannot conclude, with any degree of certainty, whether a change in outcomes occurred because of the isolated effect of a single intervention.

In experimental studies random allocation is often used - this reduces the potential for bias (intentional or unintentional) by ensuring factors other than the intervention of interest which may influence the outcome are randomly distributed between each arm in the study.
The choice of an experimental or non-experimental study design will determine key methodological issues relating to patient selection, definition of the intervention and controls, method of allocation, use of blinding and placebos, choice of outcome measure and type of analysis. The usefulness of either study design depends on the relevance of their interventions, participants and outcomes to the NHS setting and the research question. For adoption of new medicines, where relative efficacy and safety are key parameters for decision-makers, RCTs are preferred. Increasingly, RCTs are seeking to be more directly relevant to clinical practice whilst still retaining the safeguards against bias.

### 7.2 QUALITATIVE RESEARCH AND COMPARISON TO QUANTITATIVE RESEARCH

Qualitative research is used to explore and understand people's beliefs, experiences, attitudes, behaviour and interactions. Qualitative studies generate non-numerical data, e.g. a patient's description of their pain rather than a measure of pain. In healthcare, qualitative techniques have been commonly used to document patients’ experience of care and the usefulness of new technologies from a service provider perspective, and to understand issues in the adoption and maintenance of interventions. Qualitative study designs include observations, interviews and focus groups. They are essential to find out more about the views and experiences of patients, carers and service providers.

The robustness of both qualitative and quantitative methods depends on:

- Participant selection which must be well reasoned and the sample relevant to the research question;
- Appropriate analysis and interpretation of data.

Different ontological and epistemological approaches can underpin the choice between qualitative and quantitative designs. From a practical standpoint, key differences often include data collection processes, which are likely to include field observation, and interviews for qualitative studies. In contrast, quantitative studies may rely more on pre-specified forms and templates into which data are entered. Data analyses approaches for qualitative studies also differ from the statistical analyses required for quantitative studies. Common approaches to interpret qualitative data include grounded theory and various forms of thematic analysis.

With quantitative studies there are well known tests to measure the statistical significance of results and the confidence we can have in it (e.g. 95% confidence intervals). With qualitative studies no such numerical approach is used. Researchers may use triangulation, whereby the researcher uses multiple methods, sources or theories to provide evidence to increase the validity of their results. The greater the consistency across sources and analysts, the stronger the confidence one can have in the results. Quantitative studies often aim for a level of generalizability to a wide population. The findings of qualitative studies often provide very rich understandings of a specific context.
7.3 MIXED METHODS RESEARCH

The 2 methods can be used in tandem or 1 may precede the other to give a mixed methods study design. For example, a qualitative study to identify users' behaviours when provided with a new app/wearable may inform modifications to the app/wearable and associated training and documentation to optimise its effectiveness before a quantitative investigation is undertaken.

References

Good reference sources are:


7.4 ALL STUDIES

A pictorial representation of the different study types is provided at Table 2.1.
<table>
<thead>
<tr>
<th>Engineering Studies</th>
<th>Qualitative Studies</th>
<th>Experimental</th>
<th>Observational</th>
<th>Mixed Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ecological movement assessment</td>
<td>Focus group</td>
<td>Parallel RCT</td>
<td>Cohort</td>
<td>Any combination of Qualitative and Quantitative studies</td>
</tr>
<tr>
<td>Small studies</td>
<td>Interviews</td>
<td>Crossover RCT</td>
<td>Case control</td>
<td></td>
</tr>
<tr>
<td>Interrupted time series</td>
<td>Observations</td>
<td>Cluster RCT</td>
<td>Case series</td>
<td></td>
</tr>
<tr>
<td>Multiphase optimisation strategy (MOST)</td>
<td></td>
<td>Factorial RCT</td>
<td>Case studies</td>
<td></td>
</tr>
<tr>
<td>Sequential multiple assignment randomised trials (SMART)</td>
<td></td>
<td>Stepped wedge cluster RCT</td>
<td>Case studies</td>
<td></td>
</tr>
<tr>
<td>Real world</td>
<td></td>
<td>Delayed-start RCT</td>
<td>Cross-sectional</td>
<td></td>
</tr>
<tr>
<td>A/B testing</td>
<td></td>
<td>Adaptive clinical trial</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Non-randomised experimental trial</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Section 8: Randomised Controlled Trials**

### 8.1 RANDOMISED CONTROLLED TRIALS

Randomised controlled trials (RCTs) are usually considered the most rigorous method of determining the efficacy of a new intervention. One strength of the RCT lies in the process of randomisation. Generally, in an RCT, similar participants are randomly assigned to 1 of 2 or more groups; the intervention group who will receive the intervention, and a comparison group (control) who will receive standard treatment or a placebo. These groups are then followed prospectively to assess the effectiveness of the intervention compared with the standard or placebo treatment. Ideally, participants and investigators are ‘blind’ to treatment allocation and each group is managed identically, with the only difference being the intervention received. Measurement and reporting are also important. Hence, outcomes should be measured in an objective and reliable way, (e.g. not subject to clinical opinion), with results reported for all of the patients in the study (*intention to treat*) and factors such as numbers not completing the study clearly identified.

There are 2 main types of RCT study designs, pragmatic and explanatory. Pragmatic RCTs are designed to determine the relative efficacy and safety of an intervention as it would occur in routine clinical practice. In contrast, explanatory trials aim to determine whether a clinical intervention is effective under optimal circumstances (e.g. in a carefully chosen, treatment-adherent patient population). With a well-designed and rigorously conducted explanatory RCT, the possibility that an observed association is due to factors other than the intervention should be minimal. However, sometimes there are concerns that the results of explanatory trials will not generalise to the NHS setting. These can arise from several factors including:

- Characteristics of the patients (e.g. an RCT may have included younger, fitter patients, with a better prognosis than those typically presenting in the NHS;
- It may have been set in sites which adopted a care pathway which was not representative of most NHS settings (e.g. had intensive diagnostic testing and follow-up regimens);
- Outcome measures may have been intermediate outcomes so impact on outcomes relevant to patients are still uncertain;
- Follow-up may have been too short for some adverse events and outcomes such as mortality to be observed.

Pragmatic RCTs seek to reduce the difference between the conduct of the trial and how the intervention will be used in clinical practice. However, the associated variation in the patient populations and clinical practice may increase the range of responses and make it more difficult to demonstrate statistically significant changes.
Randomisation is not sufficient to ensure a good RCT. Having an adequate sample size to be able to determine the number of participants needed to detect a clinically relevant treatment effect is essential before commencing an RCT. Underpowered trials, which may fail to achieve statistically significant results should be avoided, whilst including too many people is expensive and may not be ethical. Usually a sample size is calculated using 4 parameters (population size and its variance, confidence intervals and margin of error) and statisticians can provide advice on this calculation.

There are several other factors, which must be considered to develop a high quality RCT study design. These include selection criteria of participants; randomisation methods; blinding of participants, investigators and assessors (particularly those reporting results) to the interventions (where possible - with apps such blinding may be logistically impossible to achieve); adoption of unbiased end-points; length of follow-up; reporting of interventions; loss to follow-up; reasons for drop-outs and results and no conflict of interests.

Assessing the quality of RCT design is particularly important to those who want to use their results to inform decisions. The results from a high-quality RCT are more likely to be a valid estimate of the true effect size than those from a poor quality study. Checklists are available to evaluate the quality of an individual study. Section 21 provides a series of checklists which can be used to assess the quality of the main study designs reviewed in this document. These also provide a useful tool to guide the reporting of the different clinical study types.

Strengths of RCTs include:

- A well designed RCT provides the strongest evidence of any study design about the comparative efficacy and safety of a given intervention;
- Randomisation enables investigators to control for known and unknown factors hence should minimise the risk that results are confounded;
- Enables robust statistical analysis including confidence in results;
- Enables blinding which also reduces the risk of bias;
- Can measure disease incidence and multiple outcomes;
- Can provide data to inform an economic evaluation.

Weaknesses of RCTs include:

- Need to recruit large numbers of similar people to provide sample sizes sufficient to power the study (i.e. large enough to detect statistically significant changes in outcomes). This may require study to be conducted across several settings which brings additional problems around ensuring conformity with a set protocol and requiring multiple data collection systems;
- Inefficient for rare diseases or diseases with a delayed outcome;
- Challenging and costly to set-up, run and report large trials;
- The difficulty in blinding patients, investigators or both to some interventions for ethical and practical reasons;
- Can raise ethical issues on whether it is appropriate to deny some patients active treatment;
• Problems with generalising their results to the NHS setting and cost (unless it is a pragmatic RCT).

References

Good reference sources are:


8.2 RCT STUDY DESIGNS

There are several forms of study design, which can be grouped as RCTs. The major study designs used are listed from the most to least common:

• **Parallel group** – each participant is randomly assigned to a group, and all the participants in the group receive or do not receive the intervention;
• **Crossover** – each study participant has both therapies, i.e. is randomised to intervention A first, and at the crossover point they then start intervention B;
• **Cluster** – pre-existing groups of participants (e.g. villages, schools) are randomly selected to receive (or not receive) an intervention;
• **Factorial** – each participant is randomly assigned to a group that receives a particular combination of features e.g. group A receives a weight loss app plus counselling, group B receives a drug, and group C receives the app plus counselling and the drug. The interventions could be expanded to include say dietician support with randomisation across more groups;
• **Stepped wedge cluster** – involves random and sequential crossover of clusters from control to intervention until all clusters are exposed, e.g. 1 cluster will be randomly assigned to receive the intervention at week 3, a different cluster will be randomly assigned to receive the intervention at week 6, and this continues until all clusters are receiving the intervention;
• **Delayed-start designs** – randomises participants to the intervention arm or the control in the first phase. In the second phase, the control group switch to the intervention arm and the intervention group remains on it;

• **Adaptive designs** – clinical trials can be designed with adaptive features enabling changes in design or analyses to be guided by examination of the accumulated data at an interim point in the trial. This can reduce the duration of studies or enable them to demonstrate an effect using fewer patients than other RCT designs.

### 8.3 PARALLEL GROUP RCT

A parallel group RCT is a simple and commonly used clinical design to compare an intervention to a relevant comparator such as standard practice. Participants are randomly assigned to the intervention or the comparator. Parallel group designs do not require the same number of subjects in each group, although usually similar numbers are observed.

The strengths of this design include:

• It is the most common form of RCT design, so members of the research community are familiar with its conduct and analysis;

• Advice on methodological issues such as the appropriate number of participants required to be included for a given study design to show statistically significant results (sample size) is readily available for this study design;

• Transparent and validated reporting criteria to assist readers to understand its design, conduct, analysis and interpretation are available [CONSORT (Consolidated Standards of Reporting Trials) statement] [20].

Weaknesses of the parallel design include:

• Need to recruit large numbers of similar people to provide sample sizes sufficient to power the study (i.e. large enough to detect statistically significant changes in outcomes). This may require the study to be conducted across several settings which brings additional problems around ensuring conformity with a set protocol and requiring multiple data collection systems. More efficient randomised study designs may be available;

• With apps, blinding patients and investigators may be impossible for practical reasons and hence more complex design studies (such as crossover) may be better able to address this limitation;

• Problems with generalising their results to the NHS setting (unless it is a pragmatic RCT).
An example of a study design for a two-arm parallel group RCT for an app is now described [21].

Aim – The aim of the trial was to assess whether a mobile phone intervention could help treatment-seeking patients with an alcohol use disorder to improve adherence to Naltrexone. Naltrexone is a medication for treating alcohol dependence but is not highly utilised due to concerns with non-adherence.

Trial design – Participants were randomised to parallel groups, either the intervention (n=37) or control (n=29) using a computerised algorithm that balanced groups according to self-reported drinking goal.

Participants received a one-month supply of Naltrexone (50mg per day) and an Android smartphone with 8 weeks of unlimited service. They then were instructed to return to the clinic at weeks 4 and 8.

Intervention – The intervention was called AGATE-Rx (Adaptive, Goal-directed Adherence Tracking and Enhancement). AGATE reminded participants to take the study medication via SMS text messages. Additionally, a hyperlink in the message launched the web browser to assess adherence. Over time, the frequency of medication reminders was adapted according to self-reported adherence performance. In addition to the AGATE-Rx adherence reminders/assessments, participants in the intervention group received a Smartphone Alcohol and Side Effects Diary (SASED), consisting of daily SMS prompts with hyperlinks to assess side effects and alcohol use and craving.

Comparator – Participants in the control group also received SASED prompts/assessments, but not adherence reminders/assessments.

Outcomes – The primary outcome was adherence, defined as at least 80% of prescribed doses taken during the 8-week trial.

An example of a study design for a 2-arm parallel group RCT for a wearable device is now described [22].

Aim – The aim of the trial was to compare the effectiveness of a wearable pulsed electromagnetic fields (PEMF) device in the management of pain to placebo.

Trial design – Participants were randomised to parallel groups, either the intervention (n=33) or control (n=33)

Intervention – The intervention consisted of 12 hourly daily treatment for 1 month.

Participants – 66 patients with knee osteoarthritis.
Outcomes - The primary outcome measure was the reduction in pain intensity, secondary outcomes included quality of life assessment pressure pain threshold and changes in intake of medication.

References


8.4 CROSSOVER RCTS

Crossover designs enable participants to act as their own controls, facilitating comparisons between and within groups. Participants receive a randomised sequence of the intervention(s) and comparator over periods of time, crossing over to an alternative intervention as part of the sequence. For example, at the start of the study, every patient is assigned to a sequence (e.g. AB vs. BA). All participants receive the same number of treatments but in different sequences. Hence one can examine how participant characteristics influence response to treatment. Crossover designs can be complex with multiple interventions over several time periods.

Strengths of crossover designs include:

- Crossover design removes between-participant variation as each participant serves as its own control, thereby reducing error variance and reducing sample size needed;
- It requires fewer participants than a parallel study for an equal number of treatment comparisons, because each experimental unit (i.e. participant) can be used several times. This is an economical use of resources;
- Participants can indicate preferences for one treatment versus another, because participants receive multiple treatments in a single crossover study.
Weaknesses of crossover designs include:

- They can only be used if there are no lasting participant benefits from the active treatment;
- It shortens the time period for the treatment comparisons;
- Participant responses to each intervention are likely to be correlated (e.g. a single participant’s response to treatment A is correlated with that participant’s response to treatment B), which makes analyses complex.

An example of a study design using crossover design for an app is now described [24].

Aim – The aim was to evaluate a mobile health-based remote medication adherence measurement system (mAMS) in elderly patients with increased cardiovascular risk treated for diabetes, high cholesterol and hypertension.

Trial design – Drug adherence for 4 drugs was investigated in a controlled randomised doctor-blinded study with crossover design. All participants underwent an initial run-in control phase with standard medication blisters and handwritten medication intake diaries. Participants were randomised. Group 1 used the mAMS and group 2 was a control. After 20 weeks, all participants crossed over for another 20 weeks, after which all participants used the control followed by a 4 week phase with mAMS.

Intervention – The mAMS was designed to monitor a patient’s behaviour in taking the prescribed medication and hence to measure their adherence and improve this, if found to be insufficient. In order to track the dosage and timing of medication intake, electronic medication blisters were used as add-ons to standard medication blisters.

Comparator – The control phase consisted of standard medication blisters, routine care and handwritten medication intake diaries.

Participants – 150 participants with a risk for cardiovascular conditions.

Outcomes – Primary outcome was medication intake rate in both 20 week phases. Secondary outcomes were comparison of laboratory data (e.g. fasting blood glucose concentration) and blood pressure.

An example of a study design using crossover design for a wearable device is now described [25].

Aim – To assess the performance and safety of an integrated bihormonal artificial pancreas system consisting of 1 wearable device and 2 wireless glucose sensor transmitters during short-term daily use at home.

Trial design – Patients started treatment with an artificial pancreas with a day and night in the clinical research centre, followed by 3 days at home. The control period consisted of 4 days
of insulin pump therapy at home with blinded continuous glucose monitoring for data collection. Days 2-4 were predefined as the analysis period.

Intervention – The system was an integrated bihormonal artificial pancreas.

Comparator – Standard insulin pump therapy at home.

Participants – 10 adult patients with type 1 diabetes.

Outcomes – Median glucose level.

References


8.5 CLUSTER RCTS

Cluster RCTs are studies in which participants are grouped (clustered) on the basis of a setting or geographic area (e.g. GP practices, homes for the elderly, schools), and then randomised as a cluster to either the intervention or comparator arm. The number of clusters and participants in each can vary widely but often there are a reasonably small number of clusters, each with a large number of participants, (for example, 10 GP practices each with 1,500 patients). Ideally the clusters will be populated by people with similar socioeconomic and other characteristics. The design can often indicate some issues of scaling up interventions, in addition to measuring effect size. They are frequently used to evaluate service re-design, public health education or other such interventions where the intervention alters the behaviour of an organisation which then impacts on participants rather than directly impacts participants.
Strengths include:

- The cost per participant is usually much less than an RCT enrolling individuals, particularly if individuals are not consented but rather the administrator for the cluster consents to undertake the intervention and provide the outcomes for analysis;
- Data collection usually uses existing routine systems, also saving costs;
- The results are also usually relevant to ‘real world’ decisions so generalisability is less of an issue;
- It provides an alternative methodology for assessing the comparative effectiveness of interventions where randomisation is inappropriate or impossible at the individual level.

Weaknesses include:

- Cluster trials are complex to design, size, implement and analyse and hence require experienced researchers;
- The cluster RCT design requires a greater number of participants compared to individual RCT designs because responses from participants within a cluster tend to be correlated;
- Consenting participants can be time-consuming and if there is a high refusal rate or drop-out, there is potential for bias, as those completing the study differ from those who did not, thereby breaking the benefit of randomisation;
- Biases can occur at different stages with cluster-randomised trials and may be difficult to detect. Particular biases to consider include: (i) recruitment bias; (ii) baseline imbalance; (iii) different drop-out rates across clusters causing loss of randomisation; (iv) incorrect analysis; and (v) comparability of findings with individually randomised trials.

An example of a study design using a cluster RCT with an app is now described [27].

Aim – To assess the efficacy of an application designed for smartphones as a supportive element that could assist younger smokers in their efforts to stop smoking.

Trial design – Primary care centres were randomised and stratified according to urban/rural location and number of participating professionals, and then assigned 1:1 to the intervention or control group using a computer programme. Participants or professionals were not blinded. All participants were given a smoking cessation plan.

Intervention – A smartphone app enabled users to record adherence to a smoking cessation plan and educated them on problems of smoking and provided alternatives to the user at times when he or she particularly wanted to smoke.

Comparator – Participants received the same smoking cessation plan.

Participants – 22 primary care centres participated.
Outcomes – Abstinence at 12 months confirmed by exhaled-air carbon monoxide concentration.

An example of a study design using a cluster RCT with a wearable device is now described [28].

Aim – To (i) test the hypothesis that modifying patterns of painful lumbo-pelvic movement using motion-sensor biofeedback in people with low back pain would lead to reduced pain and activity limitation compared with guidelines-based care, and (ii) facilitate sample size calculations for a fully powered trial.

Trial design – 8 clinics undertook a cluster-randomised, placebo-controlled pilot trial comparing 2 groups of patients seeking medical or physiotherapy primary care for sub-acute and chronic back pain. It was powered for longitudinal analysis, but not for adjusted single-time point comparisons. Both groups received 6–8 treatment sessions. Outcomes were measured 7 times during 10-weeks of treatment and at 12, 26 and 52 week follow-up.

Intervention – Participants received modification of movement patterns augmented by motion-sensor movement biofeedback plus guidelines-based medical or physiotherapy care.

Comparator – Participants received a placebo wearing the motion-sensors without biofeedback plus guidelines-based medical or physiotherapy care.

Participants – 8 primary care clinics and 58 individuals participated.

Outcomes – Primary outcomes were self-reported pain intensity and activity limitation (Roland Morris Disability Questionnaire and Patient Specific Functional Scale scales).

References


8.6 FACTORIAL RCTS

A factorial trial is where 2 or more interventions are simultaneously compared with a control group in the same trial. Participants are randomised to receive neither intervention (control), 1 or the other intervention, or both. An example of such a trial is an evaluation of an app to manage depression, plus a video. Participants could be randomly assigned to standard care, standard care plus app, standard care plus video, and standard care plus app plus video. Most factorial trials have 2 ‘factors’ in this way, each of which has 2 levels; these are called 2×2 factorial trials. Occasionally 3×2 trials may be conducted or trials that investigate 3, 4, or more interventions simultaneously. In most factorial trials the intention is to achieve ‘2 trials for the price of 1’, and the assumption is made that the effects of the different active interventions are independent, that is, there is no interaction (synergy). Occasionally a trial may be carried out specifically to investigate whether there is an interaction between 2 treatments. This approach has similarities to some of the engineering studies (e.g Sequential, Multiple Assignment, Randomised Trials (SMART) and Multiphase Optimisation Strategy (MOST) study designs) where different components within a product are tested to optimise the product.

Strengths include:

- Factorial trials enable investigators to evaluate the separate effects of each intervention and the benefits of receiving both (or more) interventions together;
- The design enables interactions between interventions to be investigated efficiently. An alternative design is a parallel 3-arm trial, or even a parallel 4-arm trial, but the sample size required for these is much larger.
Weaknesses include:

- They are only suitable for interventions that can be used in conjunction with one another;
- They can be more complex to design, analyse and interpret than a 2-arm RCT.

An example of a factorial study design with an app is now described [30].

Aim – The aim was to investigate the adherence and effectiveness of a human-supported and automated-supported web-based intervention for people with mild to moderate depressive symptomatology, and study the impact of 4 persuasive technology components.

Trial design – In order to identify active components of the intervention the screening phase of the MOST was used. Some components are part of the intervention itself (e.g. the programme content), and others are related to its delivery (e.g. feedback by an expert or through an automated system). The purpose of the fractional factorial design was to identify the components that were effective.

A balanced 8-arm fractional factorial design was used, which compared the human-supported web-based intervention with the automated-supported web-based intervention and screened for the effects of 4 other components. These consisted of: text message coaching (present or absent), interaction (high or low), tailoring of success stories (high or low), and personalisation (high or low). The design was balanced, meaning that each level of each component is present in half of the intervention arms (e.g. 4 intervention arms include automated support, and the other 4 include human support).

Participants were recruited through advertisements in Dutch newspapers and those eligible were automatically randomised. Participants were not blinded but did not have in-depth knowledge of the other arms. Post-intervention questionnaires were emailed to participants after 3 months, and follow-up questionnaires were emailed after 6 months.

Intervention – The intervention called ‘Living to the full’ was web-based and included 9 chronological lessons based on acceptance and commitment therapy. Each module included tests, online and offline exercises, and metaphors. Participants were instructed to complete 1 lesson per week, but had 12 weeks in total to complete the 9 lessons.
The source of support was either human or automated; all participants received feedback messages at the end of each lesson. Participants receiving human support had the opportunity to ask their counsellor questions, whereas those receiving automated support received 1 additional instant feedback message per lesson. Feedback messages were provided within the application and participants received automated email messages when feedback was received. Participants in the text messages arms received 3 text messages per week containing motivational, mindfulness and content-related information. Participants in the high interaction arms received additional multimedia and interactive material. The intervention contained a success story of each of the lessons; participants in the high tailored arms had more relevant stories based on 4 of the following: their gender, age, marital status, daily activity, most prominent symptom, reason for participating. The high personalisation arms had content that was adaptable e.g. the system would show the motto and picture selected by the participant, and gave them the option of selecting the ‘top 5’ aspects from the course that they found most important.

Participants – 239 participants were randomised to 1 of 8 intervention arms.

Outcomes – Depressive symptoms were measured with a depression scale and anxiety symptoms were measured with the Hospital Anxiety and Depression Scale at baseline, post-intervention and follow-up. Task enjoyment, involvement, trust and satisfaction were also measured post-intervention.

An example of a factorial study design with a wearable device is now described [31].

Aim – The aim was to examine whether behavioural economics-based interventions for increasing steps/day may improve cardiorespiratory fitness.

Trial design – Inactive adults were randomised to an adaptive/static goal intervention with either immediate micro-incentives or delayed incentives in a 4-month factorial randomised controlled trial to improve steps/day. Additionally, all 4 groups received daily text-message prompts-to-action and Fitbit activity monitors. Between- and within-group differences were examined using linear mixed models, adjusted for sex, age, race/ethnicity, and smoking history.

