Author's response to reviews

Title: Effect of telehealth on glycaemic control: analysis of patients with type 2 diabetes in the Whole Systems Demonstrator cluster randomised trial

Authors:

Adam Steventon (adam.steventon@nuffieldtrust.org.uk)
Martin Bardsley (martin.bardsley@nuffieldtrust.org.uk)
Helen Doll (helen.doll@oxfordoutcomes.com)
Elizabeth Tuckey (lizzie.tuckey@gmail.com)
Stanton Newman (Stanton.Newman.1@city.ac.uk)

Version: 2 Date: 15 April 2014

Author's response to reviews: see over
Effect of telehealth on glycaemic control: analysis of patients with type 2 diabetes in the Whole Systems Demonstrator cluster randomised trial

Response to reviews received on 1 April 2014

We would like to thank all of the referees for their helpful comments, which have enabled us to improve the manuscript in some important ways. We have noted our gratitude towards the referees in the acknowledgement section of the manuscript. Below, we summarise our responses to each of the points raised.

Review by Geert Goderis

General Comments. This is a well-written manuscript, part of a larger trial with already some excellent publications. The article is important in its field. Most of my questions consider minor items and deal with clarifications.

We thank the reviewer for his comment.

I have one major compulsory remark that deals with the statistical analyses. I recommend the authors to analyze the difference in change in HbA1c before and after the start of the trial between intervention and control group. As stated now, the conclusion of the manuscript “telehealth led to improvement in glycaemic control” is technically spoken not supported by the analyses because the analyses only compared intervention and controls after the start of the trial.

Difference in change (before/after start of the trial between intervention and control) can be analysed using a longitudinal mixed linear regression model (eventually with three levels: level1= practice; level2= individual; level3= time) and thus 2 cluster levels, using splines with the starting point of the trial as defined spline.

Same remark for analyses of dichotomous outcomes.
The referee raises some important points. The principal analysis used a repeated measures model with a very similar specification to the one proposed (this should be clearer now that we have clarified the description of this model as suggested below). Indeed, as intervention effects were determined using the interaction of trial period and trial arm, we essentially did the proposed difference-in-difference analysis. Thus, we controlled for differences in HbA1c readings recorded before recruitment. Although we only assumed a linear trend for HbA1c (rather than a spline), we found that adding higher-order polynomial terms produced very similar results.

The secondary analysis (which summarised multiple readings for each individual by averaging across all those readings recorded within the trial period) did not use a difference-in-difference approach, though it did adjust for the last recorded HbA1c reading. We report values recorded before and after recruitment, as follows:

"Among control patients, mean HbA1c was 8.41% (68 mmol/mol) in the year before recruitment, and 8.38% (68 mmol/mol) during the 12-month trial period. Mean HbA1c thus showed little change among control patients. However, it fell among intervention patients from 8.38% (68 mmol/mol) to 8.15% (66 mmol/mol)."

On the basis of this, it seems a fair summary to say that the intervention group experienced an improvement in control, on average. However, we have now also conducted an additional difference-in-difference analysis, to provide reassurance. This analysis found a slightly smaller effect size than the main analysis (-0.20%), which did not reach statistical significance (95% CI, -0.58% to 0.17%, p=0.29). Our conclusion is that the effect sizes were consistent throughout the range of analyses done (in the range 0.2% to 0.3%). They were not always statistically significant in the secondary analysis, though the preferred, repeated measures model, did produce statistically significant results. We have included a note in the results and modified the abstract.

Note that we have moved this and all other sensitivity analysis to an appendix, as the sensitivity analysis is now quite extensive, and another referee was concerned with the length of the manuscript.
Minor remarks


We have reviewed and completed both statements (see the table at the end of this letter).

Methodology & results & discussion

Effect of patient recruitment on results (possible bias): Less than 5 diabetes patients per practice were recruited.

Some questions:

What was the average number of Type 2 diabetes patients per practice
(minimum / maximum)

The source of information on diabetes classification (type 1 or 2) was the general practice data, so was only available for patients who were linked to those data, and therefore were included in the current study. The median number of patients per practice was 3 in each trial arm (see Figure 1).

What was the minimum and maximum number of included patients per practice?

This is reported in the CONSORT diagram (Figure 1), and discussed in the paper. The range for control practices was 1-16 patients. For intervention practices it was 1-38 patients. Although there were some larger intervention practices in this study, in the principal study with all 3230 patients, there were larger practices in the control arm (see Steventon et al., 2012, Figure 1 [1]). We have included some discussion in the
'strengths and weaknesses' section, and would be happy to add further to this. We suspect that some of these differences arose by chance.

