
**DOI link**
https://doi.org/10.1016/j.ijmedinf.2014.09.002

**ePrints link**
http://eprint.ncl.ac.uk/pub_details2.aspx?pub_id=244874

**Date deposited**
26/01/2018

**Copyright**
© 2014 The Authors. Published by Elsevier Ireland Ltd. This is an open access article under the CC BY-NC-SA license (http://creativecommons.org/licenses/by-nc-sa/3.0/).

**Licence**
This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License.
Review

Design of telehealth trials – Introducing adaptive approaches

Lisa M. Law*, James M.S. Wason

MRC Biostatistics Unit, Institute of Public Health, Forvie site, Robinson Way, Cambridge CB2 0SR, United Kingdom

ARTICLE INFO

Article history:
Received in revised form 16 January 2014
Accepted 5 September 2014

Keywords:
Telehealth
Telemedicine
Adaptive design
Interim analysis
Multi-arm trials
Sample size re-estimation
Enrichment designs
Group sequential designs

ABSTRACT

Background: The field of telehealth and telemedicine is expanding as the need to improve efficiency of health care becomes more pressing. The decision to implement a telehealth system is generally an expensive undertaking that impacts a large number of patients and other stakeholders. It is therefore extremely important that the decision is fully supported by accurate evaluation of telehealth interventions.

Objective: Numerous reviews of telehealth have described the evidence base as inconsistent. In response they call for larger, more rigorously controlled trials, and trials which go beyond evaluation of clinical effectiveness alone. The aim of this paper is to discuss various ways in which evaluation of telehealth could be improved by the use of adaptive trial designs.

Results: We discuss various adaptive design options, such as sample size reviews and changing the study hypothesis to address uncertain parameters, group sequential trials and multi-arm multi-stage trials to improve efficiency, and enrichment designs to maximise the chances of obtaining clear evidence about the telehealth intervention.

Conclusion: There is potential to address the flaws discussed in the telehealth literature through the adoption of adaptive approaches to trial design. Such designs could lead to improvements in efficiency, allow the evaluation of multiple telehealth interventions in a cost-effective way, or accurately assess a range of endpoints that are important in the overall success of a telehealth programme.

© 2014 The Authors. Published by Elsevier Ireland Ltd. This is an open access article under the CC BY-NC-SA license (http://creativecommons.org/licenses/by-nc-sa/3.0/).

Contents

1. Introduction .................................................................................................................. 871
2. Adaptive designs ......................................................................................................... 871
3. Uncertain parameters.................................................................................................... 872
   3.1. Sample size review ................................................................................................. 872
   3.2. Changing the clinically relevant effect ................................................................. 873

* Corresponding author. Tel.: +44 01223 330372.
E-mail address: lisa.law@mrc-bsu.cam.ac.uk (L.M. Law).

http://dx.doi.org/10.1016/j.ijmedinf.2014.09.002
1386-5056/© 2014 The Authors. Published by Elsevier Ireland Ltd. This is an open access article under the CC BY-NC-SA license (http://creativecommons.org/licenses/by-nc-sa/3.0/).
1. **Introduction**

Telehealth is the use of technology to allow communication of information between patient and care-provider whilst the patient is outside the clinical environment, e.g. in their own home. The range of telehealth is broad, from the monitoring of blood glucose levels in diabetics who provide measurements from home, to patients of mental illness receiving therapy treatment online. It can allow low-cost monitoring of the patient’s condition, which reduces demands on hospital resources and staff. Implementation of an effective telehealth intervention therefore has the potential to lead to improved patient care and improved efficiency in health care, which has been demonstrated in various studies [1-4]. However, widespread implementation of telehealth faces barriers due to a lack of strong, valid evidence of its benefits. This lack of evidence is discussed in Ekeland et al’s systematic review of reviews, which examined up to 80 telehealth reviews in two separate papers [5,6]. According to their findings, telehealth literature suffers from several flaws. Firstly, many of the trials carried out are too small, lacking the power to find a benefit over standard care. Second, a lack of consistency in endpoints in the literature makes it difficult to consolidate evidence. Further, there is a belief that clinical effectiveness alone is not sufficient evidence for implementing a telehealth system. Some reviews call for studies to include technical, ethical and organisational effects, including direct and indirect costs to patients and care-providers. Lastly, several reviews discuss the heterogeneity that can exist in patient responses to the same telehealth intervention, contributing to inconsistency in the evidence base. There is a lack of knowledge about why some people respond or comply well with a telehealth programme but others do not. Suggested explanations can be as objective as the patient’s personal sense of motivation. Reviews recommend exploring reasons for varying response and compliance if the telehealth intervention is to be successful.