Intervention and comparator- 4 groups were included being adaptive physical activity + immediate financial micro-incentives; adaptive physical activity + delayed financial micro; static physical activity goals (10,000 steps/day) + immediate financial micro-incentives static physical activity goals (10,000 steps/day) + delayed financial micro-incentives.

Participants – 81 inactive adults, age = 41.8 and BMI = 34.

Outcomes – Maximum volume of oxygen peak, blood pressure and vascular stiffness.
References


8.7 STEPPED WEDGE CLUSTER RCT

The stepped wedge cluster randomised controlled trial is a relatively new study design and is an alternative to parallel cluster trial designs. The design includes an initial period in which no clusters are exposed to the intervention. Subsequently, at defined intervals, 1 or more clusters are randomised to the intervention. This process continues until all clusters have crossed over to the intervention. At the end of the study there will be a period when all clusters are exposed. Data collection continues throughout the study, so that each cluster contributes observations under both control and intervention observation periods. It is a pragmatic study design and recommended if there is heterogeneity across the clusters or the clusters are large.

Strengths include:

- Implementation can be phased with sequential roll-out;
- It can detect the impact of the intervention over time;
- It increases statistical power (within and between comparisons);
- It can study heterogeneity between clusters and study the impact of different settings on effectiveness.

Weaknesses include:

- Complexity of the design requires careful design of the clusters, e.g. the randomisation pattern at each step, the length of each step and the logistics of communicating with lots of participants;
- Sample size and power calculations are not available for all design types;
- Data collection and analyses can be complex.

An example of a study design for a web-based system using a stepped wedge RCT is now described [33].
Aim – The aim was to determine whether moving from a paper-based referrals system to a web-based system with automated tracking features would result in improved access to speciality medical services.

Trial Design – 11 primary care and 25 speciality sites were randomly staggered, 1 at a time over several months using computer generated ordering. This allowed for system implementation and training. Sites served as controls before receiving the intervention and intervention sites after receiving it.

Intervention – The intervention replaced faxing of referrals by encrypted, internet-based delivery of requests to speciality offices, with prompts and reminders.

Comparator – Usual care consisted of creating and faxing orders in primary care and making the appointment.

Outcomes – The main outcome was percentage of scheduled appointments. Other outcomes included median time to appointment.

No wearable study using a stepped wedge RCT was identified.

References


8.8 DELAYED-START RCTS

In a delayed-start design (also known as “stepped-wedge designs” if the randomisation into arms is by group) participants are randomised to the intervention arm or the control in the first phase. This phase should be sufficiently long to allow the clinical or other impact of the app to be captured. In the second phase, participants in the control arm are switched to receive the intervention and followed for the same amount of time. The intervention arm remain on the intervention. At the end of the trial any differences in outcomes should be due to the impact of intervention. These forms of trial design are of particular use for interventions used to manage people with diseases or conditions that progress slowly, such as rheumatoid arthritis, Alzheimer’s or Parkinson’s disease, where the intervention is aimed at, say increasing self-management, and thereby reducing symptoms or modifying the disease.
Figure 3.1 has a graphical representation of a delayed-start design. As may be seen in Phase 1, participants in the control arm receive the placebo (P) and those in the intervention arm the intervention (A). Both sets of participants received the active treatment in phase 2. The 2 arms deteriorated over time in respect of their Parkinson’s symptoms, however, the participants in the treatment arm to a lesser extent. The disease-modifying effect is demonstrated at T4 as the difference in final symptom score between the 2 groups of participants.

Strengths:

- Enables the disease-modifying effect on symptoms to be separated from other (short-term) effects;
- All participants receive the active intervention unlike in a standard RCT, thereby overcoming some of the ethical concerns.

Weaknesses:

- Sufficient understanding of the clinical progression of the disease / condition is required in advance;
- The second phase is open-label (un-blinded) therefore may potentially introduce biases;
- There may be issues recruiting participants who would be willing to receive the control treatment for the first phase of the trial, particularly if they judge the intervention is ‘better’ than the control;
- These trials are only applicable to early / mild stage of the disease and therefore difficult to generalise to advanced disease;
- These trials are prone to high drop-out rates, thereby also increasing potential biases.

---

An example of a delayed-start RCT for an app is now described [37].

Aim - To test whether a data-driven feedback module presented as a mobile app was able to improve self-management for type I diabetes patients.

Trial design and intervention - A stepped-wedge (delayed-start design) where patients were randomised into 2 groups. In the 5 week run-in period both groups of patients entered their glycated haemoglobin levels into the mobile app at least five times per day. After this run-in period the intervention group received feedback from an additional module on the app consisting of 24-hour blood glucose readings, trends (long and short term), situation matching (to enable the participant to decide if and when to inject insulin), and additional information. The intervention group were presented with this additional module for an initial 8 weeks, after which the control arm also received the module for a further 10 weeks. A total of 30 participants were recruited and 15 randomised to each group.

Outcomes – The outcomes of interest were the high and low blood glucose counts (“out of range” values).

An example of a delayed-start RCT for a wearable device is now described [38].

Aim – To test whether a fully automated mobile health (mHealth) intervention with tracking and texting components would increase physical activity.

Trial design – Sequential randomization was used to individually evaluate the tracking and texting components of the intervention. After establishing baseline activity during a blinded run-in (week 1), in phase I (weeks 2 to 3) participants were randomized 2:1 to unblinded versus blinded tracking. The activity tracker itself did not show activity information, but continuously transmitted it in all participants. The activity data were only visible to those who were unblinded. In phase II (weeks 4 to 5), unblinded participants were randomised 1:1 to ‘smart texts’ versus no texts. Smart texts were automated, personalized, smartphone-delivered coaching messages informed by real-time activity and other factors.

Intervention – Participants used their own smartphones, activity tracking was performed using a wearable, display. Unblinded patients could continuously view their daily step count, activity time, and aerobic activity time through smartphone, web interfaces and the app also provided a history tab allowing review of data from previous days. Each participant was a patient of a study physician; the latter sent texts aiming to leverage the physician-patient relationship. Smart texts were grouped as positive reinforcement messages, sent when a participant was on track to attain or had already attained his or her daily goal, and booster messages, to motivate individuals when they were not tracking to surpass their step goal. Texts were sent 3 times/day (morning, mid-day, and evening), with exact times customized to the participant’s usual wake time, lunch time, and beginning of evening leisure time.

Comparator – Participants were blinded to activity data and later to texts.
Participants – 48 smartphone users at an ambulatory cardiology centre.

Outcomes – Primary outcome was mean change in daily step count from baseline through phase I and II. Secondary outcomes were attainment of the prescribed 10 000 steps/day goal, changes in total daily activity time and aerobic time and participant satisfaction.

Reference:


8.9 ADAPTIVE DESIGNS

Adaptive designs, often using Bayesian methods are being used increasingly in RCTs. With a Bayesian approach, current knowledge about the model parameters is expressed by placing a probability distribution on the parameters, called the "prior distribution". When new data become available, this information is combined with the prior to produce an updated probability distribution called the "posterior distribution," on which all Bayesian inference is based.

Adaptive designs – as the name implies – allow aspects of trials to be modified as they are being run on the basis of the information accumulated during the trial. Data analyses are conducted at specified points in the trial to enable adaptations to be made, ensuring that the validity and integrity of the trial are kept intact. There are 3 major forms of adaptive trial designs:

• Prospective, where planned modifications based on data accumulating within a study are permitted; examples include removing inferior treatment groups, stopping trial due to safety issues or including adaptive randomisation of participants;
• Concurrent, where modifications are made to the trial as it progresses, for instance modifying inclusion/exclusion criteria, changing hypotheses or study end-points;
• Retrospective, which are usually applied to the statistical analysis plan after data collection, but before any un-blinding.
Trial data are commonly analysed using standard, frequentist statistics, however, Bayesian analysis methods are particularly suited to analysing data from adaptive designs. Bayesian methods, as opposed to frequentist methods, use all the data accumulated during the trial and prior information (from literature reviews or previous trials) in combination for the analysis and conclusions.

Velengtas and colleagues [11] suggest that adaptive designs are useful in that they enable new information to inform the study design, thereby enabling potential reductions in sample size, trial duration and costs. These designs may be of particular use in apps where large volumes of data can often be obtained quickly. This study design enables designers to adapt studies as information is received.

Strengths:

- Flexibility to redesign clinical trials;
- Reduction in length and cost of trials;
- Potential reduction in time participants are exposed to the intervention;
- Modification to trial assumptions and parameters in a timely manner.

Weaknesses:

- Potential to introduce bias, particularly if blinding is compromised as a result;
- Potential difficulty in interpreting data;
- Potential increase in complexity of statistical analyses and interpretation.

References


Section 9: Observational Studies of Interventions

With observational studies of interventions, the investigator does not randomise participants to specific treatments within a tightly defined care pathway. Rather the investigator simply records what happens to participants. Some observational studies are conducted because it may not be ethical to randomise people to an intervention (e.g. if the efficacy of the intervention relative to standard care has already been proven), or randomisation may not be feasible, or the sample size may not be sufficient to power an RCT.

Some observational studies, especially if they have a control, are described as quasi-experimental studies. These may take the form of pre- and post-intervention designs or compare an intervention and control group. The key features are participants are not randomised and care is in a real world setting. Non-randomisation means that important confounding variables may not be effectively measured or controlled. Under these circumstances it becomes difficult to interpret the outcomes in light of the intervention applied. Despite this criticism, quasi-experimental studies remain popular, particularly due to the relative ease and low costs with which they may be implemented.

There are 3 main categories of interventional observational studies: cohort studies, cross-sectional studies and case-control studies. Case-control studies are useful to identify risk factors and rare events but seldom can demonstrate efficacy and hence unlikely to be relevant to app developers. They are not considered further in this document.

Checklists to assess the quality of each study design are available at Section 1.

9.1 COHORT STUDIES

In a cohort study, a group of individuals receive an intervention and is followed over time to determine how the intervention affects specified outcomes. The outcomes for this group can be compared with those from another group that has not received the intervention.

Cohort studies may be prospective or retrospective. In a prospective cohort study, the subjects are followed up for a period and the outcomes of interest are recorded. Neither group should have the outcome of interest at baseline. In a retrospective cohort study, both the intervention and outcome have already occurred at the outset of the study. A retrospective cohort study is usually easier to conduct, often using medical records or interviews/questionnaires but is more susceptible to bias due to incomplete records and incorrect recall of past events. These cohort designs can also inform estimates of either the incidence of an outcome or the relative risk of an outcome, based on exposure.
Follow-up of study participants in a cohort study can be challenging, requiring staff to monitor participants’ health outcomes and measure their individual exposure to the intervention under review over time. Incomplete data collection is a major problem.

Cohort studies are observational and not as reliable as RCTs, since the 2 groups may differ in ways other than in the variable under study. However, they can demonstrate causality and examine changes in multiple outcomes over time.

Strengths:

- Multiple outcomes can be measured for any 1 risk factor under review (e.g. smoking);
- One study can look at multiple risk factors (e.g. smoking and alcohol consumption);
- Good for measuring rare events, for example, among different geographical areas;
- Demonstrate direction of causality;
- Can measure incidence and prevalence.

Weaknesses:

- Can be administratively complex leading to relatively high costs;
- Prone to bias due to loss to follow-up;
- Prone to confounding as other factors cannot be controlled for;
- Being in the study may alter participant's behaviour (the Hawthorne Effect).

An example of a cohort study with an app is now described [40].

Aim – The aim was to determine whether electronic healthcare technology could be integrated into current therapy for obesity.

Trial Design – 124 children and adolescents who were admitted to hospital to undertake a structured treatment and teaching programme (STTP) entered the trial. The STTP consisted of 28 therapeutic sessions.

Intervention – The electronic health technology consisted of a mobile motion sensor integrated into a mobile phone with a digital camera. This analysed the type, intensity and duration of physical activity using algorithms. Eating habits were documented using the digital camera on the mobile phone to take pictures of all meals, which were then analysed by educators in nutrition.

Comparator – Physical activity and eating habits were assessed using self-reported questionnaires.

Outcomes – Difference between self-reported and motion sensor reported activity and energy intake.

An example of a cohort study with a wearable device is now described [41].
Aim – To test efficacy of a bio-absorbable antibacterial envelope in reducing implant related infections in high risk patients.

Trial design – This single-centre retrospective cohort study compared the prevalence of cardiac implantable electronic device infections among subjects with >2 risk factors using data collected by individual chart reviews. Results were presented for a propensity score-matched cohort of 316 recipients of either envelope or 316 controls.

Intervention – A recently developed bio-absorbable antibacterial envelope (TYRXTM-A)

Comparator – 1 control group did not receive an envelope, a second received a nonabsorbable envelope.

Participants – Patients with >2 risk factors treated with the bio-absorbable antibacterial envelope (n=135), a nonabsorbable envelope (n=353), and controls who did not receive an envelope (n=636).

Outcomes – Cardiac implantable electronic device infections.

References


9.2 CASE SERIES OR CASE REPORTS

A case series reports on a group of participants with an outcome of interest. No control group is involved. These may enrol consecutive or non-consecutive participants depending on the inclusion criteria. They can be useful in identifying unexpected consequences to an intervention and hence are used in pharmacovigilance/safety studies. The authors are often familiar with cases, so can provide an in-depth understanding of the cases. Further, they can be useful if other research designs are not possible, being relatively cheap to conduct and usually rely on routine data collection. Reporting may be for all participants or just a selection chosen by the authors. These can indicate trends but do not provide evidence of a causal relationship.
In addition to the impossibility of establishing cause-effect relationship, such studies can be vulnerable to selection bias i.e. they do not represent the wider population and hence their results may not generalise to other settings. They may also over-emphasise the unusual in that unless a case is somewhat unusual it will not be published.

A case report reports on a single participant with an outcome of interest.

Strengths:

- They allow a lot of detail to be collected that would not normally be easily obtained by other research designs. The data collected are normally a lot richer and of greater depth than can be found through other experimental designs;
- Case studies tend to be conducted on rare cases where large samples of similar participants are not available;
- Within the case study, scientific experiments can be conducted;
- Case studies can help experimenters adapt ideas and produce novel hypotheses which can be used for later testing;
- They can collect quantitative or qualitative data.

Weaknesses:

- Data are often not systematically collected;
- Data are not collected consistently across the source documents;
- Often it is not known whether the data can be generalised to the wider population, which limits its relevance and usefulness;
- Smaller case studies are usually reported by 1 researcher which can lead to bias;
- One can seldom draw a definite cause/effect from case studies.

An example of a case series with an app is now described [42].

Aim: To determine whether a parental training programme delivered via an iPad was effective in improving the language, social interaction and learning skills in children with fragile X-syndrome (FXS).

Trial Design: A case series design.

Intervention: A 16-week long programme consisting of weekly 1 hour long parental guidance sessions followed by an unguided parent-delivered training programme delivered via the iPad over 3 hours per week.

Comparator: None.

Participants: 4 children, mean age 6 years, with FXS.
Outcomes: These were language gains, social skills acquisition, progress in academic learning and positive behavioural outcomes. Parents rated their satisfaction with the educational programme in terms of their child’s outcomes.

An example of a case series with an wearable device is now described [43].

Aim – To evaluate the effectiveness of sustained acoustic medicine to alleviate pain and improve function in people with elbow or Achilles tendinopathy.

Trial design – Patients were trained to self-apply a wearable, long-duration, low-intensity ultrasonic device on their affected body part at home for 4 hours a day, at least 5 times per week over 6 weeks. Pain measurements were recorded before, during, and after daily intervention. Function of the injured limb was assessed biweekly using dynamometry.

Intervention – Wearable ultrasonic device.

Comparator – None.

Participants – 25 patients with clinician-diagnosed tendinopathy of the elbow or Achilles tendon.

Outcomes – Change in pain score and in function.

References


9.3 CROSS-SECTIONAL STUDIES

A cross-sectional study examines the relationship between a health outcome and, say, the use of an app, in a defined population at a single point in time or over a short period of time (e.g. a calendar year). These are purely descriptive and cannot determine causation. Cross-sectional studies are often used to assess the prevalence of disease; e.g. asthma among 12-14 year olds, or depression in a selected age group across a geographic area.
Strengths:

- Relatively quick and easy to conduct (no long periods of follow-up);
- Data on all variables are only collected once;
- Able to measure prevalence for all factors under investigation;
- Multiple outcomes and risk factors can be studied;
- Good for public health data on factors such as prevalence of disease or other health related characteristics;
- Good for descriptive analyses and for generating hypotheses.

Weaknesses:

- The time relationship between risk factors and events or disease may be difficult to determine;
- Not suitable for studying rare diseases or diseases with a short duration;
- Unable to measure incidence;
- Associations identified may be difficult to interpret;
- Susceptible to bias due to low response and misclassification due to recall bias.

An example of a cross-sectional study using an app is now described [45]. This app is also 1 of the case studies.

Aim – To determine the user perception of an oral health app, and to provide a basis for future research and development of oral health app technology.

Trial Design – This study employed a cross-sectional qualitative user perception questionnaire. The Brush DJ app is aimed at encouraging better oral health. The app links into the user’s playlist on their mobile device, and presents reminders and advice about oral health to the user. In this study, respondents who had downloaded and used the Brush DJ app were presented with a user experience questionnaire consisting of 9 multiple choice questions and 1 open-ended question about users’ attitudes towards the app.

Intervention – Brush DJ app.

Comparator – None.

Participants - 189 respondents completed the survey.

Outcomes – Outcomes included users’ responses to the questionnaires to identify perceived benefits and whether they would recommend the app.

An example of a cross-sectional study using a wearable device is now described [46]

Aim – To investigate how vigorous-intensity activity varies with age.
Trial design – Cross-sectional data were extracted from the International Children's Accelerometry Database and linear regression were used to investigate age-related patterns in vigorous-intensity activity; other variables included age, exposure, adjustments for monitor wear-time and study.

Intervention – Accelerometry.

Comparator – None.

Participants – 24,025 participants from 20 studies in 10 countries obtained 2008-2010.

Outcomes – Moderate-intensity activity and vigorous-intensity activity.

References


9.4 NEW-USER DESIGNS

New-user designs are recommended where the rate of intervention-related outcomes is dependent on the time that has elapsed since treatment was initiated. This is particularly relevant where the risk of the underlying disease or condition is affected by the treatment. New-user designs are non-experimental (i.e. observational studies of interventions) which attempt to eliminate the biases introduced by "prevalent users" (e.g. long-term users of a particular medicine) by excluding those participants, and only including participants with no prior use of the treatment or those who have had a minimum period ("washout period") of non-use. An example from the literature is women undergoing hormone-replacement therapy (HRT) [47] where prevalent users, i.e. those women who have been receiving long-term HRT, introduce potential biases through the underestimation of events early on in their treatment, and through the inability to control for disease risk factors affected by the treatment.

A variant of this design is to apply a different restriction (i.e. not new-user) to participants included in the trial, to seek to reduce bias. This is called a restricted design. For example, if an app is used with people who have a spectrum of disorders, as in many mental health apps, a developer may restrict the trial to a specific group, such as those with a bipolar condition.
This should enable a valid comparator to be identified. The developer may want to narrow the group down further (e.g. age, gender, disease severity) to make comparator groups as similar as possible. These methods attempt to reduce confounder variables as much as possible.

No studies of apps or wearable devices using new-user or modified designs have been identified.

References


Section 10: Safety Studies

Evidence on the safety of a new app or indeed any intervention can come from several sources including RCTs, specific adverse event case reports, safety studies and meta-analyses. The pharmaceutical sector often uses registries, observational safety studies, including cohort, case-control, nested case-control, and case-crossover studies. These usually have a longer duration than RCTs. A comparator is desirable, ideally, it will be an existing comparator or, if not possible, a historical comparator can be adopted. Alternatively, some RCTs continue to follow-up participants for adverse events after conclusion of the formal RCT; this may be for several years.

The FDA has issued guidance on ‘Best Practices for Conducting and Reporting Pharmacoepidemiologic Safety Studies Using Electronic Healthcare Data’, which provides a good overview of the factors to consider when designing studies to inform safety.

Demonstrating the safety of a device which has a potential for harm, for example, apps or wearables that affect medication use such as dose calculators, or an ingestible sensors informing adherence decisions are essential. As a minimum all algorithms and calculations must be documented, with steps taken to ensure inputs are complete and have face validity, and that outputs are calculated correctly, with recommendations for action consistent with best clinical practice. Ideally, the developer will engage with clinical experts and patients throughout the development process.

The Mersey Burns app provides a good example of the developer using a well-designed safety study to show the app intervention was not inferior to current practice (i.e. it was at least equivalent to current practice). The developer, who was a clinician, devised a robust safety study which was not onerous to conduct or report. The evidence described in the case study, together with technical documentation (e.g. design, component specifications, risk analysis and compliance with necessary standards) enabled the app to be certified as a class 1 medical device by the Medicines & Healthcare products Regulatory Agency (MHRA).

A more traditional approach using registry data is presented in [48]. This study reported findings on safety events such as ‘received inappropriate defibrillator therapy’ from The Prospective Registry of Patients Using the Wearable Defibrillator (WEARIT-II) Registry. The registry provides real-world data on the wearable cardioverter defibrillator as a strategy during a period of risk stratification and had enrolled 2,000 patients prescribed the defibrillator.

All developers seeking to connect to the NHS IT system will be required to meet the standards laid down by Health and Social Care Information Centre (HSCIC). HSCIC also provides a clinical safety assurance process for NHS IT systems. It undertakes a hazard assessment to identify a risk level and those at that level will be required to provide further safety assurance as required under ISB0129 Clinical Risk Management: its Application in the Manufacture of Health IT Systems (see Section 23:).
References

Section 11: Accuracy and Diagnostic Studies

Accuracy studies are studies which measure the extent of agreement between the outcome of a new test and the reference standard. A reference standard is considered to be the best available method for measuring the same outcome as that measured under the new test. The conduct of accuracy tests is not well documented. However, there are several sources advising on the conduct and reporting of diagnostic studies, (for example, FDA: Guidance for Industry and FDA Staff Statistical Guidance on Reporting Results from Studies Evaluating Diagnostic Tests). The principles which apply to compare different diagnostic tests are now explained. These generalise to the wider group of accuracy tests.

Diagnostic tests are usually used to identify and/or confirm a disease condition or change in a condition for individuals. For a new diagnostic test, specific criteria are used to assess its efficacy against the ‘reference standard’ including:

- Sensitivity;
- Specificity;
- Likelihood ratio;
- Receiver-Operating Characteristic (ROC) Curves.

The sensitivity of a new test is estimated as the proportion of subjects with the target condition in whom the test is positive. Similarly, the specificity of the test is estimated as the proportion of subjects without the target condition in whom the test is negative. A likelihood ratio measures the likelihood that a given test result would be expected in a patient with the target disorder compared to the likelihood that that same result would be expected in a patient without the target disorder. In a ROC curve the true positive rate (sensitivity) is plotted against the false positive rate (100-specificity) for different cut-off points of a parameter. Each point on the ROC curve represents a sensitivity/specificity pair corresponding to a particular decision threshold. The area under the ROC curve is a measure of how well a parameter can distinguish between 2 diagnostic groups (diseased/normal).

The conduct of such studies to avoid bias is also slightly different. For example, diagnostic tests perform differently in different populations, so the setting is important. The choice of the test to compare the app intervention to (the reference standard) is also important. Ideally, it will be the gold standard, or may be current practice, and in either case should have well-defined performance statistics. The test accuracy of the app intervention is based upon a comparison between its results and those of the reference standard. Any discrepancy is assumed to arise from error in the app intervention. Hence, selection of the reference standard is critical to the validity of a test accuracy study.
Advantages include:

- When performed correctly, this design does inform the accuracy of the test for the population tested.

Disadvantages include:

- The assumption of 100% accuracy for the reference standard rarely holds true in practice. This represents a fundamental flaw in the test accuracy study design, since the app intervention can never be deemed to perform better than the reference standard, and its value may therefore be underestimated;
- The statistics used to compare the tests are specialised and may not be widely understood outside a small community.

Suggested checklists to evaluate the quality of diagnostic studies are available at Section 16.

An example of a diagnostic study is now described [49].

Aim: To develop a mobile phone app to help community health workers (CHW) screen for the risk of cardiovascular disease (CVD).

Trial Design: A non-observational study where CHWs were randomised into 2 groups: paper-based risk assessment and the mobile app risk assessment tool to screen patients for CVD risk.