Are there significant differences in patient characteristics between participants and non-participants

The referee asks an important question. The evaluation team has already published a qualitative study exploring the reasons given by patients to refuse to participate in the trial. This is described in more detail in the revised manuscript. In addition, a comparison of participating and non-participating general practices has been published (see Cartwright et al, 2013, Web Appendix 4). We are undertaking further quantitative analysis, which includes a form assessment of the generalizability of the trial based on comparing the characteristics of participating and non-participating patients. This quantitative analysis is not yet available as we chose to prioritise the intention-to-treat analysis. However, the discussion section includes comparison between the included patients and other studies, and concludes that the included patients typically had less well-controlled diabetes. We would be happy to expand on the discussion if there are specific suggestions.

What were the reasons for the very high refusal rate of patients?

A qualitative study [2], referenced in the paper, explored the reasons given by patients for refusal to participate in the trial after the initial visit. This found three broad themes, relating to the requirements for technical competence and operation of equipment; threats to identity, independence and self-care; expectations and experiences of disruption to services. High refusal rates are common in telehealth trials [4] and, as we indicate in the paper, suggest that telehealth might not meet the needs of a substantial proportion of patients and a number may be reluctant to change from the current treatment they are receiving.

How did recruitment happen? Were all eligible patients contacted? Or did recruitment stop after obtaining sufficient numbers? If yes, did recruitment happen randomly or sequentially or... ?

Recruitment began by writing to all potentially eligible patients in a general practice. The need to produce policy-relevant findings meant that there was a deadline on
recruitment, of 30 September 2009, at which time slightly more than the targeted number of patients had been recruited. I believe that a large proportion of potentially eligible patients had been invited to participate by that date.

4b. Remarks & discussion: Strictly speaking, the findings are not generalizable to the whole diabetes patients but only to those patients who accepted to participate. There may be an important inclusion bias and the results may only applicable to those patients willing to adhere to that kind of intervention/care.

Please comment this in the discussion section. The discussion section only deals with the cluster effect on inclusion bias. This is too limited.

The reviewer is correct that the findings only apply to those patients who participated. The design of the WSD trial included strategies to promote generalisability—these included the use of multiple sites with varied characteristics, comparison with usual care, broad inclusion criteria, and a pragmatic approach to the intervention [5]. However, as is common for telehealth trials, a substantial proportion of eligible patients declined to participate. We had addressed these limitations in the discussion as follows:

"Approximately 80% of the individuals who were asked to participate in the WSD trial refused to do so [22]. Such high rates of patient refusal are not unusual for telehealth trials [31], and by themselves suggest that telehealth might not always meet the perceived needs of patients [20, 32]. The current study included a relatively high proportion of patients (49.7%) who had used insulin in the past year [2]. Further, compared to the previous IDEATel trial, baseline levels of HbA1c were higher in the current study: mean HbA1c was 8.4% (86 mmol/mol) compared with 7.4% (57 mmol/mol). It seems likely that the current study recruited a population with relatively less well-controlled diabetes. Further, although the current study was restricted to the subset of trial participants with type 2 diabetes, there was some comorbidity from heart failure and chronic obstructive pulmonary disease (which affected 13.0% and 16.3% of intervention patients, respectively)."
In response to your comment, we have included a more explicit statement that the recruited patients may not be fully representative of the population with type 2 diabetes. As stated above, we are conducting a formal, empirical assessment of the generalizability of this trial in relation to the primary outcomes of hospital utilization.

5. P. 7 the read codes: what are these codes, diagnostic code is not ICPC-2. Proper to EPR?

Read codes are the standard clinical terminology system used in General Practice in the United Kingdom. We have included a statement to this effect in the supplementary material.

6. p. 11 Repeated Measure model: was this an ANOVA? (I do not think so because measures are time-varying). Was it a mixed linear model with random effect? Random slope? Please specify (see also first major remark about longitudinal analyses).

Many thanks for raising this—we have added that we used a linear model. The random effects were already described.

7. p. 12 in order to summarize HbA1c, why did the authors use the mean value? I would have taken the last HbA1c value because HbA1c is already a parameter that indicates the glycaemic levels during the last three months prior to the measure (e.g., Goderis G, Van PG, Truyers C, Van C, De CE, Van Den Broeke C et al. Long-term evolution of renal function in patients with type 2 diabetes mellitus: a registry-based retrospective cohort study. BMJ Open 2013; 3(12):e004029). The problem with the 12-month average is that it may smooth the effect of the intervention. (You should not mention this reference, it's only to stimulate the debate about choices to make in the case of multiple measures). It may be helpful to conduct these supplementary analyses using the last available HbA1c value.