It may be possible to address the criticisms and issues discussed in reviews of telehealth through appropriate trial design. Ekeland et al. [6] discuss methodology used in telehealth studies, and recommend standard randomised controlled trials as the most rigorous option of trial design. In this paper we argue that the evaluation of telehealth interventions could benefit from expanding this usual approach to include adaptive trial designs. The potential advantages of adaptive designs in drug trials have been examined in several places in the literature [7,8]. The aim of this paper is to introduce adaptive designs in the context of telehealth, taking into account the specific challenges that may arise when evaluating telehealth interventions.

2. **Adaptive designs**

An adaptive design is one in which new decisions can be made about the design or progress of the trial, once the trial is already underway. This is in contrast to standard, non-adaptive designs, in which all aspects of the trial are fixed beforehand and remain fixed throughout. The changes made in an adaptive trial are based on interim analyses of the data observed so far. Examples of changes that may be made are: stopping the trial early if efficacy or futility has already been demonstrated; changing the endpoint or objective of the study, and; re-estimating the sample size. Potential changes are usually considered beforehand and incorporated into the design stage of the study, so that if these changes do need to be made, any increase in resources required is known in advance. At the end of the trial, the final analysis includes all patients recruited since the beginning of the trial. Thus the
patients observed in the early stages contribute to optimising the remainder of the trial, as well as to the final analysis. Adaptive designs have the potential to improve efficiency, power, flexibility and ethical value of a trial. So far, however, the potential of adaptive designs in telehealth studies has not been explored. We confirmed this by performing a literature search, the results of which can be seen in Table 1. Search terms for telehealth were chosen based on the range of terminology discussed in a recent review [9], and the remaining search terms were words likely to be used when describing methodology of an adaptive design. Of 156 results returned, most were found to be irrelevant upon reading their abstracts. Of the rest, some used adaptive designs but the intervention was not considered a genuine telehealth system, while others studied genuine telehealth systems and made changes at some point in the trial, but the changes did not affect the overall study design and were not considered adaptive in the strictest sense. The literature search, therefore, revealed a research gap for adaptive designs in the context of telehealth.

In the remainder of this paper, we will review the issues that can arise in telehealth trials and suggest how adaptive designs could be used to address those issues. Generally, our suggestions are made in the context of improving randomised controlled telehealth trials. Single-arm intervention trials are typically only suitable when substantial information is available on outcomes in individuals without the intervention. However, if such a scenario existed for a telehealth intervention, then many of the following adaptive designs could still be applied in a single-arm study. Similarly, adaptive designs may be useful for observational studies or other types of experiment.

3. **Uncertain parameters**

When planning a standard, fixed-sample randomised controlled trial, the parameters used to compute the required sample size include the desired power and type I error rate and the value of the intervention effect that would be clinically relevant (the clinically relevant effect, CRE). Also required is a measure of uncertainty in the outcome, which will be different depending on the type of outcome being evaluated. For example, for continuous outcomes the standard deviation of the outcome is required, for binary outcomes it is the probability of success without the intervention, or for survival outcomes it is the event rate. In a telehealth trial, the CRE and the uncertainty in the outcome can be difficult to estimate in advance. They may be difficult to obtain due to the expense of running a pilot telehealth trial. Further, in a drug trial, clinicians may be able to refer to previous, similar trials to help form an estimate of the parameters in their data. For telehealth, however, similar studies rarely exist and it is difficult to estimate the potential impact a telehealth system may have. Telehealth interventions can be complex and the underlying mechanisms which lead to successful outcomes may not be as well understood as the physiological mechanisms that contribute to successful drug treatments. Misspecification of parameters in a sample size calculation will result in a sample size that is under- or over-estimated. This will lead to inefficient scenarios in which there are too few patients and so not enough power to detect the desired intervention effect, or too many patients, which is unnecessarily expensive. There are several options of adaptive design in which an interim analysis takes place after some data has been observed. This allows the previously estimated parameters to be adjusted according to the new information.