Intervention: Mobile app risk assessment tool.


Participants: 24 CHWs were recruited: 14 into the paper-based group, 10 into the mobile app group and they screened 537 people.

Outcomes: Outcomes included time taken to train staff to use the assessment tool, screening time and errors.

Several studies compare the accuracy of competing wearable devices for activity tracking and heart rate monitoring. A good example is a comparison of the accuracy and precision of 17 wearable devices reported by [50].

An example of a diagnostic accuracy study using a wearable device to detect breast cancer is now provided. [51].

Aim – To evaluate circadian patterns of breast temperature as a method of breast cancer screening and to compare circadian pattern behavior analysis to mammography and physical examination for breast cancer screening.
Trial design – Gold standard was pathology; results from the 3 screening types were compared to pathology results. Two statistical approaches were adopted, giving slightly different results.

Intervention – Each woman wore a portable breast monitor which consisted of a microprocessor with 16 thermistors that automatically recorded temperatures every 5 minutes for 48 hours.

Comparator – Mammography and physical examination.

Participants – 138 women who had been scheduled for open-breast biopsies as a result of physical examination and mammography.

Outcomes – True and false positives.

References


Section 12: Economic Studies alongside RCTs

12.1 CONDUCTING AN ECONOMIC EVALUATION WITHIN A CLINICAL TRIAL

Healthcare decision makers increasingly require evidence of value, measured by the change in resources, their costs and change in participant outcomes resulting from adopting an intervention, compared to current practice. One potential source of this information is the clinical trial establishing safety and effectiveness. The additional steps involved in conducting an economic evaluation within a clinical trial include:

- Quantification of cost of intervention and comparator;
- Quantification of impact of the intervention compared to current practice on health and social care costs;
- Assessment of whether the costs and benefits differ among user groups;
- Using an indicator to compare the differences in costs, savings and patient benefits.

Collecting resource and cost data raises several design issues including procedures for data collection for economic endpoint data and generalisability. Pragmatic clinical trials collect data under real world conditions and hence the outcomes are more relevant to current practice. If the safety and effectiveness trial adopts a protocol which differs from the current care pathway then this limits the usefulness of the economic data collected.

The health and social care resources used by patients such as hospital admissions, consultations, treatment of any side effects or complications, and care packages required on discharge from hospital are normally recorded for each patient over the trial follow-up. The categories of resource use that are included in the study will be determined by the perspective of the analysis - whether it is confined to the health and social care system or includes broader societal costs. In England and Wales, the National Institute for Health and Care Excellence (NICE) recommends adopting a NHS and personal social services perspective. The societal perspective also considers care provided by other sectors of the economy, such as costs incurred by patients, informal care provided by family and friends, and productivity losses from morbidity and premature death.

Use of many resources can normally be recorded on trial case report forms, which increases the data collection, validation and analyses tasks within the trial. Some studies may use additional information sources such as electronic/paper medical records, patient questionnaires and diaries. Recording events outside hospital in primary and social care can be problematic. Often this information is collected using patient or carer questionnaires, which gives rise to recall bias and often may raise issues of low response rates- another potential bias.
12.2 CHALLENGES WITH APPS INTERVENTIONS

Normally economic evaluations will be conducted from the perspective of the NHS and social care sector. Thus the relevant costs of the app intervention are those that these services will incur. Initial purchase costs, fees for use in subsequent years plus any maintenance or warranty costs should be included in the economic appraisal. If the app intervention links to the NHS or social care system then any related costs which will be incurred by the services should also be included.

Discovery and development costs incurred by the developer are not relevant to the economic evaluation.

12.3 FORMS OF ECONOMIC ANALYSES

Economic analysis is the systematic appraisal of costs and benefits of interventions, normally undertaken to determine the relative economic efficiency of these. There are 5 forms:

- **Cost consequences analysis:**
  o An economic analysis that presents the impact of a new intervention, compared to current practice, on:
    i. Usage of individual healthcare resources and their related costs or savings (depending on whether resource use increases or decreases with the technology);
    ii. Health outcomes (usually several measures are used, such as change in disease progression, impact on subsequent clinical events, life expectancy and health-related quality of life) for patients;
    iii. Users of the technology, (for example, benefits to the operators in terms of ease of use or increased security, which are additional to time savings).

- **Cost-minimization analysis:**
  o An economic analysis in which the clinical and safety consequences of competing interventions are the same and in which only inputs, that is, costs are taken into consideration. The aim is to decide the least costly way of achieving the same outcome.

- **Cost-effectiveness analysis:**
  o An economic analysis in which the costs and consequences of alternative interventions are expressed as cost per unit of health outcome (e.g. cost per stroke avoided). This form is used to determine technical efficiency; i.e., comparison of costs and consequences of competing interventions for a given patient group within a given budget.
• **Cost-utility analysis:**
  o An economic analysis in which interventions, which produce different consequences, in terms of both quantity and quality of life, are expressed as ‘utilities’. These are measures, which comprise both length of life and subjective levels of well-being. The best known utility measure is the ‘quality adjusted life year’ or QALY. In this case, competing interventions are compared in terms of cost per quality adjusted life year (cost per QALY).

• **Cost-benefit analysis:**
  o An economic analysis in which all costs and consequences of an intervention are expressed in the monetary terms. This form can be used to compare different types of interventions across different sectors and serving different groups.

### 12.4 COST CONSEQUENCES ANALYSIS

Cost consequences analysis (CCA) is anticipated to be the form of analysis most relevant to app developers. It is used by NICE when judging the value for money of medical technologies and public health interventions.

A principle underpinning CCA is that there are different types of benefits, some of which cannot be measured using change in a patient’s quality of life and life expectancy, (for example, improved usability and security.) With CCA, individual benefits are listed separately from costs, often using a tabular format for both. Benefits and costs are not combined into a single ratio. A CCA format is thus transparent and readily understandable. It is also comprehensive, enabling all benefits and resources to be considered. For example, if a new app offers benefits to operators these can always be evaluated under CCA but not with other forms of economic analysis, unless the benefit can be quantified as a cost saving.

A checklist to evaluate the quality of economic evaluations is available at Section 21:

An example of a protocol for an RCT including cost-effectiveness analysis for an app is now described [55].

**Aim** - The aim was to determine whether in-person follow-up, i.e. attendance at clinics, can be circumvented through the regular monitoring of patients post-surgery through a mobile app, and whether this approach is cost-effective.

**Trial Design** - A pragmatic, single-centre, open, 2-arm parallel group superiority randomised trial. The mobile app will be used by patients to enter patient-reported outcome measures and surgery-specific questions and surgical site photographs. These will be submitted daily for the first 2 weeks and at lesser intervals for another 2 weeks.

**Intervention** - A post-operative mobile app collecting post-surgical data over a 30-day period.

**Comparator** - In-person attendance at clinic for post-surgical follow-up.
Participants - Female patients aged between 18 and 70 who had undergone breast reconstruction surgery.

Outcomes -. The primary outcome is number of visits to clinicians associated with the surgery in the first 30-days after surgery; secondary outcomes include other interactions with healthcare professionals regarding surgery (phone calls, emails), complication rates and patient satisfaction. Societal and healthcare system costs will be collected to assess the cost-effectiveness of this approach.

An example of an RCT including cost-effectiveness analysis for a wearable is now described [56]

Aim – To determine the efficacy and cost-effectiveness of the wearable cardioverter defibrillator in terminating tachyarrhythmias in patients at high risk for sudden cardiac death.

Trial design – This study was conducted on all wearable cardioverter defibrillator patients who were at high risk for sudden cardiac death but did not meet the implantation-eligibility criteria for implantable cardioverter-defibrillator. A Markov model of the cost, survival, and incremental cost-effectiveness of the wearable defibrillator compared with usual care was developed.

Intervention – Wearable cardioverter defibrillator.

Comparator – Usual care.

Participants – 66 patients.

Outcomes – Cost per patient and mortality.

References


Qualitative research aims to provide an in-depth understanding of people’s experiences, perspectives and histories in the context of their personal circumstances or settings. It employs a variety of methods, including interviews; focus groups; observations; documentary analysis. The data captured are often rich and complex. Qualitative research can also be used to observe how groups perceive apps or use (or do not use) them in practice.

Qualitative studies are often exploratory in nature and seek to generate novel insights using inductive (starting with observations and developing hypotheses) rather than deductive (starting with extant hypotheses and testing them with observations) reasoning. This section looks at some of the more common design methods and provides examples of each.

Checklists to address the quality of qualitative studies are available at Section 16.

References


National Centre for Research Methods has several videos and studies on qualitative methods. More detail at http://www.ncrm.ac.uk/

13.1 FOCUS GROUPS

Focus groups are a form of group interview, which aim to capture the level of respondents' attitudes, feelings, beliefs, experiences and reactions to a specific topic. These may not be available when using methods, such as observation or interviewing, being more likely to be revealed via the social gathering and the interaction created in a focus group. There are several guides on the 3 main components of focus groups (composition, conduct and analysis) and some authors cover the role and conduct of ‘virtual focus groups’ using the internet.
The main advantages of focus groups are:

- They are useful to obtain detailed information about personal and group feelings, perceptions and opinions;
- They can save time and money compared to individual interviews;
- They can provide a broader range of information;
- They can be a useful first step to identify potential problem areas allowing a more in-depth analysis to be planned;
- Recruitment can be based on certain discriminating criteria e.g. sex, race and age.

The limitations are:

- The researcher has less control over the data produced and little control over the interaction other than generally keeping participants focussed on the topic;
- Recruitment and assembling the focus group can be difficult (e.g. finding a date and time for busy healthcare professionals, or resistance from people who are less articulate or confident);
- Not useful for gathering statistics as it only allows you to analyse people’s views but not the number of people holding that view;
- The wrong mix within a group can cause problems and may not work effectively;
- They may not be useful if people shy away from discussing the topic because it is embarrassing or sensitive;
- They tend to provide a ‘group view’ and not a collection of individual views;
- It may be difficult to compare information between the groups and draw out high-level findings.

The ChatHealth poster provides a good example of the use of focus groups to inform app development. When developing the Chathealth app, Leicestershire NHS Trust used focus groups in 4 schools (n= 33) to identify if a safe and secure messaging service with a school nurse might make it easier for young people to access healthcare advice. The findings informed the decision to commission the app, its content and functionality.

An example for a wearable is described in Belsi et al [61]

Aim – To identify the impact the use of wearable technology could have in patients with osteoarthritis in terms of communication with healthcare providers and patients’ empowerment to manage their condition.

Trial design – Qualitative study using focus groups with patients with osteoarthritis; data from patients’ responses were analysed using Framework Methodology.

Intervention – Wearable technology.

Comparator – None.
Participants – 21 patients with knee osteoarthritis in 4 focus groups.

Outcomes – Patient responses.

References


13.2 QUALITATIVE RESEARCH INTERVIEWS AND SURVEYS

Qualitative research interviews are often differentiated as unstructured, semi-structured and structured [65].

No interview is completely unstructured but those conducted as guided conversations are usually categorised as unstructured. Usually the interviewer asks questions of the interviewee that emerge over time and are informed by responses to earlier questions. This format is suitable for individual in-depth interviews where the interviewee has a wealth of knowledge and experience which the interviewer wants to elicit to inform their research. Selecting the sample of interviewees is often iterative using a purposeful sampling frame to identify people with the relevant knowledge and experience.

Often the interviewer is seeking to capture behaviours or experiences or to probe on sensitive issues. Usually the interviewer records the interview as throughout s/he must listen carefully and test understanding, ideally building up a relationship with the interviewee throughout the process. The interview ends when no new topics are left to explore. Data analysis usually adopts the same themes as those explored in the interview and the record is a fulsome record of the conversation.
Semi-structured interviews are more organised and have as a starting point a set of predetermined open-ended questions. Responses to these lead the interviewer to ask more questions, related to the responses received. These can be conducted with an individual or in groups (see focus groups) and are usually of a shorter duration than the unstructured interview. These are well-sited to exploring complex issues with a range of people who may have different perspectives on the research question.

A structured interview is commonly employed in survey research to obtain quantitative data. The aim of this approach is to ensure that each interviewee is presented with exactly the same questions in the same order. This is important for minimising the impact of context effects, where the answers given to a survey question can depend on the nature of preceding questions. Though context effects can never be avoided, it is often desirable to hold them constant across all respondents. It is important that the interviewer plays a neutral role and does not insert his or her opinion in the interview.

The choice of answers to the questions is often fixed (closed) in advance, though open-ended questions can also be included within a structured interview. Fixed questions ensure that answers can be reliably aggregated and that comparisons can be made with confidence between sample subgroups or between different survey periods. A common technique is to adopt a Likert Scale, which enables respondents to adopt a rating scale (often 1 to 5 with 3 being neutral).

These are best suited when all parties have a well-developed understanding of the topic, avoiding the need for the interviewer to explain issues and hence potentially introduce bias. This enables researchers to create a highly structured interview guide or questionnaire that provides respondents with relevant, meaningful and appropriate response categories to choose from for each question.

Structured interviews are, therefore, best used when the topic is familiar to interviewees or following use of observational and other less structured interviewing approaches that provide the researcher with adequate understanding of a topic to construct meaningful and relevant close-ended questions.

Structured interviews can be conducted, face-to-face or by telephone, and be web-based or paper based. Data collections methods depend on the format of the study design.
Advantages:

- Useful for studying a limited number of cases in depth;
- Useful for describing complex issues;
- Provides individual case information;
- Provides understanding and description of people’s personal experiences of the issue;
- Can provide rich detail, and analyses the factors that influence variation in responses e.g. socio-economic factors, or by organisation, or by setting;
- Responses can often vividly demonstrate the depth of feeling on factors;
- Detailed and rich data can be gathered in a relatively easy and inexpensive way;
- Allows interviewer to establish rapport with the respondent and clarify questions;
- Provides an opportunity to build or strengthen relationships with important stakeholders;
- Can raise awareness, interest and enthusiasm around an issue;

Disadvantages:

- Selecting the "right" key informants may be difficult if there is a lot of heterogeneity across the population of interest;
- Can be challenging to schedule interviews with busy or hard to reach groups;
- Difficult to generalize results to the larger population unless interviewing many key informants.

An example of a structured interview app study is now described. This study also demonstrates how the design can inform usability tests [66].

Aim – To report on the usability and acceptability, to healthcare professionals and non-professionals, of Painometer – a smartphone app that helps users to assess pain intensity.

Trial Design – Non-professionals reported the maximum pain intensity they experienced in the last 3 months. All participants were asked to ‘think aloud’ while using the app, and field notes were taken with the mistakes in both groups being recorded. A qualitative usability testing approach with a semi structured interview was conducted. All participants were asked 7 open-ended questions about ease of use, efficiency, and their satisfaction using Painometer. They also filled in a questionnaire about their previous use of technology. The trial was conducted in 2 cycles. After analysing the results of the first cycle, minor changes to improve the usability of the app were made, and it was then retested in cycle 2 using the same methodology, but with different participants.

Intervention – Painometer is a smartphone app that contains 4 pain intensity scales.

Comparator – None.
Participants – 19 healthcare professionals and 14 non-professionals for the first cycle, and 15 healthcare professionals and 16 non-professionals participated in the second cycle.

Outcomes – Outcomes included ease of use, major and minor usage errors and acceptability.

An example of a structured interview wearable study conducted by Nabhani, S., et al. (2015) [67] is now described.

Aim – To explore chronic obstructive pulmonary disease patients’ perceptions on vest comfort, ease of wear and handling, willingness to use and concerns.

Trial design – A structured interview was designed to explore chronic obstructive pulmonary disease patients’ perceptions on vest comfort, ease of wear and handling, willingness to use and concerns. Interviews took place at clinics in England and Netherlands where patients were provided with a suitably sized vest to try on followed by the structured interviews.

Intervention – A vest containing wearable sensors to inform the early diagnosis of exacerbations and disease deterioration allowing for early intervention.

Comparator – None.

Participants – 15 patients in England and now interviews underway in Netherlands.

Outcomes Patient perceptions on vest comfort, ease of wear and handling, willingness to use and concerns.

Responses to structured questions may be obtained using a questionnaire or survey rather than an interview. A survey generally refers to the selection of a sample of people from a pre-determined population, followed by collection of data from this group, usually using a standardised approach with mainly closed questions. Responses are analysed to draw inferences about the wider population. The output is a snapshot of responses at a specific time point. Surveys can be repeated across different groups to provide cross-sectional analyses or over time to provide trend analyses. Most surveys are now online rather than distributed face to face at say a clinic, or posted out to recipients.

Advantages:

- Can be conducted in relatively short time and at fairly low cost;
- Researcher does not need to be skilled interviewing techniques;
- Questions and response can be designed to simplify data analysis;
- Results are less easily influenced by the researcher’s personal biases.
Disadvantages

- Some respondents may not answer accurately or honestly;
- Respondents may not answer all questions, leading to bias;
- Questions may be poorly constructed leading to bias;
- Researchers do not gain an understanding of the behaviours informing responses;
- The information might not generalise to other people or other settings;
- It might have lower credibility with some audiences such as commissioners;
- Low response rates limit validity and may cause bias if the sub-group responding differs from the wider group sampled.

An example of a study using qualitative surveys to inform app development is now described [68].

Aim – To investigate the feasibility of assisting autonomous physical training of independently living older adults with Active Lifestyle – an IT based system aiming to improve people’s balance and strength. Additionally, to investigate the adherence of participants to training plans and the effectiveness of the in-built motivation instruments.

Trial Design – Participants were taught how to use the iPad and app and how to perform balance and strength exercises supported by the app. Participants undertook a 2-week home based balance and strength training plan at the end of which they answered questionnaires regarding the perceived usefulness, usability, visual attractiveness, and the effectiveness of the motivation instruments of the app.

Intervention – Active Lifestyle app, which gives spoken, written and video instructions to guide users through exercises and provides feedback at the end of the workout.

Comparator – None.

Participants – 13 participants.

Outcomes – Outcomes included adherence to the intervention, effectiveness of the app and the apps motivation instruments, and usability of the app.

An example of a study using qualitative surveys to inform development of wearable is now described [69].

Aim – Describe first use of a digital health feedback system by practicing pharmacists to establish evidence-based blood pressure management recommendations.

Trial design – A digital health feedback system was used to identify patients where medication was not effective in controlling blood pressure. The information from the system was combined with information from qualitative surveys that used root cause analysis to identify reason for resistance in each patient.
Intervention – Pharmacists used a digital health feedback system which passively captured and shared information about medication-taking using an ingestible sensor, and daily patterns of rest, activity, and exercise using a wearable patch that incorporated an accelerometer.

Comparator – Current medicine service.

Participants – 15 pharmacies and 39 patients in the Isle of Wight.

Outcomes – Pharmaceutical resistance, inadequate medication use, need for additional pharmacological treatment, additional adherence support, pharmacist and patient experiences.

References


13.3 OBSERVATIONAL METHODS

Observational methods are those that analyse what people do and say by observing behaviours and listening to conversations when people are in their natural settings [71]. Participants are virtually always aware of the researcher’s presence which may itself influence behaviour. Covert observation, except in rare circumstances, is unethical. Conducting such studies can be time consuming because the individuals or groups to be observed may be reluctant to give consent unless there is rapport established and some sharing of objectives.
Once the study is underway the researcher usually keeps a record of the research process including of their reaction to events and this introduces a subjective aspect. Observations may also be videoed and conversations taped. The observations usually inform hypothesis which can then be tested in the setting. The observations are usually very detailed and numerous. These are often coded into different themes and should be analysed systematically which can be time consuming and interpretation may not be straightforward.

There are several potential difficulties in conducting observational studies well. Once such is that the observer begins to take an active role which means the focus on observation is lost. Moreover, the observer cannot record everything and systematic biases may arise in terms of what is recorded and these may be difficult to detect.

The methodology is most suited to organisational studies where other techniques such as interviews or surveys may not be able to capture the dynamics of the interplay between e.g. patients, doctors and nurses on a ward.

An alternative, but related approach, to study organisations is focussed ethnography [72]. Conventional ethnography, the systematic study of people and cultures was originally used by anthropologists to study, over a long time frame, the values, behaviours, beliefs and language of a group of people, often living in a remote part of the world. Focussed ethnography is an approach to study these features in a setting which is part of one’s own society.

The approach usually comprises of relatively short field visits to well-defined and quite narrow settings. The field work usually comprises intensive data gathering using audiovisual technologies, in addition to observation. The recordings are then subject to extensive and rapid data analysis. The digital data are often shared with multiple listeners and viewers to get a wide range of interpretations quickly. This reduces the risk of bias through reliance on the interpretation of a single researcher or small group. Examples include workplace studies and those related to the adoption of digital technology. Thus focused ethnography could be used to investigate the usability of an app to enhance communications in the workplace.

Advantages

- Enables observer to record what actually happens and not what people describe will happen;
- Avoids recall and other biases in responses;
- May identify behaviours that people did not realise occurred;
- Can be very powerful at identifying key drivers in the decision making process and in explaining variation in outcomes.
Disadvantages

- Settings which refuse to participate may have different characteristics to those that do and hence results cannot be generalised across settings;
- Retaining objectivity can be challenging;
- Retaining objectivity, together with systematic data collection and analyses can be challenging;
- Traditional observational studies can be lengthy to conduct.

This methodology was used by Stack, E., et al. (2016) [73] in a study to understand falls.

Aim – To investigate which in-home sensors, in which locations, could gather useful data about fall risk.

Trial design – Over 6 weeks, observers monitored 5 people at high risk of falling, at home, making field notes about falls (prior events and concerns) and recording movement with video, Kinect, and wearable sensors.

Intervention – Sensors, observations and cameras to monitor falls behaviour

Comparator – None.

Participants – 5 people with moderate or severe Parkinson's disease.

Outcomes – Common activities leading to balance loss and acceptable of methodology to participants.

References

Section 14: Mixed Methods

Mixed methods research has been defined as "integrating quantitative and qualitative data collection and analysis in a single study or a program of inquiry". The field of mixed methods has only been widely accepted for the last decade, though researchers have long been using multiple methods, just not calling them "mixed." Mixed methods research takes advantage of using multiple ways to explore a research problem. In particular, whilst quantitative research can test impact on process and outcome measures, qualitative research can help understand the quantitative data produced and unpack important behavioural aspects such as why, where, how, for whom, an intervention worked.

Characteristics include:

- Designs combine quantitative and qualitative research designs;
- The designs can adopt prospective or retrospective perspectives;
- Studies can combine experimental, pragmatic or observational designs;
- Sample sizes and populations will vary based on methods used;
- Data collection can involve any technique available to researchers;
- Each study will be analysed separately but the findings will be synthesised;
- Syntheses can be formal, using techniques such as meta-analysis for quantitative studies, or less formal, looking for consistency in trends and outcomes across a range of studies;
- Interpretation of the findings is continual and can these influence future studies in the research strategy.

The main reason for adopting a mixed methods design is to overcome the limitations of a single design. The richness of this design is it can enable a through exploration of a research topic. It can go beyond looking at changes in outcomes to what works, why, for whom and in which setting. Combining data collected from several sources also leads to greater validity in the findings. Usually the results of each study are reported in a separate sections of a report, with the discussion bringing out the salient points together in the report.

The benefits from the approach differ with different sequencing of studies. For example, if qualitative studies are conducted prior to an RCT, their findings may refine the intervention, identify relevant outcomes, explore issues to address in the protocol and indeed inform the research question. Conducting qualitative research during a trial can explore variation in delivering the intervention, identify the contribution of individual components to the observed change and explore the intervention from different perspectives e.g. those of clinicians and patients. After a trial, qualitative research can examine reasons for variations in effectiveness and their implications for the intervention’s development, future trial programme and adoption. It can also inform regulatory submissions and the value-based case.
Strengths include:

- Provides greater breadth of perspectives around a certain issue;
- Statistical results can be augmented by a narrative to improve understanding;
- Combining the 2 study designs can improve overall efficiency with the findings from influencing the design of the second stage. Hence, they enable researcher to generate and test a hypothesis within study plan;
- The range of research questions which can be answered is broader;
- Combining the results from 2 different study designs provides stronger evidence so users can have more confidence in the findings;
- Using 2 methods is also likely to provide a range of findings which together provide more explanation than either method in isolation;
- Can partially overcome a key weakness of RCTs by demonstrating results do generalise to a wider setting;
- Can inform adoption by identifying the factors that influence the use of a new technology in a real world setting;
- Can be presented in an interesting style;
- Some studies are usually set in the real world, and these can help inform the generalisability of results from the more restrictive RCT environment;
- Combining results from multiple data sources, using different study designs, (triangulation) can produce a deeper understanding of issues and increase the confidence in the results.