Many thanks—this is useful. We have done sensitivity analysis using the last HbA1c value (please see Additional File 2). This occurred around 8 months into the trial period on average, versus 6 months for the average value. Effect sizes were very similar: adjusted difference -0.29% when using the last value, versus -0.30% for the mean value.

Results p. 13 please be consequent in reporting e.g. first intervention then control, or vice versa, but do not mix both.
Many thanks. We have now done this.

9. p.14 (+ discussion): proportion of patients under or above certain cut-off values (7.5%, 6.72%, 9.85%) why 6.72% / 9.85%? were these values calculated in order to obtain a significant difference?

As stated in the paper, these cut-off values were chosen as they were used in another study,[6] which examined the relationship between HbA1c and survival in the UK. That study was particularly relevant to the current one, as it also used electronic data from general practice. We have improved the description of the approach in the methods section.

9b. again, difference in change should be analysed. If these differences were present before start of the trial, no real change happened. (method, longitudinal logistic regression with splines).

Many thanks. We have added some important information to Table 1, relating to the proportion of people under the 7.5% threshold at baseline (35.7% of controls and 33.3% of intervention patients). We have also conducted the proposed difference-in-difference analysis (see Additional File 2).

Discussion

10. What about the Hawthorne effect? The authors only deal with this problem with relation to the act of measuring HbA1c. But when patients are supposed to make self-control measures 5 times a week, the act of measuring happens 5 times a week and thus intervention patients are definitely aware that they are participating in a study... That's the Hawthorne effect. Please mention in the discussion section.

Many thanks for this. You are correct that the previous draft only mentioned the Hawthorne effect in relation to the endpoint measurement, when it is a broader issue. We have now included a more general statement in the discussion that: "In general, the recruitment and monitoring protocols used in randomised controlled trials can affect outcomes for both intervention and control groups, for example through Hawthorne effects."

11. What about sustainability of the findings?
Another referee also raised this point. We have acknowledged that this study cannot provide information about the sustainability of the improvements beyond the one-year follow-up. Indeed, it is possible that the improvements were concentrated towards the beginning of the follow-up period. Further research on this issue may be helpful.

12. Please explicitly mention “weaknesses of the study”.

We have restructured the discussion to include a section on strengths and weaknesses of the study.
Review by Susannah M cLean

The authors should be pleased that they have produced what appears to be a thorough and valid assessment of the state of the art of telehealthcare for diabetes.

We thank the referee for her comment.

Minor Essential Revisions:

1. It would be helpful to have some more examples of what precisely was done as a telehealth intervention. The authors have mentioned on several occasions that the interventions were at the discretion of the local service providers, however, if these interventions are to be rolled out we need to know what was done. A supplementary file with a table with details of the interventions at different locations would be very helpful as telehealth is far from homogenous.

The site-specific information has already been published in an open-access journal. For interested readers, we have linked to:

1) An existing document that describes the telehealth service models in each of the three sites in one particular year (see Henderson et al [7], Web Appendix).

2) A more detailed description of the telehealth interventions (Cartwright et al [3], Web Appendix 2).

We have also expanded on the description included in the current manuscript. This now includes: nature of base units; connectivity of base units and peripherals; the approaches to selecting thresholds; methods of communication between monitoring centre staff and patients; and confirmation that the participating patients were not asked to pay for the telehealth services. We believe this gives a good overview of what was done, while signposting interested readers to the more comprehensive resources.

2. In the discussion it might be useful to mention the diabetes control and complications trial (DCCT), which suggested a minimum reduction in HbA1c of -2% in order to improve
outcome. In this context, the reduction found in this study of -0.21% is unlikely to have a significant clinical impact in terms of improving outcomes.

Thank you. We have added a reference to the DCCT and quoted the effect size of 2% detected in that study. We agree that there are reasons to doubt that the effect sizes associated with telehealth interventions are large enough to create patient benefit, and have been stronger about this in the conclusions and abstract.

Research since that the DCCT has suggested that the relationship between achieved HbA1c is complex, and may depend on where in the distribution the reduction occurs as well as (as you suggest below) the sustainability of the reduction. We have described these complexities in the manuscript. Ultimately, these point to the need for studies to examine outcomes that matter to patients (as you say below), such as disease-specific quality of life and complications. A WSD analysis of disease-specific quality of life is planned, but a study of complications would need to have a longer follow-up than WSD.