3.1. **Sample size review**

A sample size review allows the opportunity to recalculate the sample size at an interim stage of the study, either at a pre-specified time, or after a pre-specified number of patients has been observed. Interim data, i.e. the data collected so far, is used to estimate the parameters of the data and these are used in a new sample size calculation. Subsequent recruitment plans are adjusted to reflect the new sample size, and

<table>
<thead>
<tr>
<th>Table 1 – Results of literature search for use of adaptive designs in telehealth studies.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>No. of results</strong></td>
</tr>
<tr>
<td>(total = 156)</td>
</tr>
<tr>
<td>144</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>1</td>
</tr>
</tbody>
</table>

a PubMed search terms: (telehealth OR telecare OR telehomecare OR telemedicine OR e-health OR ehealth) AND (adaptive OR interim OR sequential).

b With respect to the use of adaptive designs.
so the trial achieves the desired power and type I error rate. Various methods of sample size re-estimation exist, including procedures to avoid unblinding and procedures using conditional power. Gould [14] discusses these methods, along with issues involved in sample size re-estimation and situations in which they may not be suitable, for example if the recruitment process is particularly complex.

3.2. Changing the clinically relevant effect

The CRE is the effect that the trial is powered to detect. In some cases it represents the size of the intervention effect which would be important enough to make an impact in the clinical setting. In other cases any improvement at all may be of interest. The CRE is specified in advance in order to compute a sample size. Due to the difficulty of anticipating the impact of a telehealth system, the effect of the telehealth intervention observed at the interim may be far from what was expected. In this case there would be little point in continuing with the trial as it was originally planned, as the trial was powered to find a different effect. Instead, a more realistic CRE should be chosen. Then a sample size review may be carried out, incorporating this new value. The remaining recruitment would be adjusted to achieve the new sample size. Subsequent analysis of the study may need to be adapted in order to deal with the possible effects of observing the intervention effect at the interim, such as an inflated type I error rate and biased intervention effect. Such effects are discussed in a paper by Hung et al. [15], which centres on the situation where the treatment effect observed at an interim stage is far off target in terms of the original estimate. They discuss several ways in which statistical inference may be adjusted following a change in the clinically relevant effect.

4. Multiple endpoints

A complex telehealth system may involve several interacting components and is likely to impact on more than one aspect of health care. For example, the Whole Systems Demonstrator study [16] is a large randomised controlled trial to investigate a telehealth device for monitoring symptoms of patients with chronic diseases and offering them advice to manage their own health. The study aims to evaluate various endpoints, including health service utilisation, cost of care, symptom reduction, quality of life and well-being of the patient, and burden on care-provider. In such a range of endpoints it may be difficult to choose the primary endpoint of interest. If it became apparent in a study, that the intervention was not affecting the primary endpoint, but was having a greater effect on a secondary endpoint, it may be desirable to change the endpoint of interest entirely.

4.1. Changing the endpoint of interest

Changing the endpoint, i.e. changing the study hypothesis, at an interim stage of the study carries some risk. Observing the intervention effect at this stage can introduce bias into the later results of the study, multiple testing may inflate the type I error and the flexibility of being able to change the endpoint means failing trials could be manipulated into becoming successful [17]. However, with careful planning and design these risks may be minimised. Changing the hypothesis has the potential to gain more meaningful data about the intervention being studied, rather than continuing the study with the same endpoint, which may be wasteful and non-informative. Several procedures to handle bias and control of type I error rate are discussed in a paper by Hommel [18].

Alternatively, the study may plan to assess more than one endpoint from the beginning. Previous reviews recommend that telehealth trials look at the impact of telehealth on a broad range of areas. Therefore, a well-designed telehealth trial will almost certainly involve multiple endpoints. When conducting a trial that assesses multiple endpoints of equal importance, it is important to control the type I error rate whilst testing multiple hypotheses.

5. Cluster randomisation

Cluster randomisation is sometimes chosen for evaluating telehealth, such as in the Whole Systems Demonstrator study [16]. This is because often the telehealth intervention involves implementing an entire system of changes, and this is much easier to do across a whole health unit rather than attempting to expose particular individuals. Adaptive methods will have to take into account characteristics of clusters, such as within- and between-cluster variability and randomness in cluster size.