Disadvantages include:

- Integration adds to the time scale for an RCT and may not be possible within a tight timeframe;
- It is also more expensive;
- Requires multiple data collection techniques;
- Resolving discrepancies in results can be challenging;
- Different study designs generate different grades of evidence so synthesising results requires careful planning;
- There are limited examples of mixed methods studies and few recognised experts so devising valid, unbiased approach can be problematic;
- Ethics committees may prefer RCTs;
- Is likely to require a larger team of researchers who are employed over a longer time period;
- Some of the details of mixed research methodology in respect of combining results are still under development.

There are no checklists for mixed methods studies, rather the quality of each individual study should be judged using a relevant checklist (see Section 21:).
An example of an apps study using mixed methods is now described [74].

**Aim** - To evaluate the effectiveness of a weight management programme presented to participants online and supported through a mobile app.

**Trial Design** - A mixed methods design involving data collected via the web and an app alongside in-depth telephone interviews with participants.

**Intervention** - Participants were provided with fortnightly access to an online weight management intervention and access to a supplemental weight management app on alternate weeks.

**Comparator** – None.

**Participants** - 13 participants with a body mass index of at least 23 who were monitored over 4 weeks.

**Outcomes** - Participants self-reported engagement with the weight management programme daily and usage of the online programme and app was recorded automatically. At 4 weeks participants were interviewed and their experiences of the systems were explored in-depth and analysed using a thematic analysis.

An example of a wearables study using mixed methods is now described [75].

**Aim** – This mixed-methods pilot study examined whether the effectiveness of an online intervention for stress in students could be augmented by the use of prototype wearable sensors.

**Trial design** – Students who were stressed, but not depressed, were allocated to a stress management programme alone (n = 34), with sensors (n = 29), or to no intervention (n = 35). Interventions lasted 4 weeks. Participants were interviewed to gain feedback about the programme and sensors.

**Intervention** – Stress management programme with or without sensors.

**Comparator** – None.

**Participants** – 98 students.

**Outcomes** – Measures of stress, anxious, and depressive symptoms and detailed feedback about the programme, sensors, and biofeedback exercises.
References


76. Research Design Service London, part of the National Institute for Health Research provides guidance on choosing the right study design, which referencing studies on good practice for researchers considering a qualitative component to their RCT.
Section 15: Engineering Informed Study Designs

The group of study designs included in this section are informed by engineering designs, many of which are used to address optimisation problems. Such designs often seek to identify the optimal set of parameters, which maximise the probability that multiple objectives can be met, subject to constraints, which can include behavioural aspects and cost of the product. These designs have been used widely since the 1970s and are increasingly being adopted in healthcare, particularly with behavioural interventions. Such studies are usually set in the real world.

15.1 ECOLOGICAL MOMENTARY ASSESSMENT STUDIES

Ecological momentary assessment (EMA) studies enable participants to self-report on changes in symptoms e.g. pain, moods and behaviour, in near real time in their daily lives, ideally with electronic devices. EMA studies can use the longitudinal data to correlate changes in mood/behaviour to changes in interventions over time. Initially EMA studies were adopted to evaluate new clinical psychopharmacology but their adoption is expanding to use with any intervention where the key outcomes of interest are detailed measurements of mood and behaviour changes.

Often measurement is achieved by the participant using an ecological momentary intervention [EMI]. EMIs are well-suited for use with smartphones and similar devices. A typical EMI is a structured prompt e.g. a person participating in a smoking cessation intervention receives a text message on their mobile phone with tips on managing cravings at user defined times. EMIs also encompass devices which enable ambulatory monitoring of physiological information (e.g. using wearable sensors). The key feature of all EMIs is that the treatment is provided to people during their everyday lives. EMIs normally also facilitate data collection directly to inform EMAs. Real-time data are usually collected at frequent, regular intervals. Methods have been developed to enable random sampling to capture real time data. Older assessment techniques include paper diaries and behavioural observation. EMIs reduce recall bias, and their data usually generalise to the real world. Hence the findings from EMA studies using such data are externally valid.

Research findings support the reliability of EMA study measurements as these correlate with physical changes as recorded in the laboratory.
EMA studies can examine relationships across variables over time and potentially develop algorithms to indicate when there is a step change in behaviour or mood. Behavioural mediators can also be examined and self-management techniques re-enforced.

Findings from EMA studies can also inform and tailor the content and delivery of the EMI. Qualitative feedback (e.g. formal questionnaires, focus groups, and unstructured interviews) is important especially during the early stages of intervention development to optimise the interventions before say an RCT is conducted. This approach offers greatest benefit in hard to reach populations.

Strengths:

- Evidence of efficacy and reliability of EMAs is robust;
- Qualitative methods can be incorporated particularly during the early stages of the intervention development so that deficiencies in the recording systems can be identified and corrected before controlled trials are conducted;
- Enables large quantities of data to be collected cheaply and then the data set can be randomised;
- Study data can be used by clinicians and patients to monitor treatment progress and inform when changes in medication or an intervention are required;
- Efficient study design, data collection and analysis.

Disadvantages:

- Some patients/users may not be open to this form of intervention;
- Some clinicians may not be open to using this form of intervention;
- Finding a suitable control can be difficult;
- Methodology is still developing.

An example of a study using ecological momentary intervention is described in the True Colours case study. A second example for an app is now provided [77].

Aim – To examine the feasibility, usability and ecological validity of a mobile-based Ecological Momentary Assessment (mEMA) app with regard to dietary intake and physical activity among Dutch vocational education students.

Trial Design – Students from 3 schools in the Netherlands were invited to participate in the study, and participatory incentives of ten €20 coupons were randomly distributed to students at the end of the study. Participants who participated for 7 days and filled in the online evaluation doubled their chance of winning. Students from the first 2 schools were allocated to group 1 and were randomly assigned to the dietary intake or physical activity condition. Students from the third school were allocated to group 2. All participants completed an online questionnaire regarding their dietary intake or physical activity. They were then required to download the mEMA app, and use it for 7 days and complete the same short questionnaires regarding their dietary intake or physical activity in the preceding 3.5 hours, 5 times a day.
The app prompted participants to fill in the questionnaire and up to 2 reminders were sent if the diary entry was ‘missed’. Students allocated to the dietary intake condition were asked questions based around their mood, eating behaviours, and availability of food. Students in the physical activity condition were asked questions based around their mood, location and physical activity. Feasibility and usability of the app were evaluated by completing an online evaluation after 7 days of using the app.

Intervention: Ecological Momentary Assessment app.

Participants – 30 students.

Outcomes – Compliance rates as reported by the app, and self-reported compliance as measured by an online questionnaire. Feasibility and usability as assessed by questionnaire. Ecological validity assessed by multiple choice questions.

An example using EMA information from wearables is described in Blaauw et al [78].

Aim – To describe a platform which integrates sensor data from wearables and self-report questionnaire data about cognition, behaviours, and emotions.

Trial design – 2 individuals used a wearable to collect data on heart rate and steps. Psychological variables were collected using administered questionnaires over a pre-defined schedule (EMAs). Statistical analyses were performed on the combined data sets. Data were summarised to a single data point. System outputs were validated by the manual analysis of a domain expert. Physiological data were added to the physical data at the individual level to identify relationships.

Intervention – A platform for researchers that gathers and integrates data from commercially available sensors and service providers into a unified format.

Comparator – None.

Participants – 2 participants.

Outcomes – Validated and accuracy of results, time spent, and ease of use plus some associations between physical and psychological data e.g. time-lagged association between number of steps to humour and to feeling down and influence of cheerfulness on calories expended.
References


15.2 SMALL STUDIES AND N-1 STUDIES

Small study designs involve a very small number of participants (under 10) and focus on observing changes in individual behaviour or event rates over time, as they repeatedly move on or off an intervention. Hence, the design framework is essentially:

1. Study a single person or small group of persons over time;
2. Repeated measurement of the outcome;
3. Sequential application and withdrawal of (or variation in) the intervention.

The USA Agency for Healthcare Research and Quality (AHRQ) has produced a comprehensive User Guide on the ‘Design and Implementation of N-of-1 Trials’, these being studies with just 1 participant. Most of the guidance also applies to small studies [81].

The User Guide identifies key elements to consider in applying the n-of-1 trial methodology to patient-centred outcomes research, describes some of the important complexities of the method and provides checklists. N-of-1 trial designs usually incorporate multiple crossovers, are randomised and often blinded and conducted in a single patient. They have been widely used in psychology, education and social work. The results from similar n-of-1 trials can be combined to enhance their information content.

Their strengths include:

- Patients and clinicians may recognise ineffective interventions quickly, thereby reducing costs and potential exposure to adverse effects;
- They may help engage patients in their own care and thereby improve outcomes;
- They enable research to be conducted as part of clinical practice.
Weaknesses include:

- The benefits to the patient and clinician must be clear from the outset otherwise the commitment to raising the funding and developing the study design will not be there;
- Ethics boards may not accept the design as a valid form of exploratory research;
- Statistical procedures for the design and analysis of n-of-1 trials need further development;
- The interventions must be carefully selected and have no lasting benefit;
- Low external validity;
- Ethical challenges about withdrawing an intervention if it is seen to work.

An example of a small studies design for an app is now described [82].

**Aim** – To examine the potential of mobile phone technologies to increase access to cognitive behavioural therapy techniques and to provide immediate support.

**Trial Design** – This was a longitudinal field study with participants rating their mood on a mobile app (“Mood Map”) 3 times per day over a 1 month period.

**Intervention** – Participants rated their mood on a daily basis (3x) using the Mood Map app, along with a number of other mood scales. Participants were also able to activate mobile therapies (based on cognitive behavioural therapy, CBT) using the app. Participants also attended weekly interviews over the course of the study.

**Comparator** – None

**Participants** – 10 adults participated in the study although results are presented for only 5 participants.

**Outcomes** – The main outcomes were changes in mood scores and qualitative assessment of mood changes.

An example of a small studies design for a wearable, combined with an app is now described [83].

**Aim** – To test if a smart watch can be used to improve medication management and adherence for cancer patients undergoing chemotherapy.

**Trial design** – Data collected by a smart watch from a patient on temperature and symptoms were sent to a consultant’s iPad who could remotely adjust her medication dosage and provide chemotherapy treatments tailor-made to her current situation. App provided medication reminders and feedback on drug adherence

**Intervention** – App and smart watch.
Comparator – None.

Participants – A cancer patient at the Princess Royal University Hospital.

Outcomes – Patient usability and satisfaction, clinician perspective on clinical safety and quality of care, reduced hospital visits and fewer side-effects.

References


15.3 INTERRUPTED TIME SERIES

A time series is a sequence of data points made over a continuous time interval, using successive measurements across that interval and at the same unit of time between each consecutive measurement. Examples include a daily chart of blood pressure or pain scores. These examples relate to an individual’s experience but a time chart can also measure population-level events such as the incidence rates of new diseases or mortality or uptake of new public health interventions.

In the simplest interrupted time series (ITS), a time series repeated observations of a particular event collected over time is divided into 2 segments. The first segment comprises rates of the event before the intervention and the second segment is the rates after the intervention. A statistical analysis called “Segmented regression” is used to measure the changes in the rate of the outcomes and trend in the post-intervention period compared with the pre-intervention period. These studies may also include a comparator arm where no intervention occurred.

The strengths of ITS include:

- The use of trend analysis over time can pick up when the rate of occurrence of an outcome was trending down before the intervention and avoid the mistake of attributing a continuing reduction benefit to the intervention. With a simple pre-post design the change may be wrongly related to the intervention when in fact the change was due to other factors;
- Using rates of an outcome at a population level means confounding by individual-level variables will not introduce serious bias unless it occurred simultaneously with the intervention;
• Expanding the analyses to include other variables can identify any unintended consequences of an intervention. For example, when the absolute level and rate of change in an ‘independent’ variable varies simultaneously with the start of the intervention then this may indicate the intervention is impacting on this variable;
• One can conduct stratified analyses to evaluate the differential impact of an intervention on subpopulations (e.g. by age, sex, race);
• Finally ITS enables the results to be graphed to provide clear information on the shape and direction of change.

Limitations include:
• Require statistical knowledge to undertake the regression analysis (but no more than is required to interpret other study designs);
• The causality assumption is only valid if no other variable or process changes at the same time. For example a change in data collection methods introduced to gather outcome data may change the completeness of the data and introduce bias, or a change in the population will have a similar effect;
• A minimum of 8 observations is required to provide adequate power for the regression analysis;
• Interpretation is enhanced if there is a direct control but this adds to complexity.

An example of a study using ITS with an app is now described [84].

Aim – Examine adherence and ambulatory performance of an app.

Trial design – Ambulatory participants, referred for cardiac event monitoring for cryptogenic stroke or suspected atrial fibrillation were provided a study iPhone preloaded with the app. Participants were asked to use the app 3 times daily and as needed for symptoms (e.g. palpitations) over a 7-day period. Participants completed a usability questionnaire evaluating the app upon study completion.

Intervention – An atrial fibrillation detection app to acquire pulsatile time series recordings and analyse data using novel algorithms.

Comparator – None.

Participants – 16 patients.

Outcomes – Degree of adherence, detection of paroxysmal atrial fibrillation and user satisfaction.

An example of a study using ITS with a wearable is now described new [85].

Aim – To evaluate stride events by creating the time series necessary for this analysis.
Trial design – A method to extract stride cycle events from tri-axial accelerometry signals was validated using data collected from participants. Motion capture data were also collected and served as the comparison method. Statistical tests and mixed models were adopted to summarise results and make comparisons.

Intervention – All participants walked at self-selected comfortable and reduced speeds on a computer-controlled treadmill. Gait accelerometry signals were captured via a tri-axial accelerometer positioned over the L3 segment of the lumbar spine.

Comparator – Motion capture data.

Participants – 10 participants with Parkinson's disease, 11 participants with peripheral neuropathy and 14 healthy controls.

Outcomes – Accuracy of heel and toe-contact events from both feet, mean gait cycle intervals and cycle-to-cycle variability measures.

References

15.4 MULTIPHASE OPTIMISATION STRATEGY (MOST)

Multiphase optimisation strategy (MOST) is an engineering-inspired framework for optimising and evaluating multicomponent behavioural interventions. MOST includes a randomised study for intervention evaluation, but importantly includes other phases of research before the RCT. These earlier phases are aimed at intervention optimisation using selected criteria and to achieve a specific goal (e.g. to develop a cost-effective intervention, an intervention that achieves a specified level of effectiveness, or the shortest time duration for an intervention that achieves a minimum level of effectiveness).

The MOST approach seeks to use highly efficient experimental designs to optimise the various components using feedback informed by the behaviour of participants. A component is any part of an intervention that can reasonably be separated out for study. The design has some similarities with complex factorial design RCT, with the main difference being timing. These are conducted before an RCT is undertaken.

MOST comprises of 3 phases:

- The preparation phase is used to create the theoretical model to identify which intervention components to examine and what the optimisation criterion is (e.g. what is the most effective weight reduction intervention that can be implemented for £200 per person or less. Pilot testing of intervention components is highly recommended;
- During the optimisation phase, the investigator empirically tests options to identify which individual components, in combination, make up the intervention that meets the optimisation criterion. Typically, this phase uses a randomised factorial trial. The outcome from this phase is identification of the components that make up the optimised intervention, together with an understanding of its likely effectiveness. At this point a decision is required on whether there is sufficient promise to warrant undertaking an RCT. If so the evaluation phase starts; if not the investigator may return to phase 1 and pilot new components. Hence the process can be iterative;
- The evaluation phase consists of conducting a standard RCT or a factorial experiment\(^2\), to compare the optimised intervention to a suitable control. If the RCT/experiment indicates that the optimised intervention is not effective, then the investigator may return to the first phase. If it is judged effective then it can be commercialised.

Strengths of MOST designs include:

- They enable testing of an optimised intervention under randomised conditions efficiently. In contrast, conventional RCTs test whether an intervention works or not and these results may inform a revision to the intervention, requiring a further randomised trial. This approach is unlikely to achieve the optimised intervention;

---

\(^2\) Differences between factorial experiments and RCTs are explained at the methodology Centre, Penn State University. [https://methodology.psu.edu/ra/most/factorial](https://methodology.psu.edu/ra/most/factorial)
They enable the effects of the individual components and their combinations to be tested robustly, unlike a conventional RCT which just measures the efficacy of the intervention only as a whole;

Design strategies can also be flexible with few limitations, in theory, on the number of components being tested, ability to use short-term surrogate endpoints to capture behavioural change and enabling a range of design methods including factorial experiments in the evaluation;

Traditional measures of statistical significance can be applied to the results so the analyses do not lose any rigour despite the flexibility.

Limitations include:

Implementing the design can be challenging to ensure each participant receives the appropriate components, at the appropriate times, and outcomes are recorded and analysed. Thus, careful planning, good data collection and well-trained staff are essential;

Decisions may require trade-offs between outcomes, (for example, if a component looks relatively strong in terms of 1 outcome measure but not in another, then the investigator will be forced to choose which outcome is more important);

Funding bodies may be unfamiliar with the framework and unwilling to depart from the gold standard ‘RCT’ approach, particularly given the uncertainties at the start on what the intervention will look like.

The lead researcher for MOST is Linda Collins at Penn State University. The MOST website https://methodology.psu.edu/ra/most contains many resources related to conducting such studies.

An example of an app study using MOST is now described [88].

Aim – To identify active psychosocial and communication components of a web-based smoking-cessation intervention.

Trial Design – A randomised fractional factorial design. There were 5 factors (success story; source; efficacy expectations; outcome expectations; and, exposure) each with 2 levels of depth (high versus low). This gave rise to potentially 32 arms, however, due to the fractional factorial design the total number of arms was reduced by half to 16. Participants were randomised into 1 of these 16 arms.

Intervention – The intervention was a web-based smoking cessation programme delivered over a 6-month period. The primary outcome was self-reported abstinence over a 7-day period recorded at the exit interview at 6 months.

Comparator – None.

Participants – 1,866 were recruited and 1,415 completed the 6-month follow-up interview.
Outcomes – Primary outcome was 7 day point-prevalence abstinence at the 6-month follow-up.

No example of a wearable using MOST was identified.

References


15.5 SEQUENTIAL, MULTIPLE ASSIGNMENT, RANDOMISED TRIALS (SMART)

Sequential Multiple Assignment Randomised Trials (SMART) are used to develop adaptive interventions (AI) to manage complex health disorders such as cancers, addictions and depression. Managing these diseases often entails sequencing, and repeating, different potential approaches depending on participants’ responses. An AI is a sequence of decision rules that specify whether, how, when and based on which measures, to alter the intensity, type or delivery of treatments at pre-defined decision stages in the management care plan.

The aim of a SMART is to develop the optimal adaptive intervention. SMARTs are multistage randomised trial designs. Each participant in a SMART moves through multiple stages of treatment and at each stage is randomised to the next therapy/intervention. The information generated can be used to answer what is the best therapy for non-responders to a specified intervention, or what is the optimal duration of another.

SMART is a form of a factorial experimental design and hence standard data analysis methods can be applied. The effectiveness of the adaptive intervention must still be compared to current practice, ideally using an RCT. Hence, SMART only develops the optimal adaptive strategy, evaluation is still required.

The strengths and weaknesses of SMART trials are similar to those for MOST (MOST trials encompass SMART). These study designs are the only study designs which can reliably tailor AIs to individuals because of their unique randomisation features. The design does not require large sample sizes.
An example of a study using SMART is now described [91]. It is not app or wearable based but 1 of the interventions could be adapted to an app.

Aim – To construct an adaptive intervention that used combined, communication-focused and evidence-based early interventions for preschool children and varied the addition of a speech-generating device.

Trial Design – The study was a longitudinal (repeated outcome measures at baseline and weeks 12, 24 and 36), 3-site SMART design. This SMART included 2 stages of treatment, each for 12 weeks. All children were randomised to the adaptive intervention or adaptive intervention plus a speech generating device. At the end of 12 weeks, children were assessed for early response vs. slow-response to Stage 1 treatment. At the beginning of Stage 2, the subsequent treatments were adapted based on response status. All early responders continued with the same treatment for another 12 weeks. For slow-responders on the adaptive intervention plus a speech generating device, treatment was intensified. Slow-responders in the other arm were re-randomised to intensified adaptive intervention or augmented adaptive intervention plus a speech generating device.

Intervention – Adaptive intervention plus a speech generating device.

Comparator – Adaptive intervention.

Participants – 61 school-aged children with minimal verbal skills.

Outcomes – Total socially communicative utterances, total different word roots and total number of comments outcomes at week 24 and 36, for each of the 3 adaptive interventions.

References

94. The Methodology Center at Penn State University hosts a web page on SMART study with many resources related to using SMART designs. http://www.measuringu.com/blog/qual-methods.php
15.6 REAL WORLD STUDIES

Results from clinical trials may not always be a useful aid for decision-making – particularly if these do not measure the value of the intervention when used in a practical, real life setting. Studies set in real life practice are relatively new and there is no standard design for these. Rather the focus has been on collecting ‘real life’ data to demonstrate effectiveness in a naturalistic environment where factors such as intensity of usage may differ from those in the controlled environment of an RCT.

The most common sources of real life effectiveness data are:

- Databases: Information from the app may be held by the developer and available for further research. The OWise breast cancer app and website provides a good example of this approach. User data are fully anonymised and aggregated in order to conduct health outcomes research. Clinicians can also use the data, which patients record in the app to inform decision-making about treatment options, supporting a personalised treatment approach;
- Patient and population surveys;
- Patient chart reviews: Used to reflect particular insights in patient management;
- Observational data from cohort and case studies;
- Registries: These involve registering and subsequently analysing all patients treated at a particular centre for a particular condition on a continuous basis, using pre-defined fields to capture treatment options, clinical measurements, clinical outcomes and adverse events.

The advantages of real life data include that they:

- Should contain a more representative patient population;
- Can assess clinically relevant endpoints, rather than the often more short-term surrogate endpoints used in clinical trials;
- Can provide longer-term information on safety, resource use and costs;
- Are dynamic so the comparators can be flexed as clinical practice changes.

Disadvantages include:

- The cost of setting up a registry or database are high;
- Coverage may be limited to 1 care setting only, for example, consultant-led care and thus omit activity in the primary care setting;
- Incentives to enter data may be weak so missing or poor quality data can be problematic;
- Definitions for events or the threshold to include events in say a patient record may differ across users;
- Vital information needed to interpret results such as severity of condition may not be captured;
- Unintentional biases may creep into the data, confounding causal analyses.
A real life study combining app and wearable technology is described in Tzallas, A. T., et al. (2014) [95].

Aim – To describe a system for the continuous remote monitoring and management of Parkinson's disease patients and compare to standard care.

Trial design – The patients' physician compared the performance of the system compared to routine examination has been carried out.

Intervention – The system is an intelligent closed-loop system that seamlessly integrates a wide range of wearable sensors constantly monitoring several motor signals of patients. Data acquired are pre-processed by advanced knowledge processing methods, integrated by fusion algorithms to allow health professionals to remotely monitor the overall status of the patients, adjust medication schedules and personalize treatment. The information collected by the sensors (accelerometers and gyroscopes) is processed by several classifiers. Based on information on motor symptoms, together with information derived from tests performed with a virtual reality glove and information about the medication and food intake, a patient specific profile can be built. This is compared to profiles collected recently to assess if status is stable, improving or worsening. Based on that, the system analyses whether a medication change is needed--always under medical supervision--and in this case, information about the medication change proposal is sent to the patient.

Comparator – Standard routine clinical evaluation.

Participants – 12 patients with Parkinson's disease.

Outcomes – Accuracy and acceptability of the system to patients and healthcare professionals, tremor, bradykinesia, freezing of gait and levodopa-induced dyskinesia.

References


15.7 A/B TESTING

Most apps, particularly those that are web-based will have been subject to formal or informal A/B testing. A/B testing offers a randomised experiment with 2 variants, A and B, which are normally called the control and the variation. However, in this context, the control and variant are usually iteration of the intervention rather than a comparison between the intervention and standard practice. User responses to each version are compared to inform on their relative effectiveness. With web applications, the variables are often web pages which essentially are the same except for variations in individual elements like copy text, layouts, images or colours. In such online settings, the goal of A/B testing is to identify changes to web pages that increase or maximize an outcome of interest (e.g. click-through rate for a banner advertisement). Formally the current web page is associated with the null hypothesis.