Discretionary revisions:

1. In 2011 Liang et al. studied mobile phones for diabetes self-management in a meta-analysis and found a reduction in HbA1c values of a mean of -0.5%. There was some evidence that studies of longer duration demonstrated less tight control, indicating perhaps a novelty effect of the intervention that wore off with time. Perhaps the authors could comment on whether this might be a problem with their study.

The referee raises an important point. Due to the way that HbA1c readings were pooled, we assessed the average effect of telehealth on HbA1c over a 12-month period. We have now acknowledged this point in the strengths and weaknesses section. Although the evidence from the Liang review was weak (p = 0.49, not correcting for multiple comparisons made), we raise the possibility that patients might have been, for example, more adherent to the monitoring schedule in the early days of the trial. Unfortunately we have no data on this, but it is plausible.

We have amended the conclusions section of the manuscript, to make clear that this study (and indeed, most others in this area) unfortunately cannot provide information about the sustainability of improvement in HbA1c beyond the 12-month
follow-up. Indeed, it is possible that the improvements were concentrated towards the beginning of the follow-up period.

2. The authors may be interested to know of other systematic reviews of telehealthcare which include glycaemic control, many of which have found only small effects (< 0.5% HbA1c) which may not be clinically relevant and several of which have found non-significant impacts. They are all cited in our 2012 PLOS One systematic overview of telehealthcare interventions, as follows: Dellifane 2008, Haley 2003, Jackson 2005, Liang 2011, Montani 2001, Montori 2004, Pare 2007, Polisena 2009, Ramada 2001, Tian 2008, Verhoeven 2010 and Wu 2010. In this context we may say that there is a substantial body of evidence that telehealth is not very effective in improving HbA1c control. What is needed is more longer term trials of clinical outcomes such as amputations, blindness and cardiac events to quantify the effect on the outcomes that really matter to patients. This kind of research is increasingly feasible using routine data techniques similar to the kind used in the authors’ paper.

Many thanks for sharing these references. We had previously cited Polisena 2009 and have added Verhoeven 2010, Liang 2011, and your useful systematic overview. We do not cite Wu 2010 as it was concerned with telephone follow-up that did not necessarily involve transmission of physiological information. The other papers are older and can be found through your overview. As you say, the current study is in line with the reviews, which find only modest effects of telehealth on HbA1c.
Review by Neil Smart

Major Pulmonary Revisions

From reading the abstract, it is unclear whether or not this manuscript is a sub-study of the WSD study?

Was the WSD 3230 patients or the sub-study?

We have amended the title of the paper to reflect the nested nature of this study. We believe we have been sufficiently clear in the abstract, which states that (emphasis added):

...In the current study, we tested whether telehealth involving regular transmission of blood glucose information led to changes in glycosylated haemoglobin (HbA1c) among the subset of patients with type 2 diabetes.

Methods: The trial recruited 3,230 patients with diabetes, heart failure and chronic obstructive pulmonary disease; the current analysis was restricted to those patients with type 2 diabetes (n=513).

Please let us know if you think additional information is needed in the abstract.

How was usual care standardized?

Usual care was not standardized although there is common practice in the UK and rewards in place for monitoring at least annually. We have clarified that sites were asked to provide usual care to patients in the control arm, however this was defined.

Minor

In Abstract results why were 95% CIs omitted for some data but not others?

We have added the missing confidence interval.

Discretionary

The work is well written and presented but this is perhaps a reflection of the article length. As this is an open access journal I am presuming article length does not need to be curtailed?
As the manuscript is rather lengthy and may be shortened without compromising reporting quality.

We have reviewed the manuscript and removed unnecessary words in places.

Many thanks for your comments.
### CONSORT statement (cluster randomised trials)

<table>
<thead>
<tr>
<th>Title and abstract</th>
<th>CONSORT statement</th>
<th>Extension for cluster randomised trials</th>
<th>Extension for pragmatic trials</th>
<th>The current study</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a</td>
<td>Identification as a randomised trial in the title</td>
<td>Identification as a cluster randomised trial in the title</td>
<td></td>
<td>See title</td>
</tr>
<tr>
<td>1b</td>
<td>Structured summary of trial design, methods, results, and conclusions (for specific guidance see CONSORT for abstracts)</td>
<td>See table 2</td>
<td></td>
<td>See abstract</td>
</tr>
</tbody>
</table>