5.1. Sample size review for cluster randomised trials

Lake et al. propose methodology to allow a sample size review to take place in a cluster randomised trial [19]. Planning the sample size for a cluster randomised trial requires reliable estimates of within- and between-cluster variability and variability in cluster size. These estimates are difficult to estimate in advance as it is unlikely that previous studies will exist that examined the same clusters. A solution to this is to conduct a sample size review at the interim, as discussed in a previous section. Lake et al’s approach is particularly suited to trials where the cluster sizes are small, e.g. family units, and where recruitment will take place over an extended length of time. The telehealth setting is more likely to have larger sizes of clusters, e.g. GP practices. Extra consideration may need to be given to whether a multi-stage study would be appropriate in this case.

6. Efficiency and ethics

Efficiency is important in telehealth trials because of the cost involved in evaluating a telehealth system. It is also important that trials are conducted in the best interest of the patient, i.e. ethically. As in drug studies, it would be unethical to continue with a treatment that is having adverse effects, or to deny control participants a treatment that was showing a clear benefit. Several design options are available that are ethical and aim to increase efficiency.
6.1. **Group sequential design**

A group sequential design includes one or more interim analyses which offer the opportunity of stopping the trial early, either because the intervention has already been shown to work (stopping for efficacy), or because the evidence suggests the intervention does not work (stopping for futility). Stopping rules are derived for each interim, in which a test statistic beyond a specified level means there is already significant evidence of an effect and recruitment to the study does not need to continue. Likewise, if the effect is negative and falls below a specified threshold then there is significant evidence that the intervention does not work and the trial must stop. Otherwise, the trial continues until the next interim, in which the decision will be made again using more data. The main advantage of such a trial is that the expected sample size is generally smaller and so is less costly on average. It is also, arguably, the more ethical option, because if the intervention is unsuccessful then fewer patients have been exposed to it, and if the intervention is successful, then the trial ends sooner and the patients on standard treatment can be switched to the superior treatment. The stopping rules are essentially boundaries with which the observed intervention effect is compared at each interim analysis. Research into various types of stopping rule has been on-going for many years. Stopping rules and general methodology of group sequential trials are thoroughly explored in books by Jennison and Turnbull and by Whitehead [20,21]. Examples of common designs are illustrated in Fig. 1, in which the boundaries indicate where interim test statistics must fall in order for the study to be stopped with sufficient evidence. Panel A of Fig. 1 shows a design by O’Brien and Fleming [22]. Its shape has the advantage of allowing early stopping when initial results are extreme. Panel B of Fig. 1 shows the triangular design by Whitehead and Stratton [23]. The shape of this design is suited for studies in which it is not expected that the trial will stop for futility, but rather it is more important to allow stopping for efficacy. This may be suitable for evaluating telehealth interventions in which causing harm is not reasonably expected. Such an intervention could be one in which the telehealth programme involves additional monitoring on top of usual care. For a more experimental telehealth programme that replaces usual care, a group sequential design would need to allow for the intervention being harmful, and thus have suitable futility boundaries.

More recently, Wason et al. [24] showed that efficiency of multi-stage trials could be further enhanced. Their paper aims to minimise the maximum expected sample size of the overall study.

6.2. **Multi-arm multi-stage design**

The complex nature of telehealth interventions means that there could be many design or implementation possibilities. It may be useful to evaluate several versions of a telehealth intervention at the same time. In fact, this could be more useful than a comparison between a single telehealth programme and usual care as it may already be fair to anticipate that a telehealth programme that involves extra patient monitoring will lead to better outcomes than usual care. It is sensible, then, to search for the best telehealth programme among a number of options. An adaptive trial that is suitable for this situation is a multi-arm, multi-stage trial. The multi-arm feature means that patients are allocated to a control group or one of a number of interventions. Compared to carrying out separate two-arm trials to evaluate each intervention, this has the advantage of increased efficiency as only one control group is required. The multiple stages allow interim analyses to take place, after which decisions can be made on how the study should proceed. These decisions could be: (1) to drop arms for futility, or stop early for efficacy if an effective treatment is found, as in group sequential designs; (2) to drop the intervention that is performing the worst; a so-called drop-the-loser design, or; (3) change the allocation ratio for randomisation so that more patients are allocated to the best performing interventions, known as adaptive randomisation. These three approaches are discussed further in an article by Wason [8]. In the case of the telehealth intervention being expensive compared to the control, further savings could be made by changing the allocation ratio such that a higher proportion of patients is recruited to the control group, compared to each of the telehealth intervention groups. This may be done whilst maintaining the level of statistical power [25].