A/B testing is evidence-based practice: the process includes generating a hypothesis to test, developing the comparators, running the trial to collect data and then analysing these. These tests have randomisation in common with RCTs but their objectives and conduct are very different.

Strengths include:

- It can be performed continuously and the ability to run A/B tests on an on-going basis is designed into many software products;
- The results can be actioned immediately enabling the site to be refreshed regularly;
- The planning and conduct of such tests is reasonably straightforward and not that expensive;
- Existing website users provide real-time randomised focus groups at a fraction of the cost of doing direct recruitment.

Weaknesses include:

- Optimisation requires that only 1 aspect is changed at a time, so sequencing testing over time can be complex;
- Only 1 goal can be maximised at a time;
- Inefficient data collection, as none of the information from a previous test can be reused to draw conclusions about other variables in future.

An example of A/B testing as used to develop RunKeeper app is now described [97].

Aim – RunKeeper is an app that can be used to track activities such as running and cycling. In response to user feedback, the RunKeeper team wanted to expand their fitness tracking features to include other activities such as yoga and weightlifting. This would involve redesigning the apps’ home screen. Hence, the developers decided to conduct A/B testing on the app’s functionality in order to test how the implementation of a new home screen and features might affect user experience.
Trial design – RunKeeper teamed up with mobile A/B testing organisation, Vessel and Localytics, and conducted a 1 month A/B test. RunKeeper randomly assigned Android users to 1 of 2 groups. When Group A users opened the app they saw the apps’ usual start screen which featured a basic form allowing users to input workout information, and the activity would be automatically set to running by default unless changed by the user. When Group B users opened the app an activity was not automatically selected for them, and they saw a new home screen featuring eight different activities to choose from.

Outcomes – Outcomes of interest included percentage increase in non-running activities logged for each group, and percentage of GPS tracked walk, run or bike activities to check whether there was a decline in this type of activity logged with the new interface.

A/B testing is less widely used in developing wearable products and no relevant examples were found.

References


Section 16: Patient-Reported Outcomes and Experiences Studies

Over the last decade, there has been an increasing focus on capturing the patient experience in their interactions with health and social care systems from their perspectives. Patient-reported outcomes (PROs) are individuals’ ratings of their own quality of life, symptoms, treatment effects, functioning and so on, which are elicited from the individuals directly without any input from their family, carers or healthcare professionals. Although frequently used in RCTs and other study designs, PRO measures (PROMs) are also increasingly being incorporated into clinical practice to help monitor patients’ symptoms and experiences, the effectiveness of treatment, as well as enhancing doctor-patient communication. Hence this section examples different types of such measures which can be adopted into any study design.

There are different types of PROMs used to capture different types of information from patients. For example, the SF-36 is a generic PROM, which is not specific to any particular disease or condition and contains questions about patients’ psychological and physical health, social functioning and pain. Other types of PROM are the disease- or condition-specific measures, which are, for example, used to assess symptoms or limitations associated with particular medical conditions such as cancer, chronic kidney disease and osteoarthritis.

Since 2009, the use of PROMs has been mandated by NHS England to determine patient-reported health gains following surgery for 4 procedures: hip and knee replacement, varicose vein and groin hernia repair (https://www.england.nhs.uk/statistics/statistical-areas/proms/). Patients undergoing these procedures complete a condition-specific PROM before surgery and post-surgery. For instance, patients undergoing a hip replacement will complete the Oxford Hip Scale (OHS). This is a condition-specific PROM with specific questions relating to, for instance, pain in the affected joint(s), ability to walk, mobility issues, etc.

Finally, the other main types of PROMs are the preference-based or utility measures, such as the EuroQol-5D (EQ-5D). The EQ-5D is used extensively in clinical trials, and in the NHS PROMs programme. Patients’ responses to questions on the EQ-5D are used to form a profile score which is converted into an index score based on societal preferences for a given health state (rather than the patient’s own health rating). Index scores from the EQ-5D may be used to calculate quality-adjusted life-years (QALYs), which are utilised by bodies such as the National Institute for Health and Care Excellence (NICE) to determine the cost-effectiveness of some medical interventions.
There are published databases available to help select PROMs from the extensive range available. One of these databases is PROQOLID (Patient-Reported Outcome and Quality of Life Instrument Database) (https://eprovide.mapi-trust.org/) which contains information on over 1,000 PRO measures (NB: this is a commercial database and a subscription is required in order to access the information.) Emery and colleagues (2005) have provided an accessible introduction and detailed overview of the information available in this database. Information on PROMs is also available from Oxford University’s Patient-Reported Outcome Measurement Group (http://phi.uhce.ox.ac.uk/newpubs.php). This group has published a number of systematic reviews commissioned by the Department of Health on both generic and condition-specific PROMs for use with a number of commonly occurring conditions, including asthma, chronic kidney disease, heart failure and stroke. Recent developments such as the COSMIN (Consensus-based Standards for the selection of health Measurement Instruments) provide a useful tool for assessing the quality of PROMs in conditions beyond those contained in the Oxford repository (www.cosmin.nl).

Closely related to PROMs are patient-reported experience measures (PREMs) which can be used to monitor quality and safety in healthcare systems, as well contribute to service development and improvement. Taking the example from the NHS PROMs Programme, in addition to completing the EQ-5D and OHS, following surgery patients are also asked to provide information about their experience of the operation (allergy / reaction to drug, urinary problems, bleeding, wound problems), readmission, and subsequent surgery.

Examples of PREMs in other contexts include surveys used to capture older patients’ experiences in their interactions with community care services [100], as well as a PREMs designed to capture patients’ experiences of rheumatoid arthritis and other rheumatic conditions. The latter is currently being used in the UK National Clinical Audit of Rheumatoid and Early Inflammatory Arthritis [101]. Developers interested in designing apps to capture patient experience (as well as outcomes) are referred to a comprehensive guide published by the Health Foundation: (http://www.health.org.uk/sites/default/files/MeasuringPatientExperience.pdf).

Alongside the traditional paper version of PROMs and PREMs, these instruments are now commonly available in electronic format (the so-called, electronic PRO or ePROs), including handheld devices, desktop, laptop and tablet computers, digital pens, and increasingly as apps. The apps have been used for a variety of purposes including health intervention and symptoms capture. Other examples of healthcare apps with PROMs include the “mobile Oncology Symptom tracker” (mOST) used to collect daily self-reported symptoms (e.g. pain, nausea, vomiting, sleep quality) from adolescent and young adult cancer patients [102], self-reported sleep disturbances in breast cancer patients undergoing chemotherapy [103] and TrueColours.
Examples of their use with wearables include:

- Jordan, C. et al. (2015) [104] which used a Patient Reported Outcome Measurement Information System (PROMIS) 10 - Global Health Survey to measure quality of life in postmenopausal women with clinical stage I or II breast cancer who were given a Fitbit Flex and wellness coaching.
- Shinde, A. M. et al. (2016) [105]. Who sought to combine patient reported outcomes around function and activity, with data from wearable biometric sensors and oncologists’ assessments of patients’ performance status, to inform appropriate chemotherapy decisions.
- Viers, B. et al. (2016) [106] who examined patient reported acceptance and perceptions of physical activity monitors within an ambulatory urology setting.

One final point, there is a need for developers to ensure, when developing ePROs, that patients are interpreting and responding to PROs on an app in the same way as on paper, i.e. that there is measurement equivalence. Although this principle applies in general to the development of any ePRO, it is of particular importance when developing apps for this purpose. The International Society for Pharmacoeconomics and Outcomes Research (ISPOR) has published a number of guidance documents regarding ePROs in general, that are of relevance to PROs on apps (www.ispor.org).

In terms of mobile apps, the Mobile App Rating Scale (MARS) has recently been developed to assess the quality of mobile apps for health. The MARS is a 23-item instrument covering areas such as engagement, functionality, aesthetics and information quality [2]. A further development has been the Quality of Experience in mHealth instrument which has been specifically designed for developers in order to evaluate the quality of healthcare apps and wearables. [107].

An example of an apps study using PROMs is now described [108].

Aim – To test whether adding mobile application coaching and patient/provider web portals to community primary care compared with standard diabetes management would reduce glycated haemoglobin levels in patients with type 2 diabetes.

Trial Design – A cluster-randomised clinical trial, randomly assigned 26 primary care practices to 1 of 3 stepped treatment groups (coach-only; coach plus portal access; coach plus portal access plus decision support software) or a control group (usual care).

Intervention – A patient-coaching system (a mobile diabetes management software application and a web portal) and provider clinical decision support.

Comparator – Standard diabetes management.

Participants – 163.
Outcomes – Primary outcome was change in glycated haemoglobin levels. Secondary outcomes included changes in the Patient Health Questionnaire-9, the 9-item version of the Self-Completion Patient Outcome Instrument and the 17-item Diabetes Distress Scale.

An example of a wearables study using a Patient Reported Outcome Measurement is now provided [104].

Aim – To examine impact of regular physical activity to improve mental and physical health during treatment and survivorship for hormone-receptor positive breast cancer patients.

Trial design – This was a prospective study with intervention beginning at the surgical consultation. Visits occurred again post-operatively, 6 months after consultation, and again at 1 year. At each visit, a record of activity, and physical and mental health indicators was made.

Intervention – Each patient was given a Fitbit Flex with the expectation to complete 7,000 steps daily, and weekly wellness coaching.

Comparator – None.

Participants – 19 postmenopausal women with clinical stage I or II breast cancer.

Outcomes – Patient Reported Outcome Measurement Information System (PROMIS) 10 – Global Health Survey, steps, heartrate, blood pressure, and body mass index.

References


Section 17: Feasibility and Pilot Studies

The previous sections on EMAs, SMART and MOST designs have described the benefit of undertaking early studies, particularly qualitative studies, prior to an RCT.

The National Institute for Health Research (NIHR) has adopted quite a narrow definition of a pilot study, being a version of the main study that is run in miniature to test whether the components of the main study can all work together. They would thus focus on the processes of the main study, for example, to ensure recruitment, randomisation, the intervention including all components and follow-up assessments all run smoothly. In some cases the pilot will be the first phase of the RCT; more often it will inform revisions to the draft protocol or intervention.

In contrast NIHR defines feasibility studies as pieces of research done before a main study in order to answer the question “Can this study be done?” They are used to estimate important parameters that are needed to design the main study.

More generally, the purpose of conducting early studies is to examine the feasibility of an approach that is intended to ultimately be used in a larger scale study. This applies to all types of research studies. Their purpose is to increase the probability of success in the larger subsequent studies. The intervention or other aspects of the study design that are judged to be unsatisfactory or sub-optimal can be modified before the subsequent trial or removed altogether. The case studies highlight the value of undertaking early studies (see Mersey Burns, GDm-health, Sleepio, and TrueColours).

Examples of using a pilot or feasibility study to inform the main study for an app intervention are provided by several of the case studies, for example, Mersey Burns, True Colours and GDm-health.

An example of a pilot used to inform a wearable intervention is now described [110].

Aim – To compare patients with Parkinson’s disease undergoing remote motion sensor-based monitoring using objective, wearable sensors to patients receiving standard care to determine the impact on the advanced therapy referral rate and if algorithms could be developed to screen patients for referral.

Trial design – Participants were followed for 1 year. Using the clinician referral as the gold standard, objective motor features representing symptom severity, dyskinesia severity, and fluctuations were used as inputs to classification algorithms developed post-hoc to predict which patients should be referred for advanced therapy.

Intervention – Remote monitoring technology used once a month.
Comparator – Standard care.

Participants – 40 individuals with advanced Parkinson's disease.

Outcomes – Referral rates for advanced therapies.

References


Section 18: Innovative Study Designs

18.1 INSTRUMENTAL VARIABLES

Instrumental variables (IV) estimation is a technique used to control for confounding variables and measurement error. Studies that are not randomised, such as observational studies, may be subject to biases from confounding variables, which limit the ability to interpret causality between intervention and outcomes. Most confounding variables are observable, and may be controlled for statistically in regression models to minimise their impact on the results, however some are unobservable. The IV approach tackles this issue of unobservable confounding variables by identifying the variables or instruments that have 2 properties: 1) they affect the treatment variable, i.e. the intervention, but 2) have no direct effect on the outcome variable. Figure 13.1 shows a schematic diagram of the IV approach. Once the IV has/have been identified they can be entered into the regression models (along with the other confounding variables) to determine the effect caused by the intervention.

Figure 13.1: Schematic of the IV approach (from Newhouse & Maclellan, 1998)

18.2 THE USE OF BIG DATA IN HEALTHCARE

Big data are those that exceed the “processing capacity of conventional database systems” (p1. Wang & Alexander, 2015) and that may be characterised by 6 “V” features: 1) Volume, 2) Velocity, 3) Variety, 4) Value, 5) Variability and 6) Veracity.

In terms of healthcare, Big Data refers to large, complex electronic datasets that are difficult to manipulate with existing methods and software. It may include genomic, clinical, and behavioural data, as well publication and reference data, administrative and business data.
Big Data may be derived and integrated from a wide range of sources beyond clinical data management systems and public health observatories, including social media, and the internet. Sophisticated analysis techniques may be applied to this integrated data to provide individual patient profiles alongside cost and clinical efficacy markers, enabling the most clinically and cost-effective interventions to be identified and tailored to individual patients. Wang & Alexander [113] highlight 3 areas where Big Data could be utilised to reduce waste and inefficiency:

- Clinical operations;
- Research & development;
- Public health.

Some of the benefits these authors identify in terms of medical devices and healthcare include:

- Improved health outcomes due to more accurate and precise diagnostics (tailored to individuals);
- Cost reductions through earlier identification of disease;
- Predicting and managing public health, such as obesity and diabetes, more effectively;
- Reducing unplanned and emergency admission through better healthcare management.

One example where these types of data may be garnered is through patient’s own mobile devices, allowing, for instance, automated monitoring of subjective scores (mood, pain), as well as clinical markers (weight, blood pressure, glucose levels).

Although there are still a number of barriers, such as staffing, budget and infrastructure, as well issues surrounding data confidentiality and privacy to overcome before Big Data become fully integrated in healthcare systems, there is a clear potential for data such as these, collected via apps or other digital devices, to play an important role in the delivery of innovative and existing health and social care.

An example of the use of big data is Care.data, which is led by NHS England and the Health and Social Care Information Centre and aims to bring together securely, health and social care information from different settings in order to see what is working well in the NHS and areas for improvement.

Related topics include data mining and data farming. Data mining refers to the process of using healthcare data collected for various objectives, such as randomised controlled trials and routine data, from a wide range of sources. The scope and extent of the internet means that vast amounts of data collected via mobile apps, the web and social media, can be used to generate evidence to inform and shape the treatment and care of patients, through the process of data mining or farming. The processes and methods used for collecting, transforming, quality assessing and analysing the data are referred to as data farming [5].
References

Section 19:  APP Development and Evaluation

19.1 EVIDENCE ON THE CYCLE OF DEVELOPMENT

The focus of the Sections on individual study designs is to introduce app developers to these different methodologies, with the aim of encouraging them to undertake primary research, to create evidence, to inform evaluation. However, evaluation is just 1 stage in the development process. This section reviews the main findings from 2 recent, high-quality, guidance documents which discuss the cycle of development for complex interventions. This definition includes most apps, particularly those aimed at changing behaviours. These documents provide a context for evaluation within a development cycle, leading to implementation of evidence-based apps.

The first document is by the Medical Research Council (MRC) [115], which offers guidance on the development, evaluation and implementation of complex interventions to improve health. The second document by West & Michie, 2016 [8] builds on the MRC’s work and other evidence sources, to provide guidance on developing and evaluating digital behaviour change interventions in healthcare. The audience for both sets of guidance includes researchers, evaluators and commissioners.

This section identifies the cycle of development contained in both publications and its implications for commissioning research. West R & Michie S, 2016 [8] adapted the MRC’s cycle of development to that shown in Figure 14.1.

**Figure 14.1:** Cycle of development for developing and evaluating complex interventions. [Reproduced with permission from West R & Michie S (2016) [8].]
The cycle suggests development is on-going, informed by evidence from external sources, feedback from piloting and, later in the process, from users’ experiences post-implementation. Development also informs the piloting stage, with evaluation of the results informing further testing or implementation. Implementation itself is accompanied by on-going evaluation and further optimisation. Abandonment is not depicted but is an option at all stages.

19.2 IMPLICATIONS OF THE EVIDENCE ON THE CYCLE OF DEVELOPMENT

The cycle of development aims to transform a novel concept to an optimised app, which commissioners adopt. The findings from the 2 guidance documents have been integrated with our findings on research questions and study designs to identify some of the implications of the development cycle. These include:

- There is no linear process from development to adoption, rather these occur in parallel and inform each other;
- As a starting point in the development phase, establish evidence from the literature, and possibly expert opinion, ideally informed by practice, that the hypothetical theory of change associated with the concept can deliver change in practice. This will usually require undertaking a systematic literature review to identify existing studies and should also identify the key outcomes relevant to the research problem;
- Undertake appropriate development, piloting and evaluation using the criteria of interest. These could be based on the 6 criteria identified as designed to foster innovation, provide care that is evidence-based and meets users’ needs (safety, effectiveness, user-centred, timely, efficient and equitable);
- Initially, pilot studies may focus on improving usability and effectiveness and at a later stage, modelling its potential impact on NHS and social care resources and costs;
- Adopting a Multiphase Optimisation Strategy (MOST) during the piloting phase to identify the impact of different combinations of components, their intensity and sequencing on the evaluation criteria is encouraged;
- The process of evaluating user engagement, usability and effectiveness is likely to be highly iterative, with continual testing and evaluation at every stage. Hence, it does not have a well-defined duration, rather it is more of a continuum and extends beyond implementation;
- Expanding the groups of participants who are testing the app may require trade-offs to be made; hence, participants from the main user groups should be involved in early testing;
- Early modelling, informed by implementation issues, should be undertaken before full-scale testing. This is likely to include analysis of potential impact on clinical pathways, resources and costs;
• The development cycle may move to implementation without full pivotal trials to estimate effect size. This reflects the rapidly changing technology. Hence if a developer can demonstrate a strong theoretical underpinning for the app, has evidence that it works in practice and where risks to commissioners are low (financially and in terms of potential harms for users) then adoption may take place without comparative evidence;
• Further optimisation of the app is likely to take place as the app is scaled-up;
• At some point, when the app is reasonably stable, comparative evidence of clinical and cost-effectiveness testing is likely to be required to be able to demonstrate the app is indeed having the claimed effect. The associated evidence may support wider roll-out, further development or abandonment;
• Comparative testing should consider both process and outcome measures and include qualitative research to help understand the quantitative data produced;
• Randomisation should always be considered for pivotal studies. This is the most robust method of demonstrating causality between intervention and effect. However, randomisation may be unnecessary if the effects of the intervention are so large or immediate that confounding or underlying trends are unlikely to explain differences in outcomes before and after exposure;
• Results must be accessible to decision-makers, with publication in a peer-reviewed journal enhancing their credibility;
• Implementation needs to consider aspects such as scaling-up and sustainability, including obtaining long-term funding. Such funding may be easier to obtain if there is quality evidence that the app meets the criteria, which inform commissioners’ decision-making;
• Constant data collection and evaluation underpins the development cycle, with different study designs and different durations required at different stages in the cycle and their sequencing will vary with the evidence collected;
• There is no single process or design, which should be used to answer a particular research question; rather different designs suit different questions and different circumstances. However, an awareness of the range of potential approaches should lead to more appropriate methodological choices.

19.3 A RESEARCH STRATEGY FOR CARE APPS

Table 14.1 examples a high-quality mobile health application’s research strategy, developed by a team of researchers from New Zealand. Whilst no one app will be expected to complete all of these steps it indicates that developers should expect to use a mix of study designs at different stages in the life cycle of an app.
**Table 14.1:** Summary of the research and evaluation steps in the development of a mobile health intervention

<table>
<thead>
<tr>
<th>Research methods</th>
<th>Purpose</th>
<th>Example measures</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Formative research</strong></td>
<td>• Focus groups; • Online surveys.</td>
<td>To inform the development of the intervention content and process.</td>
</tr>
<tr>
<td><strong>Pretesting</strong></td>
<td>• Online surveys; • Focus groups; • Individual interviews.</td>
<td>• To determine acceptability of proposed intervention to target audience; • To improve and refine intervention on basis of feedback.</td>
</tr>
<tr>
<td><strong>Pilot study</strong></td>
<td>Small and nonrandomised.</td>
<td>• To test content and regimen of intervention; • To test processes.</td>
</tr>
<tr>
<td><strong>Randomised control trial</strong></td>
<td>Pragmatic community-based randomised control trial.</td>
<td>To test the effect of the intervention in comparison with a control group.</td>
</tr>
<tr>
<td><strong>Qualitative research</strong></td>
<td>Semi-structured interviews.</td>
<td>• To improve the intervention further; • To determine implementation issues and methods.</td>
</tr>
<tr>
<td><strong>Evaluation of implementation impact</strong></td>
<td>• Phone / online surveys; • Semi-structured interviews.</td>
<td>To determine the effect of the intervention once scaled up.</td>
</tr>
</tbody>
</table>

References


Section 20: Wearable Devices

A wearable sensor is a small electronic device containing 1 or more sensors which can transduce information related to the physiology of the device user, the ambient environment, or the users' interaction with the environment. Common sensors used in wearable monitoring systems include those for measuring movement and position, such as accelerometers and gyroscopes, GPS, and sensors for assessing electrophysiological and chemophysiological function, or other physiological properties such as body temperature [118].

Successful biosensors should be able to demonstrate they are accurate, precise and reproducible. The reader device should generate and display results in a user-friendly way. Ideally the wearable will also be small, unobtrusive, portable, easy to operate, have no toxic or antigenic effects and reasonably cheap. Example include fitbits, blood pressure monitors, ingestible devices and glucose monitors.

Realising the clinical benefits of using data from a wearable sensor may require it to be integrated with other health data in order to improve clinical decision making. A recent European Commission report [119] highlighted the benefits but also the barriers that such technologies face when seeking to integrate with NHS systems. For example, the ideal service for a sensor measuring glucose levels in pregnant women may be for the results to be uploaded into the woman's electronic patient record, with an alert to clinicians provided when these deviate from a pre-defined range. Thus deterioration can be investigated and action taken to prevent further decline and hence avoid adverse events for the woman and baby. In common with other mHealth providers, achieving such integration requires companies developing wearables to overcome several barriers - for example on meeting standards on data protection, security and privacy and the requirements on interoperability and overcoming cultural resistance. However, without such links, providing evidence of the benefit of using the output from the sensor to achieve change in clinical outcomes or NHS resource use will be difficult.

Wearables often provide the capability to record data continuously and hence the potential to generate large, complex datasets which must be interpreted to provide clinically relevant information which patients and clinicians can use to inform clinical judgements and influence future care. Torous et al [120] add that accessing the information should not increase the workload of clinicians. Other characteristics identified by these authors to support clinicians in evaluating the benefit of using data from wearables, assuming transmission of data by apps, include:

- Evidence of efficacy and safety to enable clinicians to balance risks and harms and consider potential unintended consequences. They noted that in a recent survey, physicians cited a lack of evidence based content as one primary concern in recommending apps for patients;
• Transparency in reporting how data are collected, analysed, stored, transmitted and shared, including a transparent privacy policy; and
• The need for encryption and other such measures to ensure data security.

A related article by Armontrout et al. [121] addressed some of the potential liabilities which clinicians may be exposed to if they recommend an app (and by extension wearables) which has not undergone rigorous safety and efficacy testing. Their discussion is specifically around mobile mental health apps but the concerns generalise to other clinical settings and to wearables.

A literature review, undertaken to inform this chapter, identified a small but growing number of systematic reviews which summarise the evidence base for specific wearables. For example, Urrea et al [122] reviewed recent clinical studies using smartphone devices, social media, and wearable health tracking devices designed to facilitate behavioural interventions to improve risk factors from smoking, physical inactivity, and sub-optimal nutrition. The authors’ assessment of the evidence led them to conclude that these technologies have the potential to provide ‘low-cost, scalable, and individualized tools to improve management of these important cardiovascular disease risk factors.’