6.3. **Seamless phase II/III designs**

Drug treatments go through several stages of development, known as phases. After evaluating safety in a small phase I study, a trial is carried out to see if the drug is effective (phase II), often by evaluating a short-term endpoint related to survival. This is followed by a larger trial to confirm efficacy (phase III). Phase III trials are generally more expensive than phase II and so there should be reasonable evidence from phase II before the drug progresses to phase III. Each trial is independent, i.e. it does not use the data observed in the previous trial. It would be more efficient to make use of the data collected in the early stages, in later trials. A seamless phase II/III design aims to carry out one uninterrupted study that incorporates the exploratory aspects of a phase II trial and the confirmatory aspects of a phase III trial [26]. The study would contain a planned interim analysis where a decision is made as to whether the treatment should continue into the next stage of the study, or, if multiple treatments are being studied, which treatment should progress further. In other words, results observed during the phase II stage can be used to shape the design of the phase III stage. Although telehealth technology does not have the formal development process that drug treatments have, the principles of a seamless phase II/III design could be applied in relevant contexts to add efficiency to the evaluation of telehealth. For example, if we liken pilot studies and subsequent, substantive telehealth studies to phases II and III, then it becomes clear to see that the same approach could explore the potential of several options of telehealth programme in the beginning of a trial, and then take the most promising programme forward for the remainder of the trial, following an interim analysis. This has the advantage of requiring fewer patients and taking less time than two separate trials. This is very similar to the multi-arm multi-stage design discussed in the previous section, in which decisions can be made about several interventions at the interim.
7. Compliance and usability

Reviews on telehealth often recommend studying compliance to and usability of telehealth programmes, reasoning that the success of the telehealth intervention is just as dependent on these factors as on clinical benefit [27]. Non-compliance may arise if the telehealth programme is too demanding or time-consuming [28]. It may also arise because of poor usability. For example, Johnston et al. [29] point out that an intervention that relies on the computer literacy of the user may not be suitable for groups such as the elderly, who are less likely to have these skills.

Evaluating clinical outcomes of telehealth whilst not being aware of the compliance rate among the patients may lead to a biased estimate of the intervention effect. Non-compliance could also reduce the power of the study, as there are less people representing the intervention group. Adaptive approaches to trial design may help address these issues whilst at the same time evaluating effectiveness for clinical outcomes.

7.1. Multi-arm multi-stage, or seamless phase II/II designs

These designs could be useful for telehealth systems for which there is uncertainty about how well patients will adhere to the intervention. Several versions of the intervention may be evaluated in the first stage of the trial, in terms of patient compliance. The best performing interventions may be taken forward to the next stage, where their clinical efficacy is evaluated. The rationale is that there is little point in trying to evaluate clinical efficacy if the patients are not complying with the intervention, perhaps because of usability issues or because it is too demanding. To our knowledge, no one has yet considered a trial design which incorporates both non-compliance, in an exploratory stage, and clinical efficacy in later stages. A trial that achieves this may be particularly useful for evaluating telehealth and should be researched further.

8. Heterogeneity among responses to telehealth

Another issue discussed in the telehealth literature is that of heterogeneity of the intervention effect, across subgroups of patients. Reviews reported a lack of knowledge about why patients can respond differently to the same intervention. Sometimes this is a compliance issue. For example Azar et al. [30] discuss how the success of a telemonitoring programme in diabetics was partly dependent on the personal motivation of the participants to adhere to the programme requirements. The subjective nature of this heterogeneity could be difficult to address. In other trials, however, the source of the heterogeneity may be less subjective. For example, Bewick et al. [31] showed that web-based interventions designed to reduce alcohol consumption had varying effect sizes depending on the level of alcohol consumption at baseline. Some interventions only became effective beyond a certain level of alcohol use. Gathering information on which subgroups will experience a treatment benefit is useful for both running a trial and implementing the system in practice. In a trial, including participants in whom the intervention is not likely to work weakens the effect to be evaluated. Instead, trials could target the subgroup of patients in whom the intervention is likely to work. Trials which do this are called enrichment trials, as...
the recruitment is designed to enhance the chances of finding a significant intervention effect. Enrichment trials are risky when information on which subgroups are likely to respond well is unreliable. For example, it may be that the intervention works in a larger than anticipated group of patients. To address this problem, an adaptive enrichment trial may be more suitable.

8.1. Adaptive enrichment design

An adaptive enrichment trial aims to gain information about which patients respond well in the initial stage of the trial, and then uses this information during the recruitment for the later stage of the trial [32]. At an interim analysis, the adaptive enrichment trial may change the inclusion/exclusion criteria of its participants. The first stage of the trial will require the collection of information that is potentially useful in indicating the probability of success of the intervention. In drug trials, this usually involves the measurement of biomarkers related to the effects of the drug. In telehealth trials, a problem may lie in deciding which variables should be measured, i.e. deciding which variables are potentially predictive of a successful response to the intervention. Fig. 2 illustrates the concept of an adaptive enrichment design. It is based on a scenario in which prior to the study it is suspected that a particular biomarker will be associated with response to the intervention. We will define the participants as “biomarker-positive” or “biomarker-negative”, but designs in the continuous biomarker case are also possible. Initial recruitment to the study is open to all participants. After a follow-up period an interim analysis takes place, in which the response to the intervention is compared in the two groups of participants. If there is an indication that the intervention works in all patients, the trial continues to recruit all patients. Otherwise, recruitment may be restricted to one of the subgroups, or the trial may be terminated for futility. These scenarios are numbered one to three in Fig. 2. The figure demonstrates when the study continues with the same recruitment strategy (scenario 1), when it continues with restricted recruitment (scenario 2) and when it must be stopped due to a lack of success (scenario 3). We can relate Fig. 2 back to Bewick et al’s alcohol use study described earlier [31]. An adaptive enrichment approach to their trial may have revealed the difference in successful intervention between those with high alcohol consumption and those with low alcohol consumption. The remainder of the study could then have focussed on the more responsive subgroup, to obtain evidence for the intervention more efficiently.

The importance of targeting appropriate patients was discussed in a qualitative study of the Whole System Demonstrator trial. The Whole Systems Demonstrator trial evaluated a telehealth service for chronic disease sufferers, enabling them to take measurements of their symptoms, which would be monitored remotely by healthcare staff. The qualitative study, by Sanders et al. [33], conducted interviews with individuals who had declined to participate in the Whole System Demonstrator study, or who had withdrawn after the study began. Sanders et al. revealed several common reasons for non-participation in those interviewed, including belief that they would have difficulty using the technology; negatively

---

Fig. 2 – Visual representation of adaptive enrichment design. Following the first stage of the study, the figure shows three possible scenarios that may be observed in the results of the interim analysis, and will dictate how the study progresses.
associating telehealth with a lack of independence; the feeling that the frequent use of the telehealth service would make their illness seem too present in their lives; not wanting to interfere with existing services, which they felt comfortable with, and; the pressure of having to take their own measurements every day. Anticipating non-participation for these reasons, in the same way we have described in the adaptive enrichment design in Fig. 1, may not be possible unless there are baseline covariates that are suitably predictive. Nonetheless the presence of such patients in a study is important as it may skew the effect of telehealth on clinical outcomes. If enrichment designs could not be applied, several versions of a telehealth system could be evaluated in the first stage of a multi-arm multi-stage trial, as described in the previous section on compliance. Whatever is most successful in terms of compliance or satisfaction may be taken forward for further study, to examine its clinical benefits.