No quality assessment of the studies informing their conclusions was presented. However, of the reviews which assessed quality of evidence, virtually all identified some limitations. Examples include:

• A review by Ryan et al [123] examined the best sensor types to assess balance and stability in people with Parkinson’s disease. Of 26 included articles, 31% were of low methodological quality, 58% were of moderate quality and 11% were of high methodological quality. It concluded further high-quality research was needed, adding that future studies should consider the internal and external validity of their methods and provide an appropriate sample size calculation;
• A review by Lewis et al [124] synthesised evidence on the efficacy of wearable electronic activity monitor systems. Of the 11 included articles, 8 were of medium quality, 2 were of low quality, and 1 of high quality. It identified more high-quality, adequately powered, randomized controlled trials were needed;
• A review by Uyei J et al [125] on wearable defibrillators, included 36 articles and conference abstracts and concluded the quality of evidence was ‘low to very low’ giving rise to low confidence in its results.

The exception was a review by Dhillon et al [126] which reviewed the accuracy of physical activity measurement strategies in adults with chronic lung disease. Of the 9 studies included, 8 were judged of high quality.

These findings are consistent with those from Steinhubl et al. [127] who noted several recent systematic reviews had concluded that high-quality evidence was lacking for the use of mHealth to effect behavioural changes or to manage chronic diseases, inpatient care, or health care delivery.
As Torous et al [120] noted lack of evidence might be one of greatest barriers to increased use of apps, and by extension wearables, at present.

Piwek et al [128], provided an overview of the potential for wearables to give patients direct access to personal measurements that can enhance prevention behaviours and help in managing disease and also considers the potential benefit to medicine. A key focus is on the need for robust evidence. Findings include:

- The current empirical evidence does not support the hypothesis that wearables change behaviour. The authors note much of the current evidence comes from single-subject reports of users describing their experiences, adding these cannot be treated as reliable scientific evidence. Rather what is required is longitudinal, randomised controlled studies focusing on the impact of wearables on healthy users' behaviour;
- Evidence on pedometers was the exception and the authors accepted that there was good evidence from a well-conducted study that these (and consultations) increased physical activity among older people. They raise the issue of durability of response, noting use drops off after 6 months to 1 year;
- Developers should improve their use of ‘user experience principles’ in product development, adding: ‘Those who market and develop consumer level devices may underestimate the distance between designing a product that appears to be associated with a healthy lifestyle and providing [RCT] evidence to support this underlying assumption.’
- For chronic conditions such as sleep apnoea and depression, wearables can provide detailed longitudinal data to monitor patients’ progress;
- Wearables can also inform "predictive preventive diagnosis". For example, an analysis of movement data can be used to detect early symptoms of Parkinson disease.
- Wearables can assist in maintaining people with long-term conditions in their own homes but improving their use over a sustained period is required;
- Such potential applications are still in the early stages of development, have not been approved for medical use and are poorly explored in a real-world context;
- Further research is needed to establish if wearable-generated data are safe, reliable, and consistent with privacy and security of personal data requirements;
- The authors suggest creating a simple regulatory framework that does not suppress innovation but which assists in validating wearable devices may increase the development of products of value to the NHS;
- The introduction of industry wide standards to address issues of reliability, safety, and security of patient data, combined with the appropriate regulatory framework ‘has the potential to accelerate high-quality, large-scale randomised controlled trials in order to deconstruct complex causal interactions and better understand how to make wearables safer and more useful if they are to be adopted in health care.’
- Developing ‘intelligent and personalised explanatory feedback’ could improve user experience and data reliability. This would require developers to work closely with clinicians and test impact by undertaking RCTs.
References


Section 21: Systematic Reviews and Meta-Analyses

This section describes the features of a systematic review and meta-analysis. These are important to inform recommendations for adopting new interventions, using an evidence based medicine methodology. Developers should however be aware of the information used to inform these products so that they can publish evidence, which is of sufficient quality to be included in such reviews, and hence influence decision makers. They are also an important source of evidence to inform Evidence on the Cycle of Development.

21.1 SYSTEMATIC REVIEWS

A systematic review seeks to identify, select and collate all evidence that fits pre-specified eligibility criteria in order to address a specific research question. It uses explicit, systematic methods that are selected with a view to minimizing bias, thus providing more reliable findings from which conclusions can be drawn and decisions made. A major use of systematic reviews is to inform policy and practice about the effectiveness of healthcare interventions.

A robust and widely respected methodology setting out how to conduct systematic reviews is detailed in Cochrane Handbook [18]. This defines the key characteristics of a systematic review to be:

- A clearly stated set of objectives with pre-defined eligibility criteria for studies;
- An explicit, reproducible methodology;
- A systematic search that attempts to identify all studies that would meet the eligibility criteria;
- An assessment of the validity of the findings of the included studies, for example through the assessment of risk of bias;
- A systematic presentation, and synthesis, of the characteristics and findings of the included studies.

There are now a range of systematic reviews published on healthcare apps with several examples provided in the references.

The findings of literature reviews can inform the development process by identifying the evidence on the impact of different behavioural change theories in any specific content. Hence, the app development process can be based on techniques, which have been shown to work elsewhere and generalising these findings to the specific situation. Some of the recent reviews have reported the reverse - that many current apps are not evidence-based and thus of questionable quality (see West R & Michie S, 2016 [8]).
21.2 META-ANALYSES

Many systematic reviews contain meta-analyses. These adopt statistical methods to summarize the results of the independent studies included in the review. Most meta-analysis methods provide a weighted average of the effect estimates from the different studies. By combining this information from all relevant studies, meta-analyses can provide more precise estimates of the effects of an intervention than those derived from the individual studies. They also facilitate investigations of the consistency of evidence and the exploration of differences across studies.

Undertaking meta-analyses is useful provided certain criteria are met. Benefits include increasing the explanatory power of studies, improving the precision of the estimated effect size and providing a graphical representation of the range of effect sizes so outliers can easily be identified. However, they should only be conducted if the included studies are reasonably similar in terms of designs, participants, conduct and outcomes. Combining heterogeneous studies may result in incorrect results. Specific tests are available to test for statistical heterogeneity, but it is also important, especially in the field of apps, to consider ‘clinical’ heterogeneity – it may not be appropriate to combine the results of similar studies if the interventions themselves are not sufficiently similar.

Information on the potential effect size and its variability can inform trial design, inclusion criteria and sample size and indeed support the case that a trial is warranted and ethical [115].

References


Section 22: Critical Appraisal and Associated Checklists to Assess the Quality of Studies

Critical appraisal is the process of systematically examining research evidence to judge its trustworthiness and relevance in a particular context. The quality of a study's overall design, its execution, the validity of its results and their generalisability to the research question, determine its relevance to the decision problem.

This process is relevant to mHealth developers as they create an evidence base because most evaluators critically appraise evidence and only use evidence from studies, which are judged to be of reasonable quality to inform their decisions. Thus undertaking poor trials, which give results which do not inform the expected effectiveness or safety of an intervention in the real world, is a poor use of resources and potentially highly misleading.

Critical appraisal checklists are useful tools to undertake this process. These usually address the conduct of the study, its reporting and relevance to the research question. Checklists operate by asking questions and most provide prompts to remind the user why the question is important. Ideally the quality of a checklist will have been evaluated by a process of piloting drafts with a number of independent experts. Each main study design has a checklist tailored to its relevant features.

One checklist specific to the reporting of mobile health interventions is now available [136] (see http://www.bmj.com/content/352/bmj.i1174). This 16 item checklist identifies a minimum set of information needed to define what the mHealth intervention is (content), where it is being implemented (context), and how it was implemented (technical features), to support replication of the intervention. The checklist is accompanied by a detailed explanation for each item, with illustrative reporting examples.

Information on appraising different generic study designs is now provided. However, before appraising the evidence, it is necessary to identify the research design used in a study so an appropriate critical appraisal checklist can be completed. Such a tool such is available from the Centre for Evidence Based Medicine (see www.cebm.net/index.aspx?o4r=1039).
The following is a non-exhaustive list of potentially useful tools available to aid critical appraisal of a wide variety of study types (alphabetically listed).

- Centre for Evidence Based Medicine has developed several critical appraisal checklists, accessible at [www.cebm.net/index.aspx?o=1913](http://www.cebm.net/index.aspx?o=1913). These cover the following study designs:
  - Systematic review;
  - Diagnostic study;
  - RCT;
  - Prognostic study.

- Critical Appraisal Skills Programmes (CASP), available at [www.casp-uk.net](http://www.casp-uk.net). This site gives general advice about evidence-based medicine and critical appraisal, and provides checklists for the following study types:
  - Randomised controlled trial (RCT);
  - Systematic review and meta-analysis;
  - Cohort study;
  - Qualitative research;
  - Economic evaluation;
  - Diagnostic test;
  - Clinical prediction rule.

- Consolidated standards Of Reporting Trials (CONSORT), available at [www.consort-statement.org](http://www.consort-statement.org) aims to improve the reporting of RCTs. The site includes a 25 item checklist for the critical appraisal of RCTs, as well as a downloadable flow diagram to assist tracking participants through an RCT;

- Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement updates previous health economic evaluation guidelines and checklists and sets the standard required for publication in most medical journals;

- Downs and Black, available at [http://jech.bmj.com/content/52/6/377.full.pdf+html](http://jech.bmj.com/content/52/6/377.full.pdf+html), is a checklist to assess randomised and non-randomised studies. It provides an overall score for study quality and a profile of scores for the quality of reporting, internal validity (bias and confounding), power and external validity;

- Graphic Appraisal Tool for Epidemiology (GATE), available at [www.fmhs.auckland.ac.nz/soph/depts/epi/epiq_provides](http://www.fmhs.auckland.ac.nz/soph/depts/epi/epiq_provides) graphical checklists to appraise:
  - RCTs;
  - Cohort studies;
  - Cross-sectional studies;
  - Diagnostic accuracy studies.

- Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA), available at [www.prisma-statement.org](http://www.prisma-statement.org) provides a 27 item checklist for the critical appraisal of systematic reviews and meta-analyses. PRISMA has also published a flow diagram useful to document the exclusion of trials from systematic reviews and meta-analyses;
- Quality Assessment of Diagnostic Accuracy Studies (QUADAS II), available from [www.bris.ac.uk/quadas/quadas-2](http://www.bris.ac.uk/quadas/quadas-2) provides a 4-domain checklist to evaluate diagnostic accuracy studies;
- Scottish Intercollegiate Guidelines Network (SIGN), available at [www.sign.ac.uk/methodology/checklists](http://www.sign.ac.uk/methodology/checklists) has a range of checklists, covering:
  - Systematic reviews and meta-analysis;
  - RCT;
  - Cohort study;
  - Case control study;
  - Diagnostic study;
  - Economic study.
- Strengthening the Reporting of Observational studies in Epidemiology (STROBE), available at [www.strobe-statement.org](http://www.strobe-statement.org). STROBE provides critical appraisal checklists for the following observational study designs:
  - Cohort study;
  - Case-control study;
  - Cross-sectional study;
  - Conference abstract.
- Standards for the Reporting of Diagnostic accuracy studies (STARD), available at [www.stard-statement.org](http://www.stard-statement.org) provides critical appraisal checklists and flow diagrams for diagnostic accuracy studies;

The following table summarises the previously described checklists by study design.
Table 16.1: Checklists for different study designs

<table>
<thead>
<tr>
<th>Study design</th>
<th>Critical appraisal checklist</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Systematic reviews and meta-analyses</td>
<td>CASP</td>
<td><a href="http://www.casp-uk.net">www.casp-uk.net</a></td>
</tr>
<tr>
<td></td>
<td>CEBM</td>
<td><a href="http://www.cebm.net/index.aspx?o=1913">www.cebm.net/index.aspx?o=1913</a></td>
</tr>
<tr>
<td></td>
<td>PRISMA</td>
<td><a href="http://www.prisma-statement.org">www.prisma-statement.org</a></td>
</tr>
<tr>
<td></td>
<td>SIGN</td>
<td><a href="http://www.sign.ac.uk/methodology/checklists">www.sign.ac.uk/methodology/checklists</a></td>
</tr>
<tr>
<td>RCTs</td>
<td>CASP</td>
<td><a href="http://www.casp-uk.net">www.casp-uk.net</a></td>
</tr>
<tr>
<td></td>
<td>CEBM</td>
<td><a href="http://www.cebm.net/index.aspx?o=1913">www.cebm.net/index.aspx?o=1913</a></td>
</tr>
<tr>
<td></td>
<td>CONSORT</td>
<td><a href="http://www.consort-statement.org">www.consort-statement.org</a></td>
</tr>
<tr>
<td></td>
<td>GATE</td>
<td><a href="http://www.fmhs.auckland.ac.nz/soph/depts/epi/epiq">www.fmhs.auckland.ac.nz/soph/depts/epi/epiq</a></td>
</tr>
<tr>
<td></td>
<td>SIGN</td>
<td><a href="http://www.sign.ac.uk/methodology/checklists">www.sign.ac.uk/methodology/checklists</a></td>
</tr>
<tr>
<td>Non-randomised controlled studies</td>
<td>TREND</td>
<td><a href="http://www.cdc.gov/trendstatement">www.cdc.gov/trendstatement</a></td>
</tr>
<tr>
<td>Cross-sectional studies</td>
<td>GATE</td>
<td><a href="http://www.fmhs.auckland.ac.nz/soph/depts/epi/epiq">www.fmhs.auckland.ac.nz/soph/depts/epi/epiq</a></td>
</tr>
<tr>
<td></td>
<td>STROBE</td>
<td><a href="http://www.strobe-statement.org">www.strobe-statement.org</a></td>
</tr>
<tr>
<td>Cohort studies</td>
<td>CASP</td>
<td><a href="http://www.casp-uk.net">www.casp-uk.net</a></td>
</tr>
<tr>
<td></td>
<td>GATE</td>
<td><a href="http://www.fmhs.auckland.ac.nz/soph/depts/epi/epiq">www.fmhs.auckland.ac.nz/soph/depts/epi/epiq</a></td>
</tr>
<tr>
<td></td>
<td>SIGN</td>
<td><a href="http://www.sign.ac.uk/methodology/checklists">www.sign.ac.uk/methodology/checklists</a></td>
</tr>
<tr>
<td></td>
<td>STROBE</td>
<td><a href="http://www.strobe-statement.org">www.strobe-statement.org</a></td>
</tr>
<tr>
<td>Case control studies</td>
<td>CASP</td>
<td><a href="http://www.casp-uk.net">www.casp-uk.net</a></td>
</tr>
<tr>
<td></td>
<td>SIGN</td>
<td><a href="http://www.sign.ac.uk/methodology/checklists">www.sign.ac.uk/methodology/checklists</a></td>
</tr>
<tr>
<td></td>
<td>STROBE</td>
<td><a href="http://www.strobe-statement.org">www.strobe-statement.org</a></td>
</tr>
<tr>
<td>Diagnostic accuracy studies</td>
<td>CASP</td>
<td><a href="http://www.casp-uk.net">www.casp-uk.net</a></td>
</tr>
<tr>
<td></td>
<td>CEBM</td>
<td><a href="http://www.cebm.net/index.aspx?o=1913">www.cebm.net/index.aspx?o=1913</a></td>
</tr>
<tr>
<td></td>
<td>GATE</td>
<td><a href="http://www.fmhs.auckland.ac.nz/soph/depts/epi/epiq">www.fmhs.auckland.ac.nz/soph/depts/epi/epiq</a></td>
</tr>
<tr>
<td></td>
<td>QUADAS II</td>
<td><a href="http://www.bris.ac.uk/quadas/quadas-2">www.bris.ac.uk/quadas/quadas-2</a></td>
</tr>
<tr>
<td></td>
<td>SIGN</td>
<td><a href="http://www.sign.ac.uk/methodology/checklists">www.sign.ac.uk/methodology/checklists</a></td>
</tr>
<tr>
<td></td>
<td>STARD</td>
<td><a href="http://www.stard-statement.org">www.stard-statement.org</a></td>
</tr>
<tr>
<td>Qualitative studies</td>
<td>CASP</td>
<td><a href="http://www.casp-uk.net">www.casp-uk.net</a></td>
</tr>
<tr>
<td></td>
<td>SIGN</td>
<td><a href="http://www.sign.ac.uk/methodology/checklists">www.sign.ac.uk/methodology/checklists</a></td>
</tr>
<tr>
<td>Case studies and cases series</td>
<td>Downs &amp; Black</td>
<td><a href="http://jtech.bmj.com/content/52/6/377.full.pdf+html">http://jtech.bmj.com/content/52/6/377.full.pdf+html</a></td>
</tr>
</tbody>
</table>

Reference


Section 23: Hierarchy of Evidence

Assessing a body of evidence for its quality and volume, and their implications for informing recommendations, has received increased focus in recent years. Sources of evidence range from case reports that are prone to bias to well-designed RCTs that have minimised bias. Since poor quality evidence can lead to recommendations that are not in patients’ best interests, it is essential to know whether a recommendation is strong (we can be confident about the recommendation) or weak (we cannot be confident). In 1995 Guyatt and his colleagues [138] advocated a “hierarchy of evidence” to rank the evidence generated from individual studies. A modified version of this has been adopted by Scottish Intercollegiate Guidelines (SIGN) (http://www.sign.ac.uk/pdf/SIGN147.pdf) and is shown in Figure 17.1.

Figure 17.1: Hierarchy of evidence

<table>
<thead>
<tr>
<th>LEVELS OF EVIDENCE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1</strong></td>
</tr>
<tr>
<td><strong>1</strong></td>
</tr>
<tr>
<td><strong>1</strong></td>
</tr>
<tr>
<td><strong>2</strong></td>
</tr>
<tr>
<td><strong>2</strong></td>
</tr>
<tr>
<td><strong>2</strong></td>
</tr>
<tr>
<td><strong>2</strong></td>
</tr>
<tr>
<td><strong>3</strong></td>
</tr>
<tr>
<td><strong>4</strong></td>
</tr>
</tbody>
</table>

High quality systematic reviews, meta-analyses and RCTs with a very low risk of bias constitute the highest level of evidence. Further down this hierarchy are case-control and cohort studies, with expert opinion being the lowest level of evidence.
The GRADE system (Grading of Recommendations Assessment, Development and Evaluation) is another approach to rating the quality of evidence and is mainly used in grading the strength of recommendations for specific interventions in guidelines or other evidence based products (e.g. Guyatt et al., 2011 [138]). Hence, GRADE is used to assess the quality of an outcome across studies, rather than the quality of individual studies. The quality of evidence for each outcome is usually rated on a 4 point scale ranging from “high” to “very low quality”. The following aspects are rated using this system:

- Study limitations;
- Inconsistency of results;
- Indirectness of evidence;
- Imprecision;
- Reporting bias.

A score is derived for each of these features. An example (from Guyatt et al., 2011 [138]) is shown in Figure 17.2.

**Figure 17.2: Example of quality evaluation**

<table>
<thead>
<tr>
<th>Study Design</th>
<th>Quality of Evidence</th>
<th>Lower if</th>
<th>Higher if</th>
</tr>
</thead>
<tbody>
<tr>
<td>Randomized trial</td>
<td>High</td>
<td>Risk of bias</td>
<td>Large effect</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-1 Serious</td>
<td>+1 Large</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-2 Very serious</td>
<td>+2 Very large</td>
</tr>
<tr>
<td></td>
<td>Moderate</td>
<td>Inconsistency</td>
<td>Dose response</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-1 Serious</td>
<td>+1 Evidence of a gradient</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-2 Very serious</td>
<td></td>
</tr>
<tr>
<td>Observational study</td>
<td>Low</td>
<td>Indirectness</td>
<td>All plausible confounding</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-1 Serious</td>
<td>+1 Would reduce a demonstrated effect or</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-2 Very serious</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Very low</td>
<td>Imprecision</td>
<td>+1 Would suggest a spurious effect when results show no effect</td>
</tr>
</tbody>
</table>

Scores depend on the quality of each study. So a poorly implemented RCT which has a high likelihood of bias or inconsistent results may be scored as ‘low’ or ‘very low’ quality of evidence. Conversely, scores on an observational study may be increased for large effect sizes. After going through the process of grading all evidence for a particular outcome, the overall quality will be categorised as ‘high’, ‘moderate’, ‘low’ or ‘very low’.

The GRADE approach is now used by clinical guideline developers across the UK. Hence, it will be applied to apps which seek to be adopted as recommended clinical practice within clinical guidelines.
References


Section 24: Health and Social Care Information Centre

24.1 ROLE OF HEALTH AND SOCIAL CARE INFORMATION CENTRE

The Health and Social Care Information Centre (HSCIC) supports the delivery of information technology (IT) infrastructure, information systems and governance to the NHS and social care systems. It also works with others to develop standards to ensure information flows efficiently and securely across these systems.

Developers seeking to access NHS networks must conform to relevant policies and standards, in addition to legal requirements on data safety and security. Those building standalone apps may find it advantageous to design systems consistent with HSCIC and NHS nomenclature, clinical and other coding systems and deploy robust quality assurance, potentially using tools available from the HSCIC site. This section advises on the range of services available to mHealth developers from HSCIC. One page summaries of selected work programmes are also available from the website.

24.2 HSCIC’S ROLE IN SUPPORTING CLINICAL SAFETY

The HSCIC has developed Clinical Risk Management Standards which are designed to promote and help embed clinically safer working practice methods and patient safety solutions, enabled by IT, consistently across the NHS.

The ISB 0129 Clinical Risk Management: its Application in the Manufacture of Health IT Systems standard sets clinical risk management requirements for manufacturers of health IT systems. It requires a manufacturer to establish a framework within which clinical risks associated with the design and development of a new health IT system, or the modification of an existing system, are properly managed.

The main activities defined within the standard are:

- Identification of any potential hazards the health IT system may present to patients;
- Assessment of the severity and likelihood of each hazard and hence the clinical risk;
- Evaluation of each clinical risk to determine whether it is acceptable against defined risk acceptability criteria;
- Implementation of suitable clinical risk control measures to reduce or mitigate unacceptable clinical risks.
It also identifies the main documents required to provide evidence of compliance being:

- Clinical Risk Management Plan, which defines the organisation’s approach to clinical risk management for a particular project;
- Hazard Log to record and communicate the on-going identification and resolution of hazards associated with a Health IT System;
- Clinical Safety Case Report which seeks to demonstrate the safety of the Health IT System;
- Safety Incident Management Log, which supports the communication and resolution of raised incidents.

Compliance with this standard ensures that the manufacturer has implemented an effective, best practice, clinical safety programme during the design and build of a health IT system.

The second standard applies to health organisations responsible for the use of Health IT Systems within the NHS and covers similar activities (*ISB 0160 Clinical Risk Management: its Application in the Deployment and Use of Health IT Systems*).

### 24.3 HSCIC’S ROLE IN SUPPORTING IT INFRASTRUCTURE

HSCIC is responsible for developing, delivering and maintaining the NHS national IT infrastructure. The website describes a range of national systems and services, including a full description of the Health and Social Care Network (HSCN). In 2017, the HSCN will replace the current N3 system. It will enable health and care organisations to access and exchange electronic information by standardising networks, enabling service sharing and reducing complexity.

The website also provides information on a wide range of infrastructure related topics such as:

- **Automatic Identification and Data Capture** using machine readable bar codes;
- **Choose and Book** the e-referral system;
- **Digital delivery service** which oversees key national applications and infrastructure;
- **Interoperability**;
- **NHS mail**;
- **NHS Number**;
- **Data transfer tools**.
24.4 HSCIC’S ROLE IN SUPPORTING INFORMATION SYSTEMS AND INFORMATION GOVERNANCE

The HSCIC website has several resources addressing information systems and information governance, which may be relevant to apps developers:

- **Information Standards** explains the products and services provided by HSCIC to assist the collection, management and sharing of health and social care information. The following topics are addressed:
  - Common User Interface (CUI) provides a portfolio of standards and guidance relating to the design of user interfaces for healthcare computing systems. The core objectives include: increasing patient safety, the clinical take-up of health IT and reducing training costs ([http://systems.hscic.gov.uk/data/cui](http://systems.hscic.gov.uk/data/cui));
  - Information standards specify rules for the collection, processing, management and sharing of information to support patient care. These rules may include technical standards, data standards or information governance standards. The purpose and content of each of the 3 types of standards (Standards Framework, Reference Standard and Operational Standards) are described ([http://systems.hscic.gov.uk/data/learn](http://systems.hscic.gov.uk/data/learn));
  - UK Terminology Centre (UKTC) is responsible for the UK management of SNOMED CT, Read codes and other healthcare terminology products. The UKTC maintains the NHS Dictionary of Medicines and Devices (dm+d) in partnership with the NHS Business Service Authority; ([http://systems.hscic.gov.uk/data/uktc](http://systems.hscic.gov.uk/data/uktc));
  - Clinical Classifications Service is the definitive source of clinical coding guidance and sets the national standards used by the NHS in coding clinical data ([http://systems.hscic.gov.uk/data/clinicalcoding](http://systems.hscic.gov.uk/data/clinicalcoding));
  - NHS Data Model and Dictionary Service provides a reference point for information standards to support healthcare activities within the NHS in England and has a useful Frequently Asked Questions section, a data model and dictionary and useful links to related sites; ([http://systems.hscic.gov.uk/data/nhsdmds](http://systems.hscic.gov.uk/data/nhsdmds)).