9. Monitoring data to predict adverse events

One innovative type of telehealth involves frequent and regular monitoring of patient symptoms, with the aim of predicting an adverse event in that patient. For example, the system could be designed to measure a continuous variable and if the variable crosses a pre-specified threshold then this indicates an increased likelihood of an adverse event. The measurement could also be a risk score, computed via the combination of several variables. In either case, once the monitored data indicates high risk of an adverse event, care-providers will respond to the patient. For example, the telehealth system in the Whole Systems Demonstrator study contains an alert system of several levels, which indicates to the healthcare staff what action they should take to deal with the reported symptoms. The purpose of such systems is usually to reduce the number of hospital admissions that may result when adverse events are not dealt with immediately. Deciding when alerts should be made may be difficult. Having too high a threshold may miss patients who genuinely require attention, whereas too low a threshold could result in saturating the health care services with patients whose needs were not urgent. The alert system could be seen as a kind of diagnostic test, aiming to identify people who require admission to hospital. In this case, the aim of a trial would be to assess whether the system correctly identified patients who required an admission to hospital.

9.1. Group sequential designs for diagnostic tests

These trials are the same as the group sequential trials described in an earlier section. One or more interim analyses take place to allow the opportunity to stop early if there is already evidence to demonstrate an effect or lack of effect. With diagnostic tests, the intervention effect to be tested will be related to the sensitivity and specificity of the prediction system, e.g. area under the ROC curve (AUROC). Methodology for this case is demonstrated in a paper by Mazumdar and Liu [34].

It may be the case that the outcome being predicted by the telehealth device is prevented through appropriate action taken by healthcare providers. In this situation, it would not be possible to determine whether the patient would have definitely experienced that outcome. Without being able to define the true status of the patient, measures of specificity and sensitivity cannot be computed and so diagnostic methods are irrelevant. Instead, a trial may look at the impact of the prognostic telehealth device on healthcare utilisation, such as number of hospital admissions. Group sequential trials may still be applied, where at each interim the decision can be made as to whether there is evidence for a clinically relevant difference in hospital admissions between the telehealth and control arms.

9.2. Adaptive enrichment design

As evaluation of telehealth technology can be expensive, it may not be cost-effective in a study regarding adverse events to include participants who are very unlikely to experience such an event. An adaptive enrichment study may be useful to discover the high-risk individuals and use this information to restrict the inclusion criteria to the high-risk individuals for the remainder of the trial. The same methods discussed in the previous section may be applied in this context.

10. Discussion

We have discussed ways in which adaptive designs have the potential to increase the accuracy and efficiency of telehealth trials. Adopting these approaches may strengthen the evidence to justify implementation of a telehealth system. Benefits over a standard, non-adaptive trial include: better estimation of the parameters required to adequately power the study; flexibility to change the study hypothesis if the impact from telehealth is not as expected; increased efficiency due to the option of stopping early if sufficient evidence has been obtained, and; the ability to evaluate several interventions at once and choose which ones should be studied further. More accurate evaluation could avoid inconsistencies across trials, as in some of the current literature.

We have described characteristics of telehealth which make it a promising area for utilising adaptive designs. Firstly, the complexity of telehealth interventions and their interaction with multiple endpoints calls for a non-standard approach. Secondly, multiple elements of the telehealth programme, along with varying conditions between health care sites, means that there may be more than one option for how telehealth is implemented, in terms of staff roles, organisation of data monitoring and adjustments to existing care procedures. In a telehealth edition of the Health Service Journal supplement, the Department of Health discusses the necessity of an overall service re-design alongside the addition of telehealth [35]. Adaptive, multi-arm trials could help choose between implementation options. Thirdly, anticipating the impact of a telehealth system may be difficult. Feedback during a trial could lead to the desire to make
changes, which would be problematic if the study protocol did not make provisions for changes. This was an issue in the Whole System Demonstrator trial, after which an organisational analysis discussed the frustration of gaining important knowledge about service integration throughout the trial but not being able to take action until the trial was completed [36]. Use of adaptive designs could allow flexibility to make changes to the trial, provided that the possibility of changes was thoroughly considered at the design stage. Lastly, the quality of existing telehealth studies in the literature has been called into question in several recent reviews. Adaptive designs offer efficient ways to obtain evidence more rigorously.

However, there will be some types of telehealth intervention which have characteristics unsuited to an adaptive design. For example, if long-term endpoints are of interest, and there is no short-term endpoint that is informative, then adaptive designs that try to achieve a lower expected sample size will do little to add efficiency. This is because such a trial offers the opportunity to stop early and avoid further recruitment to the trial. But by the time a long-term endpoint is observed, the majority of the planned sample size is likely to have been recruited already and so there is little opportunity to make changes in the interim. There are non-adaptive methods we have discussed, such as a multi-arm trial without multiple stages, which may still add efficiency over a standard parallel group design.

In all of the design options described in this paper, special considerations may need to be given to particular characteristics of telehealth studies. For example, a group sequential design could be chosen in order to provide the opportunity of stopping early, but it is likely that the telehealth intervention will involve other factors to be addressed, such as patient compliance, multiple endpoints or cluster randomisation. Consequently, the best approach could be a combination of several types of design described in this paper. This may require methodological adjustments of the single approaches.

The main challenge in using adaptive trials for the evaluation of telehealth interventions is the added complexity to the design and analysis process. One example is the stringent need to maintain the type I error rate. This could act as a deterrent to proponents of telehealth who are not confident with the statistical methodology. The potential advantages, however, are evident when considering the criticisms of previous trials. The fact that so many reviews have commented on the lack of consistent evidence justifies a more rigorous approach being considered, regardless of its complexity. Further, regulators look favourably on the merits of adaptive designs, and the Food and Drug Administration (FDA) have prepared specific guidelines to assist researchers in planning such studies [37]. Even so, a necessary aspect of choosing adaptive designs over more standard designs will be the inclusion of a statistical expert alongside the investigators. Furthermore, members of review boards and data monitoring committees will need to have sufficient knowledge of the adaptive design being used in order to make informed decisions on the ethics of the study. This is out of the hands of the investigator; we can only hope that continued education and discussion such as in this paper will increase the awareness of adaptive designs among members of research committees. Provided that review board members consist of members who are acquainted with the methodology, the monitoring of an adaptive design will often be more straightforward compared to a standard trial. This is because potential changes to the trial - for example, sample size recalculations, changes in recruitment criteria - will have been incorporated into the design stage of the study and will have pre-specified decision rules. When the trial is approved, the potential changes are also approved, such that if it is decided that the changes should be carried out at the interim, these modifications do not need to be sent for review again. Of course, this will be true only for the changes that have been pre-specified. It may be the case that unforeseen modifications are required, as in a standard, non-adaptive study.

The FDA is also positive regarding Bayesian approaches to evaluation of interventions. Although they have not been discussed in this paper, Bayesian methods have the advantage of incorporating prior knowledge into the design and analysis of trials. As discussed, such prior knowledge is often absent for telehealth technologies, but there could still be a place for Bayesian approaches in the adaptive designs we have described, as they accumulate information throughout the study. For many of our suggested designs, the addition of Bayesian features has been proposed in the literature and so there is potential to explore the use of Bayesian methods in telehealth studies.

11. Conclusion

Previous methodological approaches to the evaluation of telehealth interventions have contributed to inconsistent evidence in the literature. We suggest approaching telehealth trial design with an adaptive framework. Although this will likely involve further research into modifying current drug-based trial designs, the added flexibility of adaptive trials has the potential to produce more accurate and efficient studies, with clearer evidence to support decisions for implementation.

Author contributions

LML contributed to the conception and text of the paper and performed the literature search. JMSW contributed to the conception and text of the paper.

Conflict of interest statement

The authors have no conflicts of interest to declare.

Acknowledgements

This work was funded by an EPSRC iCase studentship and by the MRC (grant number G0800860).
Summary points
What was already known on this topic:

- Telehealth programmes have the potential to add efficiency to the provision of healthcare and improve patient outcomes through increased monitoring of symptoms and self-management of health.
- Supporting the implementation of telehealth interventions on a large scale is difficult due to inconsistent evidence across studies.

What this study added to our knowledge:

- No investigator of telehealth has yet carried out an adaptive trial design for the evaluation of a telehealth intervention.
- Adaptive trial designs such as sample size re-estimation, group sequential designs, adaptive enrichment designs, and multi-arm multi-stage designs, may be useful in telehealth trials.
- Adaptive designs could improve the efficiency, cost-effectiveness, accuracy and flexibility of telehealth trials.
- Characteristics of telehealth interventions, such as their impact on a range of endpoints, the lack of prior information about a new telehealth programme, and the heterogeneity in patient response, make them suitable for an adaptive approach to evaluation.

REFERENCES