- **Information Governance** (IG) guidance is provided to help health and care organisations meet the standards required to handle care information. It covers a range of topics including:
  - Standards of practice for health record confidentiality;
  - IG standards and guidance for NHS and partner organisations;
  - IG requirements for organisations accessing NHS digital services including N3;
  - Information security Safeguards and guidelines for protecting patient data;
  - NHS Codes of Practice and legal obligations;
  - Confidentiality, information security management and NHS records management;
  - A link to the Information Governance Alliance, a group of national independent care organisations promoting good IG.
Supporting open data and transparency (http://www.hscic.gov.uk/transparency);
Indicator Assurance Service which provides tools to ensure health and social care indicators are well defined, based on good data and transparent methodologies and provides a quality assurance service. There is also a National Library of Quality Assured Indicators where users can find indicators that have been assessed for their robustness and quality (http://www.hscic.gov.uk/article/1674/Indicator-Assurance-Service).

24.5 INFORMATION STANDARDS AND COLLECTIONS

The Standardisation Committee for Care Information (SCCI) oversees the development, assurance and approval of information standards, data collections and data extractions. All current SCCI and Information Standards Board (ISB) standards and collections (n=129) are available at: https://groups.ic.nhs.uk/SCCIDsupport/dashboard/Lists/ISCEportfolio/Home.aspx

The scope of these standards ranges from the provision of accessible information to those with a disability, impairment or sensory loss to data collection exercises, disease specific and screening datasets, common user interface, management of sensitive data and secure use of emails. Email support for enquires is also available.
Section 25: Ethical and Other Approvals

25.1 ETHICAL PRINCIPLES

The key ethical principles governing healthcare research were set out in 2000, when the World Medical Association published the Declaration of Helsinki. These cover all medical research involving human subjects. Fundamental principles include respect for the individual, their rights to self-determination and to make informed decisions regarding participation in research, that the investigator’s duty is solely to the patient or volunteer, whose welfare takes precedence over the interests of science and society and that ethical considerations take precedence over laws and regulations.

In 2002, the Council for International Organizations of Medical Sciences (CIOMS) published International Ethical Guidelines for Biomedical Research Involving Human Subjects. These relate mainly to ethical justification and scientific validity of research; ethical review; informed consent; vulnerability of individuals and groups; equity; choice of control in clinical trials; confidentiality; compensation for injury; and obligations of sponsors.

25.2 RESEARCH GOVERNANCE IN NHS AND SOCIAL CARE

In 2005, the Department of Health published the Research Governance Framework for Health and Social Care. The framework applies to all research undertaken by NHS or social care staff and any research undertaken by industry, charities, and research councils and universities within health and social care systems. It describes the standards required, responsibilities of participants to deliver these and quality assurance requirement. A core principle is protection of participants’ rights including by ensuring researchers are appropriately qualified, with relevant skills and experience and funding is adequate.

The conduct of clinical trials in the UK is also governed by the EU Regulation ‘EU No 536/2014 of the European Parliament and of the Council on clinical trials on medicinal products for human use’. Its stated aim is to create an environment that is favourable for conducting clinical trials, whilst providing the highest standards of patient safety.
25.3 APPROVALS FOR STUDIES TO BE CONDUCTED IN THE NHS IN ENGLAND

Regulating health research aims to provide participants with assurance that the research that they take part in is of high quality, safe and ethical. It also ensures that the results of research can be relied upon and used as evidence to inform future decisions about healthcare and treatment.

Health Research Authority (HRA) approval is a new approval process for the NHS in England that comprises a review by a NHS Research Ethics Committee (where required) as well as an assessment of regulatory compliance and related matters undertaken by dedicated HRA staff. HRA staff will seek evidence on matters including:

- Compliance with legislation and HRA Assessment standards;
- Contract assurance;
- Study delivery arrangements.

NHS Research Ethics Committees (REC) review applications for research and give an opinion about the proposed participant involvement and whether the research is ethical within 60 days. The HRA has published an algorithm to assist in determining whether a project requires ethical review by a REC. For example, if a project is a clinical audit, service evaluation, case study, satisfaction survey or equipment/systems testing then ethical approval will not be required. Research requiring ethics approval includes medicinal products or devices involving:

- Patients and users of the NHS;
- Individuals identified as potential research participants because of their status as relatives or carers of patients and users of the NHS;
- NHS staff – recruited as research participants by virtue of their professional role.

A Department of Health publication, ‘Governance arrangements for NHS Research Ethics Committees’ defines roles and responsibilities of the principal investigator and NHS organisation, and the issues considered by RECs for each research proposal.

The publication recommends that for research in health and social care occurring outside the NHS, an opinion is obtained from a NHS REC, or from a REC meeting the general standards for NHS RECs laid down in that document. If requested to do so, a NHS REC may also provide an opinion on the ethics of research studies carried out, for example, by private sector companies, the Medical Research Council (or other public sector organisations), charities or Universities in the NHS.

Under the new arrangements, local Research and Development (R&D) organisations are responsible for working with sponsors to ensure sites are suitable for a study and putting the arrangements in place to set up and oversee delivery.
HRA approval is based on a single application package to the HRA for both HRA assessment and the independent REC review. Documents are uploaded to Integrated Research Approval System, a web-based system provided by the HRA.

The HRA approval section of the HRA website (http://www.hra.nhs.uk/about-the-hra/our-plans-and-projects/assessment-approval/) provides a step-by-step guide to the new process, has leaflets for the research community and other groups, and provides contact information for the approval team.

An ethical opinion cannot be given retrospectively. If the research has already started within the NHS, without first obtaining a favourable ethical opinion, this is a breach of research governance and, in the case of a clinical trial of an investigational medicinal product; a criminal offence may also have been committed. Hence, developers should always seek an ethical opinion prior to the start of the research.

25.4 APPROVALS FOR STUDIES TO BE CONDUCTED IN UNIVERSITIES

Universities require researchers and students who conduct research involving human subjects, materials or data not in the public domain to apply for ethical approval. Each university will have its own Research Ethics Committee, which will adopt similar criteria to NHS RECs. Most such committees will also consider legal requirements on access to and use of research resources and data.

25.5 PUBLICATION AND ETHICS

The Committee on Publication Ethics (COPE) has over 10,000 members including editors of academic journals and others interested in publication ethics. COPE provides advice to members on all aspects of publication ethics including setting minimum standards of behaviour for editors and publishers through the Code of Conduct and Best Practice Guidelines for Journal Editors and Code of Conduct for Journal Publishers. Publishers and editors undertake to set journal policies addressing factors including research ethics, such as confidentiality, consent, with special requirements for human and animal research. Failure to comply with good research ethics practice may result in manuscripts being rejected.

Guidance is also available to authors reporting clinical trials, systematic reviews and to facilitate complete and transparent reporting, reduce the influence of bias on results, and aiding critical appraisal and interpretation. For example, the CONSORT Statement is a minimum set of recommendations for reporting randomized trials.
References


Section 26: General Sources of Information

This section provides links to a range of targeted websites which offer guidance to mHealth developers seeking to develop products for the health and care sectors. The organisations are grouped into 6 themes being:

- Conduct of research;
- Health and social care organisations;
- Independent organisations;
- Regulatory bodies;
- Standards development agencies;
- Platform developers.

This list is not comprehensive and can be added to and edited by the user community.

26.1 CONDUCT AND REPORTING OF RESEARCH

Agarwal S et al. [136] have developed guidelines on reporting of health interventions using mobile phones.

The PRECIS-2 tool developed by Loudon K et al. [147] is intended for use at the design stage of a trial to help trialists make the purpose of their trial explicit and to ensure that their design choices are concordant with their intended purpose. It identifies when to use a highly pragmatic trial which would maximise applicability of the intervention to usual care across a range of local and distant settings and when a highly explanatory trial is appropriate.

In December 2015 a consultation commenced on a new UK Policy Framework for Health and Social Care Research which sets out principles of good practice in the management and conduct of health and social care research that takes appropriate account of legal requirements and other standards [http://www.hra.nhs.uk/documents/2015/12/uk-policy-framework-health-social-care-research.pdf].

Medical Research Council published a toolkit on planning trials, with a focus on patient and public involvement. It also provides a guide to national context for trials and links to other information sources [http://www.ct-toolkit.ac.uk/routemap/trial-planning-and-design].
The Medical Research Council has also produced guidance on the development and evaluation of complex interventions. This is intended to help researchers to choose appropriate methods, research funders to understand the constraints on evaluation design, and users of evaluation to weigh up the available evidence in the light of these methodological and practical constraints http://www.mrc.ac.uk/documents/pdf/complex-interventions-guidance/.

ClinicalTrials.gov is a registry and results database of research studies conducted in the United States and other countries. It has several thousand reports on studies of apps. Each record includes a summary of the study protocol, including the purpose, recruitment status, and eligibility criteria. Results are posted as they become available. More general information about clinical studies is also available https://clinicaltrials.gov/ct2/home.

The European trials registry only includes studies of medicines: https://www.clinicaltrialsregister.eu/about.html.

26.2 HEALTH AND SOCIAL CARE ORGANISATIONS

The Health Developer Network
The Health Developer Network is an NHS England website that contains technical information, software tools, source code and data for use by software developers who are designing and building software for use in health and social care. Available at: http://developer.nhs.uk/
**Academic Health Science** Networks

The 15 Academic Health Science Networks (AHSNs) use partnership working and collaboration between the NHS, academia, private sector and other external partners to identify, exploit and commercialise innovations that will have national and international significance. Each supports local apps developers through a range of tools such as networking, funding, education and training, advising on study design and can assist in links with the NHS and patient/public engagement. The AHSN Network identifies the work of each AHSN and is available at: [http://www.ahsnnetwork.com/](http://www.ahsnnetwork.com/).

DigitalHealth London (DHL) aims to speed up the use of new digital health technologies by bringing together clinicians, healthcare providers, research institutes, entrepreneurs and industry to give companies a clearer route to market based on the needs of patients and the NHS. It is a collaboration between MedCity and London’s 3 AHSNs: Imperial College Health Partners, UCLPartners and the Health Innovation Network. DHL provides financial support at different stages in the life cycle. It also offers support for evidence generation, for example, it offers ‘a wealth of health professionals who will be able to help build the case and run clinical trials where necessary on your behalf. Its website also has as section on standards, best practices and regulations in this area [http://digitalhealth.london/](http://digitalhealth.london/).

**NHS England**

NHS England has a number of teams and initiatives supporting innovation including:

- The Innovation Team, which aims to increase the rate of innovation to enable the NHS commissioning system to deliver better outcomes for patients and improve productivity. It works in partnership with industry, and technical and commercial experts to generate, identify, adopt and diffuse innovative ideas, products and technologies;
- Innovation Connect, which helps innovators in the health service and industry to realise their ideas, embed them into clinical practice and exploit new opportunities in international markets. It provides a fast-track for emerging healthcare innovations by overcoming barriers which innovators may encounter and puts them in touch with the right people and organisations;
- Innovation Compass, which aims to strengthen Leadership and Accountability for innovation at Board level throughout the NHS;
- NHS Innovation Accelerator (NIA), is a fellowship programme delivered collaboratively by NHS England, UCLPartners, The Health Foundation and Academic Health Science Networks. It aims to create the conditions and cultural change necessary for proven innovations to be adopted faster and more systematically through the NHS, and to deliver examples into practice for demonstrable patient benefit;
- NHS Innovation Challenge Prizes encourage, recognise and reward front line innovation and drive spread and adoption of these innovations across the NHS;
- The Regional Innovation Fund (RIF) is an annual awards scheme which provides resources to support and promote the adoption and spread of innovation across the NHS;
- Small Business Research Initiative for Healthcare runs a series of competitions for businesses to address major unmet health needs. Winners can receive 100% funding in excess of £1m and retain the intellectual property. More detail available at [http://sbrihealthcare.co.uk/](http://sbrihealthcare.co.uk/);
- Seven test beds have been set up to evaluate the use of novel combinations of interconnected devices such as wearable monitors, data analysis and ways of working which to help patients stay well and self-monitor their conditions at home.
No specific publications were identified in the search. More information is available at: https://www.england.nhs.uk/ourwork/innovation/

**NHS Innovations South East**
This body was established in 2004 to champion innovation throughout the NHS. It has an industry engagement consultancy service to support companies in the development, adoption and diffusion of innovative healthcare solutions. Relevant publications include “App Development: An NHS guide for developing mobile healthcare applications”. This covers the stages of developing an app including pre-development and testing, information on validation methods and references further studies for validation methods. It also covers MHRA guidance on how to identify which apps are classified as medical devices. It contains references on quality and security. Available at: http://innovationssoutheast.nhs.uk/files/4214/0075/4193/98533_NHS_INN_AppDevRoad.pdf

**Code4Health**
Code4Health is a programme managed by NHS England, which aims to educate and provide knowledge and skills to develop and implement high quality digital solutions. It provides information on learning to code, and has a platform for people who want to test their software, available at: https://code-4-health.org/.

**NHS Health Research Authority (HRA)**
The HRA aims to promote and protect the interests of patients in health research and to streamline the regulation of research. The website explains the single HRA Assessment which aims to align Research Ethics Committee approvals process with NHS R&D approvals. Its publication “UK policy framework for health and social care research” sets out principles of good practice in the management and conduct of health and social care research including legal requirements and other standards to facilitate high-quality research. Available at: http://www.hra.nhs.uk/documents/2015/12/uk-policy-framework-health-social-care-research.pdf.

**National Institute for Health Research (NIHR)**
NIHR aims to provide a health research system in which the NHS supports leading-edge research focussed on the needs of patients and the public. It commissions research directly and in partnership with other funding bodies such as Economic and Social Research Council. Its website provides information on conducting research and has links to external bodies that either regulate aspects of research, or from which permission needs to be sought to carry out research studies.

NIHR has funded ten regionally based research design services. These services together form a national Research Design Service (RDS), liaising with each other to develop common processes and a consistent service to the research community. Each service provides guidance to researchers on the formulation of workshops, research questions, study design, sources of funding and collaborators http://www.nihr.ac.uk/documents/about-NIHR/Briefing-Documents/3.4-Research-Design-Service.pdf.
Publications include:

- Clinical Trials Guide for Trainees;
- RDS patient and public involvement Handbook;
- Funding Opportunities Booklet.

All are available at: http://www.nihr.ac.uk/about/nihr-publications.htm.

NIHR also has a Clinical Research Network to help researchers set up clinical studies quickly and effectively and supports the life-sciences industry to deliver their research programmes. More details are available at https://www.crn.nihr.ac.uk/.

The South Central RDS provides an overview of many of the study designs referred to in this document including cluster randomised trials, cross-over trials and factorial design trials: http://www.rds-sc.nihr.ac.uk/planning-a-study/study-design/quantitative-studies/clinical-trials/1017-2/.

Other resources and links are available at: http://www.rds-sc.nihr.ac.uk/resources-and-links/.

Health and Social Care Information Centre

Section 27:
Section 28: Health and Social Care Information Centre (HSCIC) is England’s national provider of information, data and IT systems on a range of standards required such as open source software, safety and privacy. All separate section on
Royal College of Physicians

The Royal College of Physicians (RCP) is a multifaceted organisation, with a focus on developing physicians and improving safety and efficacy using clinical audits and service accreditation. It also contributes to national policy making.

Its publication “Guidance on using apps in clinical practice” provides guidance for doctors on the use of apps, includes information on what is considered a medical app. General advice is that apps that are not CE marked should not be used, and contains a link to MHRA guidance on how medical devices are regulated. Available at: https://www.rcplondon.ac.uk/guidelines-policy/using-apps-clinical-practice-guidance.
28.1 INDEPENDENT ORGANISATIONS

Healthcare Services Platform Consortium
The Healthcare Services Platform Consortium (HSPC) is a provider-driven organisation that aims to accelerate the delivery of innovative healthcare applications to improve health and healthcare. The website hosts several blogs e.g. there is an ‘Apps developer’s guide’ and provides links to many resources. Available at: https://healthservices.atlassian.net/wiki/display/HSPC/App+Developer’s+Guide.

Digital Health and Care Alliance
Digital Health and Care Alliance (DHACA) is a free-to-join, member-driven, organisation dedicated to improving interoperability and reducing duplication in health and care systems. It has a focus on apps and seeks to share best practice service designs, technologies and business models that improve outcomes. Publications include a ‘Medical apps – processes Guide - D019’ which describes the app commissioning process, and addresses procurement issues such as risk assessment, regulations and development, plus summarises the concerns that users have when seeking to select an app. Available at: http://dhaca.org.uk/wp-content/uploads/2014/11/DHACA-Medical-Apps-Process-Document-interim-final.pdf

Academy of Medical Sciences & Royal Academy of Engineering
The Academy of Medical Sciences is an independent body to promote medical science and its translation into benefits for society. Findings from a meeting on health apps address key issues including complexity of current regulation, vigilance and monitoring, obstacles to app use and uptake, generating & evaluating evidence of clinical utility, and software development practices are identified, but solutions are not given. The document also contains case studies of apps that have successfully navigated regulatory pathways. Available at: http://www.raeng.org.uk/publications/reports/health-apps-regulation-and-quality-control

Devices 4 Limited
Devices 4 Limited is a non-profit organisation, which works to provide health professionals with information on technology to encourage greater use of mobile devices and apps in the NHS to improve patient safety and outcomes. It has produced an overview of the regulation of health apps, which provides information for apps developers on: intellectual property rights, design, building and testing, managing liability, managing app use within a health organisation, recognising app risk, managing the app environment, mobile device management, promoting the use of apps across the organisation and limitations of public app stores. Available at: http://www.d4.org.uk/research/regulation-of-health-apps-a-practical-guide-January-2012.pdf
Knowledge for Health
Knowledge for Health is funded by the United States Agency for International Development and 1 of its purposes is to conduct research on knowledge management-driven global health interventions. Its publication “The mhealth planning guide: Key considerations for integrating mobile technology into health programmes” provides information on the planning process and outlines key considerations for planning, development and implementation of mhealth solutions. Available at: https://www.k4health.org/printpdf/book/export/html/20939.

The Healthcare App Network for Development and Innovation (HANDI)
HANDI is a non-profit company that provides support for developers of health and care apps. It has a community of clinicians, developers, health informaticians and others who will work with developers to support new apps. It runs an annual conference dedicated to health and care apps and lightweight digital tools. More details are at http://handihealth.org/.

28.2 REGULATORY BODIES

European Commission (EC)
The EC legislates to develop the legal framework for medicinal products for human use to ensure high standards of quality and safety of medicinal products, including devices. The agency authorising products is the European Medicines Agency. Publications potentially relevant to mHealth developers include:

- Green paper on mHealth which reports on potential issues with mobile health apps http://ec.europa.eu/transparency/regdoc/rep/1/2014/EN/1-2014-219-EN-F1-1.Pdf;
- ‘mHealth in Europe: next steps’ which follows on from the Green paper and contains polls conducted on the main issues identified with health apps http://ec.europa.eu/digital-agenda/en/news/mhealth-green-paper-next-steps;
- Guidelines to improve mHealth apps data quality are currently in process and are expected to be published by the end of 2016. These guidelines will aim to provide common quality criteria and assessment methodologies that could help users and developers in assessing the validity and reliability of mobile health apps http://ec.europa.eu/digital-agenda/en/news/new-eu-working-group-aims-draft-guidelines-improve-mhealth-apps-data-quality;
• Guidance on medical devices, some apps will be considered medical devices
• Medical devices guidance document: ‘Qualification and classification of standalone
game software’ and provides guidance on what is considered a medical device
u/DocsRoom/documents/10362/attachments/1/translations/en/renditions/native+&c
d=2&hl=en&ct=clnk&gl=uk;
• Manual on borderline and classification in the community regulatory framework for
  medical devices contains a chapter on software and mobile applications, with help on
classifying whether an app is a medical device
  http://ec.europa.eu/growth/sectors/medical-devices/specific-areas-
development/index_en.htm;
• Draft code of conduct on privacy for mobile health applications which covers user
  consent and transparency, data rights, data minimisation and privacy, security and
  confidentiality, transferring and uses of data
mhealth-apps.

**Medicines & Healthcare products Regulatory Agency**
The Medicines & Healthcare products Regulatory Agency (MHRA) regulates medicines,
medical devices and blood components for transfusion in the UK. Its publications include
“Guidance: Medical device standalone software including apps” which contains guidance on
what apps are considered medical devices and has links to other guidance. Available at:
https://www.gov.uk/government/publications/medical-devices-software-applications-
apps/medical-device-stand-alone-software-including-apps

**Information Commissioner’s Office**
The Information Commissioner’s Office (ICO) is the UK’s independent body set up to uphold
information rights. Publications include “Privacy in mobile apps: Guidance for app developers”
which considers the risks of apps collecting and transferring personal data and contains links
to further advice. Available at: https://ico.org.uk/media/for-
US Food and Drug Administration
The Food and Drug Administration (FDA) is responsible, amongst other things, for protecting public health by assuring that medicines, medical devices, vaccines and other medical products intended for human use are safe and effective. Its publications include:

- “Mobile Medical Applications – Guidance for industry and food and drug administration staff” which describes what is considered a medical app, and what types of app would be subject to regulation. It also lists regulatory requirements and provides links to further information on them. This document only applies to medical apps which would be classified as a device
  http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/HowtoMarketYourDevice/default.htm#step3;
- “FDASIA Health IT Report – Proposed strategy and recommendations for a risk-based framework” which is not specifically for apps but for health IT in general
  http://www.fda.gov/downloads/AboutFDA/CentersOffices/OfficeofMedicalProductsandTobacco/CDRH/CDRHHreports/UCM391521.pdf;
- “General principles of Software Validation; final guidance for industry and FDA staff”
  http://www.fda.gov/RegulatoryInformation/Guidances/ucm085281.htm;
- “Guidance for industry: Evidence-based review system for the scientific evaluation of health claims” which contains information on study designs and evidence based review systems for backing up health claims

28.3 STANDARDS DEVELOPMENT AGENCIES

British Standards Institute
The British Standards Institute (BSI) is an international business standards company that promotes excellence in business through the development of international standards. It is the UK National Standards Body. It recently published “PAS 277: Health and wellness apps: Quality criteria across the life cycle - Code of practice.” This guide sets out principles for app developers to ensure their products are high quality and fit-for-purpose. Recommendations are given for quality criteria of health and wellness apps, and the app project life cycle is covered including development, testing, releasing and updating of an app. Guidance available from:
European Association for Medical Devices of Notified Bodies (Team-NB)
Team NB is an association of 23 European safety bodies (BSI is the UK representative) which seeks to achieve high technical standards and good communications within the EC, industry users and other authorities in respect of medical devices. It provides information on what constitutes a medical device and on CE marking and classification of devices is available from:
http://www.team-nb.org/passport-to-europe-for-medical-devices/

International Organisation for Standardisation
The International Organisation for Standardisation (ISO) is an independent, non-governmental international organisation with a membership of 161 national standards bodies. It develops voluntary, consensus-based, market relevant International Standards (19,000 to date) that aim to support innovation whilst ensuring quality, safety and efficiency. Relevant standards include:

- IEC 62304:2006 which is a document on Medical device software life cycle processes. The set of processes, activities, and tasks described in this standard establishes a common framework for medical device software life cycle processes. Available to purchase from:
  http://www.iso.org/iso/catalogue_detail.htm?csnumber=38421
- ISO SO/TR 17522:2015 which applies to developing smart health applications and supporting new health businesses based on the smart devices. This Technical Report is to investigate the areas of ongoing developments and analyses of emerging interoperability standards for smart mobile devices. Available to purchase from:
  http://www.iso.org/iso/catalogue_detail.htm?csnumber=38421

International Health Terminology Standard Development Organisation
The International Health Terminology Standards Development Organisation determines global standards for health terms and developed SNOMED CT as the global common language for health terms. Available from: http://www.ihtsdo.org/.

The International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH)
The ICH issues standards which are accepted by most worldwide regulatory authorities. These address quality, efficiency, safety and other ad hoc topics such as data standards and electronic standards to transfer regulatory information. More information is available at:
28.4 PLATFORM DEVELOPERS

**Apple**


**Google**
Google Play, formerly the Android Market, is a digital distribution platform for applications for the Android operating system and an online electronics and digital media store, operated by Google. The site provides security tips for developers, covers storing data, content providers, requesting permissions, networking, and handling user data. Available at: [http://developer.android.com/training/articles/security-tips.html](http://developer.android.com/training/articles/security-tips.html)

Best practices for testing apps available at: [http://developer.android.com/training/testing/index.html](http://developer.android.com/training/testing/index.html)

**Microsoft App developer**

**References**


Section 29: Academic Health Science Networks

In spring 2013, 15 Academic Health Science Networks (AHSNs) received 5-year licences with the remit of putting innovation at the heart of the NHS, thereby improving patient outcomes and contributing to economic growth. A common feature is bringing together universities, industry and the NHS to achieve rapid clinical innovation adoption. Each has developed its own approach to partnership working, and there is a range of support available to app developers. This section advises on the strategy of two, East Midlands and Oxford AHSNs. These were selected because each has a close association with 2 of the case studies (Big White Wall and ChatHealth with East Midlands AHSN; and True Colours and GDm-health with Oxford AHSN). They also provide contrasting approaches to innovation adoption.

East Midlands AHSN has positioned itself to successfully deploy proven innovations throughout the East Midlands and nationally. Local Clinical Commissioning Groups (CCGs) and the AHSN have identified clinical priority areas including cancer, diabetes, mental health and patient safety. Innovators are invited to respond to calls for proposals to select new healthcare innovations in these priority areas and successful applications are supported. Exceptional innovations in other areas may also apply. Support includes:

- Undertaking real life evaluations, including recommending study designs and measurement tools;
- Providing health economics support to inform the business case to inform the CCGs purchasing decisions;
- Tailored advice and seminars to help navigate through the complexity of the healthcare system, identify CCGs’ needs, meet buyers etc.;
- Public and Patient Involvement which can link developers to local groups and third sector agencies to enhance user feedback;
- Health analytics and informatics which provide information on the population and their epidemiology;
- Advice on quality assurance testing, CE marking and other regulatory requirements;
- Project management and training on skills such as ‘Putting into Practice’;
- Patient and public involvement;
- Health economics and business case development;
- Communications;
- Healthcare system navigation;
- Access to research;
- Seed funding and signposting to other funding sources.
The AHSN may charge for some of these services. More information is available at: http://emahsn.org.uk/

The Oxford AHSN takes a more active role in generating real-world evidence to demonstrate the effectiveness of an app and other mobile health innovations, in addition to providing support for adoption of proven applications. It prioritises 3 attributes within the projects it supports, being clinical leadership, good data/evidence and excellent programme management.

It has plans to roll-out a testbed environment which implements apps, other new technologies and digital services across the region, enabling widespread, and rapid adoption. Aims include enabling robust evaluation, with endpoints capturing change in resource use along the pathway from primary, secondary and social care and hence providing the information to develop the economic and business cases, as well as measuring changes in the quality of health and social care and quality of life for patients.

In November 2014 it published a framework for digital technologies ‘Personalised Health and Care 2020: Using data and technology to transform outcomes for patients and citizens”. Actions include developing a community of interest in digital healthcare, to improve communications between innovators and clinicians, patients and commissioners. It has also produced a roadmap to help mobile health developers maximise patient benefits and commercial potential.

It also has a Clinical Innovation Adoption (CIA) programme which aims to identify the evidence based, high impact, market-ready innovations which it then seeks to adopt at scale across the region.

The Oxford AHSN undertakes a similar role to the East Midlands AHSN in offering a range of services to help support companies seeking clinical adoption for their products. This support ranges from signposting to intellectual property advice, health economics, market analysis studies, care pathway analysis, regulatory advice, and signposting to, and application support for, funding. Potential sources include various angel investor and venture capitalist groups. It can also facilitate access to patient and public involvement groups and clinicians during the development of the innovation. Some of these services may be charged for. More information is available at: http://www.oxfordahsn.org/
<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Applicability*</td>
<td>See “external validity.”</td>
</tr>
<tr>
<td>Association*</td>
<td>A relationship between 2 variables, such that as 1 changes, the other changes in a predictable way. A positive association occurs when 1 variable increases as another 1 increases. A negative association occurs when 1 variable increases as the other variable decreases. Association does not imply causation. Also called correlation.</td>
</tr>
<tr>
<td>Bias*</td>
<td>A systematic error in study design that results in a distorted assessment of the intervention’s impact on the measured outcomes. In clinical trials, the main types of bias arise from systematic differences in study groups that are compared (selection bias), exposure to factors apart from the intervention of interest (performance bias), participant withdrawal or exclusion (attrition bias), or assessment of outcomes (detection bias). Reviews of studies may also be particularly affected by reporting bias, where a biased subset of all the relevant data is available.</td>
</tr>
<tr>
<td>Blinding*</td>
<td>A randomised trial is “blind” if the participant is unaware of which arm of the trial he is in. Double blind means that both participants and investigators do not know which treatment the participants receive.</td>
</tr>
<tr>
<td>Case-control study*</td>
<td>A nonexperimental study that compares individuals with a specific disease or outcome of interest (cases) to individuals from the same population without that disease or outcome (controls) and seeks to find associations between the outcome and prior exposure to particular risk factors. This design is particularly useful where the outcome is rare and past exposure can be reliably measured. Case-control studies can be retrospective or prospective.</td>
</tr>
<tr>
<td>Causality*</td>
<td>An association between 2 characteristics that is demonstrated to be due to cause and effect (i.e., a change in 1 causes change in the other).</td>
</tr>
<tr>
<td>Causation</td>
<td>See “causality.”</td>
</tr>
<tr>
<td>Clinical trial</td>
<td>A clinical trial is a research investigation in a clinical setting, designed to supply data on the efficacy and/or safety of a drug, device, treatment or other healthcare intervention. Clinical trials may be sponsored by a manufacturer, governmental organisation, an academic research institute, or a body such as a charity. A trial can be conducted only after safety and ethics approval have been granted. Trials may involve healthy volunteers or patients, and their size should be determined by power calculations. Clinical trials are recorded in a variety of databases including ClinicalTrials.gov (USA), the EU Clinical Trials Register, and a range of national databases accessed by the WHO International Clinical Trials Registry Platform.</td>
</tr>
<tr>
<td>Cohort study*</td>
<td>A nonexperimental study with a defined group of participants (the cohort) that is followed over time. Outcomes are compared between subsets of this cohort who were exposed or not exposed (or exposed at different levels) to a particular intervention or other factors of interest. Cohort studies can either be retrospective or prospective.</td>
</tr>
<tr>
<td>Term</td>
<td>Definition</td>
</tr>
<tr>
<td>-------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Comorbidity*</td>
<td>A medical condition that exists simultaneously with another medical condition.</td>
</tr>
<tr>
<td>Comparative Effectiveness Research (CER)*</td>
<td>The generation and synthesis of evidence that compares the benefits and harms of alternative methods to prevent, diagnose, treat, and monitor a clinical condition or to improve the delivery of care. The purpose of CER is to assist consumers, clinicians, purchasers, and policy makers to make informed decisions that will improve healthcare at both the individual and population levels.</td>
</tr>
<tr>
<td>Comparison group*</td>
<td>See “control group.”</td>
</tr>
<tr>
<td>Confidence interval*</td>
<td>A range of values for a variable of interest, e.g., a rate, constructed so that this range has a specified probability of including the true value of the variable. The specified probability is called the confidence level, and the end points of the confidence interval are called the confidence limits.</td>
</tr>
<tr>
<td>Confounder*</td>
<td>See “confounding variable.”</td>
</tr>
<tr>
<td>Confounding variable*</td>
<td>A variable (or characteristic) more likely to be present in 1 group of participants than another that is related to the outcome of interest and may potentially confuse (confound) the results. For example, if individuals in the experimental group of a controlled trial are younger than those in the control group, it will be difficult to determine whether a lower risk of death in 1 group is due to the intervention or the difference in ages (age is the confounding variable). Confounding is a major concern in non-randomised studies. Also called confounder.</td>
</tr>
<tr>
<td>Control group*</td>
<td>Participants in the control arm of a study.</td>
</tr>
<tr>
<td>Correlation*</td>
<td>See “association.”</td>
</tr>
<tr>
<td>Cost benefit analysis</td>
<td>In healthcare evaluation cost-benefit analysis is a comparison of treatments (and their consequences) in which both costs and resulting benefits (including, but not restricted to health outcomes) are expressed in monetary terms. This enables 2 or more treatment alternatives to be compared using net monetary benefit, that is, the difference between the benefits of the treatments (expressed in pounds) less the costs of each (also expressed in pounds).</td>
</tr>
<tr>
<td>Cost-effectiveness analysis</td>
<td>Cost-effectiveness analysis is used to evaluate the effectiveness of an intervention relative to its cost, compared with current practice for the population of interest. The aim of the decision maker when assessing a new intervention is to maximise outcomes and minimise costs. Interventions that are both more effective at producing health benefits and are associated with net cost savings (i.e. any additional cost of the intervention is outweighed by the cost savings elsewhere) compared with current practice are said to be “dominant” and should (other things being equal) be implemented. Conversely interventions that produce fewer health benefits and are more costly overall are said to be ‘dominated’ and should not be implemented. In many situations an intervention is more effective and more costly than the relevant alternative. In these cases decision makers weigh up the additional costs against the additional health outcomes.</td>
</tr>
<tr>
<td>Cost minimisation analysis</td>
<td>Cost minimisation analysis is a method of comparing the costs of comparative treatments (and their consequences), which are known, or assumed, to have an equivalent medical effect. This type of analysis can be used to determine which of the treatment alternatives provides the</td>
</tr>
<tr>
<td>Term</td>
<td>Definition</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Cost-utility analysis</td>
<td>Cost-utility analyses are a type of cost-effectiveness analysis in which the cost per quality adjusted life year (QALY), or other unit of utility, is estimated. Two treatment methods are assessed by comparing the how many additional QALYs are gained and at what additional cost. Cost-utility analyses are frequently required by health technology assessment agencies, such as the National Institute for Health and Care Excellence (NICE) in the UK.</td>
</tr>
<tr>
<td>Covariate*</td>
<td>An independent variable not manipulated by the study that affects the response.</td>
</tr>
<tr>
<td>Economic evaluation</td>
<td>Economic evaluation is the comparison of aspects of different health strategies in order to aid decision-making about their future use. It encompasses a number of different types of widely used and discussed methodologies.</td>
</tr>
<tr>
<td>Effect modification*</td>
<td>A situation in which the measure of effect is different across values of another variable (e.g., the measure of effect is different across race, age, etc.).</td>
</tr>
<tr>
<td>Efficacy/ Effectiveness</td>
<td>The measure of an intervention to produce a result. Effectiveness refers to the ability of an intervention (drug, device, treatment, test, pathway etc.) to provide the desired outcome(s) in the relevant patient population. Efficacy is typically assessed through a clinical trial, effectiveness is tested in a real-world setting.</td>
</tr>
<tr>
<td>Endpoint*</td>
<td>See “outcome.”</td>
</tr>
<tr>
<td>Evidence-based medicine</td>
<td>Evidence-based medicine is the deliberate and explicit use of the current best evidence in combination with clinical knowledge and experience when making decisions on patient care, rather than basing clinical decisions solely on tradition or theoretical reasoning.</td>
</tr>
<tr>
<td>Experimental study*</td>
<td>A study in which the investigators actively intervene to test a hypothesis. It is called a trial or clinical trial when human participants are involved.</td>
</tr>
<tr>
<td>Explanatory trials*</td>
<td>A controlled trial that seeks to measure the benefits of an intervention in an ideal setting (efficacy) by testing a causal research hypotheses. Trials of healthcare interventions are often described as either explanatory or pragmatic. See also “pragmatic trial.”</td>
</tr>
<tr>
<td>External validity*</td>
<td>The extent to which results provide a correct basis for generalizations to other populations or settings. Also called generalizability, applicability. See also “applicability.”</td>
</tr>
<tr>
<td>Generalizability*</td>
<td>See &quot;external validity.&quot;</td>
</tr>
<tr>
<td>Health economics</td>
<td>Health economics is branch of economics concerned with issues related to efficiency, effectiveness, value and behaviour in the production and consumption of health and healthcare. In broad terms, health economists study the functioning of healthcare systems and health-affecting behaviour.</td>
</tr>
<tr>
<td>Homogeneous*</td>
<td>Having similarity of participants, interventions, and measurement of outcomes across a set of studies.</td>
</tr>
<tr>
<td>Hypothesis testing*</td>
<td>An objective framework to determine the probability that a hypothesis is true.</td>
</tr>
<tr>
<td>Incidence*</td>
<td>The number of new cases of an event that develop within a given time period in a defined population at risk, expressed as a proportion.</td>
</tr>
<tr>
<td>Incidence rate*</td>
<td>A measure of the frequency with which an event, such as a new case of illness, occurs in a population over a period of time. The</td>
</tr>
<tr>
<td>Term</td>
<td>Definition</td>
</tr>
<tr>
<td>-------------------------------------------</td>
<td>------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Term</td>
<td>Definition</td>
</tr>
<tr>
<td>denominator is the population at risk; the numerator is the number of new cases occurring during a given time period.</td>
<td></td>
</tr>
<tr>
<td>Internal validity*</td>
<td>The extent that the design and conduct of a study are likely to have prevented bias. More rigorously designed and conducted trials are more likely to yield results that are closer to the truth. See “bias.”</td>
</tr>
<tr>
<td>Masked*</td>
<td>See “blinding.”</td>
</tr>
<tr>
<td>Matching*</td>
<td>When individual cases are “matched” with controls that have similar confounding factors (age, sex, BMI, etc.) to reduce the effect of the confounding factors on the association being investigated.</td>
</tr>
<tr>
<td>Multivariate analysis*</td>
<td>Involves the construction of a mathematical model that describes the association between the exposure, disease, and confounders.</td>
</tr>
<tr>
<td>Nested case-control study*</td>
<td>A study where cases and controls are selected from patients in a cohort study (a case-control study “nested” within a cohort study).</td>
</tr>
<tr>
<td>New-user design*</td>
<td>A type of study that restricts the analysis to persons under observation at the start of the current course of treatment.</td>
</tr>
<tr>
<td>Nonexperimental study design*</td>
<td>A study in which investigators observe the course of events and do not assign participants to the intervention. Also called an observational study.</td>
</tr>
<tr>
<td>Observational study*</td>
<td>See “Nonexperimental study design.”</td>
</tr>
<tr>
<td>Odds ratio</td>
<td>An odds ratio is a measure of the excess risk of an event in a population compared with the risk in another population. When the populations are defined by intervention (but otherwise identical, as in a clinical trial) this gives a measure of the relative effect of an intervention. The odds ratio is the odds of an event occurring in the intervention group divided by the odds of the same event occurring in the comparison (control) group.</td>
</tr>
<tr>
<td>Outcome*</td>
<td>The result of an experimental study that is used to assess the efficacy of an intervention. Also called endpoint.</td>
</tr>
<tr>
<td>Patient registry*</td>
<td>An organized system that uses nonexperimental study methods to collect uniform data (clinical or other) to evaluate specified outcomes for a population defined by a particular disease, condition, or exposure, and that serves 1 or more predetermined scientific, clinical, or policy purposes.</td>
</tr>
<tr>
<td>Patient-reported experience measure (PREM)</td>
<td>Patient-reported experience measures (PREMs) are psychometrically validated tools (e.g. questionnaires) used to capture patients’ interactions with healthcare systems and the degree to which their needs are being met. PREMs are designed to determine whether patients have experienced certain care processes rather than their satisfaction with the care received. A PREM may, for instance, be used to collect information on the patient experience of hospital admission. Data derived from this could be used to inform service development and configuration.</td>
</tr>
<tr>
<td>Patient-reported outcome measures (PROMs)</td>
<td>Patient-reported outcome measures (PROMS) are psychometrically validated tools, such as questionnaires used to collect patient-reported outcomes.</td>
</tr>
<tr>
<td>Patient-reported outcomes (PROs)</td>
<td>Patient-reported outcomes (PROs) are any reports directly from patients on their health. PROs encompass single or multidimensional aspects of patients’ symptoms, health, quality of life, treatment satisfaction etc. PROs are often recorded in clinical trials, using validated instruments to measure the impact of the intervention as perceived by the patient.</td>
</tr>
<tr>
<td>Term</td>
<td>Definition</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Perspective</td>
<td>The perspective is the point of view adopted when deciding which types of costs and health benefits are included in an economic evaluation. The societal viewpoint is the broadest perspective, as this aims to reflect a full range of social costs associated with different interventions. For example, it includes productivity losses arising from patients’ inability to work. In England, most decision makers adopt a NHS and Personal Social Services perspective. This perspective includes NHS treatment costs such as medicine costs, GP visits, hospital admissions and all social care costs including home help and residence in care homes when paid by local authorities. Costs incurred by patient, and their families are excluded from this perspective.</td>
</tr>
<tr>
<td>Placebo*</td>
<td>A placebo is an inactive drug, therapy or procedure that has no treatment value. In clinical trials, experimental treatments are often compared with placebos to assess the treatment’s effectiveness.</td>
</tr>
<tr>
<td>Power*</td>
<td>The probability of rejecting the null hypothesis when a specific alternative hypothesis is true. The power of a hypothesis test is 1 minus the probability of Type II error. In clinical trials, power is the probability that a trial will detect an intervention effect of a specified size that is statistically significant. Studies with a given number of participants have more power to detect large effects than small effects. In general, power is set at 80% or greater when calculating sample size. Also called statistical power.</td>
</tr>
<tr>
<td>Precision*</td>
<td>In statistics, precision is the degree of certainty surrounding an effect estimate for a given outcome. The greater the precision, the less the measurement error. Confidence intervals around the estimate of effect are one way of expressing precision, with a narrower confidence interval defining a higher precision.</td>
</tr>
<tr>
<td>Prevalence*</td>
<td>The proportion of a population that is affected by a given disease or condition at a specified point in time. It is not truly a rate, although it is often incorrectly called prevalence rate.</td>
</tr>
<tr>
<td>Probability*</td>
<td>The probability (likelihood) of an event is the relative frequency of the event over an infinite number of trials.</td>
</tr>
<tr>
<td>Prospective study*</td>
<td>A study in which exposures are measured by the investigator before the outcome events occur.</td>
</tr>
<tr>
<td>P-value*</td>
<td>The probability (ranging from zero to one) that the results observed in a study (or more extreme results) could have occurred by chance.</td>
</tr>
<tr>
<td>Quality of life</td>
<td>Quality of life is a broad, multidimensional concept of an individual’s subjective evaluation of aspects of their life as diverse as their physical, social, spiritual and emotional well-being, as well touching on others areas such as their environment, employment, education and leisure time. A narrower definition is health-related quality of life (HRQOL), which reflects the impact a medical condition and/or treatment has on a patient’s functioning and well-being. HRQOL is increasingly being measured in clinical trials alongside other outcome measures to evaluate the effect of an intervention.</td>
</tr>
<tr>
<td>Quasi-experimental*</td>
<td>A study that is similar to a true experiment except that it lacks random assignment of participants to treatment and control groups. A quasi-experimental design may be used to reveal a causal relationship in situations where the researcher is unable to control all factors that might affect the outcome. Because full experimental control is lacking, the researcher must thoroughly consider threats to validity and uncontrolled variables that may account for the results.</td>
</tr>
<tr>
<td>Term</td>
<td>Definition</td>
</tr>
<tr>
<td>------</td>
<td>------------</td>
</tr>
<tr>
<td>Randomisation*</td>
<td>The process of randomly assigning participants to 1 of the arms of a controlled trial. There are 2 components to randomisation: generation of a random sequence and its implementation, ideally in such a way that those enrolling participants into the study are not aware of the sequence (concealment of allocation). The objective is to minimise differences between groups.</td>
</tr>
<tr>
<td>Randomised controlled trial (RCT)</td>
<td>A randomised controlled trial (RCT) is an experiment designed by investigators to study the efficacy and safety of at least 2 interventions in groups of randomly assigned subjects. The main value of randomisation is that the study groups are designed to be as near identical as possible in measurable and unmeasurable characteristics that may impact on the study outcomes. Where possible patients and investigators are 'blinded' as to which subjects are receiving which intervention. Patients are recruited using pre-specified inclusion and exclusion criteria. Ethical approval is required before a study can start. All aspects of the study are set out in a study protocol, including planned statistical analysis and a pre-specified criterion for success. An important distinction is made between 'explanatory' trials, which use very tight protocols and intense follow-up to measure safety and efficacy with the minimal possible opportunity for bias, and pragmatic trials which aim to mimic usual practice, with broader inclusion criteria and fewer protocol-related activities (and so generate results which may be considered more generalisable to routine practice).</td>
</tr>
<tr>
<td>Regression analysis*</td>
<td>A statistical modelling technique used to estimate or predict the influence of 1 or more independent variables on a dependent variable.</td>
</tr>
<tr>
<td>Residual confounding*</td>
<td>Confounding by unmeasured variables in a study.</td>
</tr>
<tr>
<td>Restriction*</td>
<td>Limiting of cohort entry to individuals with a certain range of values for a confounding factor (e.g., age, race, etc.) to reduce the effect of the confounding factor.</td>
</tr>
<tr>
<td>Retrospective study*</td>
<td>A study in which exposures are measured by the investigator after the outcome events have occurred.</td>
</tr>
<tr>
<td>Standard error*</td>
<td>The standard deviation of a theoretical distribution of sample means about the true population mean.</td>
</tr>
<tr>
<td>Type I error*</td>
<td>The error that results if a true null hypothesis is rejected or if a difference is concluded when no difference exists. Also called alpha error, false alarm and false positive.</td>
</tr>
<tr>
<td>Type II error*</td>
<td>The error that results if a false null hypothesis is not rejected or if a difference is not detected when a difference exists. Also called beta error, missed opportunity, and false negative.</td>
</tr>
<tr>
<td>Variance*</td>
<td>A measure of the dispersion shown by a set of observations, defined by the sum of the squares of deviations from the mean, divided by the number of degrees of freedom in the set of observations.</td>
</tr>
</tbody>
</table>

* Definitions extracted from Velengtas P et al. [11].
References


21. Stoner SA, Arenella PB, Hendershot CS. Randomized Controlled Trial of a Mobile Phone Intervention for Improving Adherence to Naltrexone for Alcohol Use Disorders. Plos One 2015;10:e0124613
of people with sub-acute or chronic low back pain—a multicentre, cluster-randomised, placebo-controlled, pilot trial. Bmc Musculoskel Dis 2015;16


37. Skrovseth SO, Arsand E, Godtliebsen F, Joakimsen RM. Data-Driven Personalized Feedback to Patients with Type 1 Diabetes: A Randomized Trial. Diabetes Technol Ther 2015;17:482-9


References 133


55. Armstrong KA, Coyte PC, Bhatia RS, Semple JL. The effect of mobile app home monitoring on number of in-person visits following ambulatory surgery: protocol for a randomized controlled trial. JMIR Res Protoc 2015;4:e65

Petrou S, Gray A. Economic evaluation alongside randomised controlled trials: design, conduct, analysis, and reporting. BMJ 2011;342:d1548


Belsi A, Papi E, McGregor AH. Impact of wearable technology on psychosocial factors of osteoarthritis management: A qualitative study. BMJ Open 2016;6 (2) (no pagination)


100. Teale EA, Young JB. A Patient Reported Experience Measure (PREM) for use by older people in community services. Age Ageing 2015;44:667-72


Lewis ZH, Lyons EJ, Jarvis JM, Baillargeon J. Using an electronic activity monitor system as an intervention modality: A systematic review. BMC Public Health 2015;15


Steinhubl SR, Muse ED, Topol EJ. The emerging field of mobile health. Sci Transl Med 2015;7:283rv3