### Introduction

#### Background and objectives

| 2a | Scientific background and explanation of rationale | Rationale for using a cluster design | Describe the health or health service problem that the intervention is intended to address and other interventions that may commonly be aimed at this problem | The WS trial was established to answer the questions of policymakers, for example whether telehealth might be effective at improving the quality of life of patients with long-term health conditions, while preventing hospital admissions. Many other interventions have been proposed (see the first work cited in the introduction). |
A cluster design was used as individual randomisation was found to be unacceptable to stakeholders. See methods Design of the Whole Systems Demonstration trial.

| 2b | Specific objectives or hypotheses | Whether objectives pertain to the cluster level, the individual participant level or both | See Introduction. The trial and evaluation as a whole had several aims. We aimed to test whether telehealth was associated with changes in HbA1c for patients. |

**Methods**

<table>
<thead>
<tr>
<th>Trial design</th>
<th>3a</th>
<th>Description of trial design (such as parallel, factorial) including allocation ratio</th>
<th>Definition of cluster and description of how the design features apply to the clusters</th>
<th>General practices were the clusters. There was no pre-defined allocation ratio.</th>
</tr>
</thead>
<tbody>
<tr>
<td>3b</td>
<td>Important changes to methods after trial commencement (such as eligibility criteria), with reasons</td>
<td></td>
<td>Data collection for HbA1c switched from bespoke collection to the use of routine data sets.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Participants</th>
<th>4a</th>
<th>Eligibility criteria for participants</th>
<th>Eligibility criteria for clusters</th>
<th>Eligibility criteria should be explicitly framed to show the degree to which they include The three sites were selected after the competitive process described in the methods</th>
</tr>
</thead>
</table>
Within each site, eligibility criteria were broad. All practices in the sites were eligible to participate. Individual criteria were age 18 or over plus a diagnosis of COPD, HF or diabetes. This particular study is conducted on the subset of patients with type-2 diabetes, as identified in general practice electronic data.

### Settings and locations where the data were collected

Three areas in the south of England (Cornwall, Kent and Newham in East London).

### Interventions

**The interventions for each group with sufficient details to allow replication, including how and when they were actually administered**

**Whether interventions pertain to the cluster level, the individual participant level or both**

**Describe extra resources added to (or resources removed from) usual settings in order to implement intervention. Indicate if efforts were made to standardise the intervention or if the intervention and its delivery were allowed to vary between participants, practitioners, or study sites. Describe the comparator in similar detail to interventions broadly, and interested readers are referred to existing publications for more detailed...**

<table>
<thead>
<tr>
<th>Section</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>4b</td>
<td>Settings and locations where the data were collected</td>
</tr>
<tr>
<td>5</td>
<td>The interventions for each group with sufficient details to allow replication, including how and when they were actually administered</td>
</tr>
<tr>
<td></td>
<td>Whether interventions pertain to the cluster level, the individual participant level or both</td>
</tr>
<tr>
<td></td>
<td>Describe extra resources added to (or resources removed from) usual settings in order to implement intervention. Indicate if efforts were made to standardise the intervention or if the intervention and its delivery were allowed to vary between participants, practitioners, or study sites. Describe the comparator in similar detail to interventions broadly, and interested readers are referred to existing publications for more detailed...</td>
</tr>
<tr>
<td>Section</td>
<td>Question</td>
</tr>
<tr>
<td>---------</td>
<td>----------</td>
</tr>
<tr>
<td>6a</td>
<td>Outcomes</td>
</tr>
<tr>
<td>6b</td>
<td>Any changes to trial outcomes after the trial commenced, with reasons</td>
</tr>
<tr>
<td>7a</td>
<td>Sample size</td>
</tr>
<tr>
<td>7b</td>
<td>Whether equal or unequal cluster sizes are assumed, cluster size, a coefficient of intercluster correlation (ICC or k), and an indication of its uncertainty</td>
</tr>
<tr>
<td>7b</td>
<td>When applicable, explanation of any interim analyses and stopping guidelines</td>
</tr>
</tbody>
</table>

**Randomization:**

| 8a | Sequence generation | Method used to generate the random allocation sequence | M inimization algorithm. This is described in more detail in the primary publication [1]. |
| 8b | Type of randomization; details of any restriction (such as blocking and block size) | Details of stratification or matching if used | M inimization algorithm. |

**Allocation concealment mechanism:**

<p>| 9 | Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned | Specification that allocation was based on clusters rather than individuals and whether allocation concealment (if any) was at the cluster level, | The algorithm was centrally administered; allocations were concealed until after general practices had agreed to participate, but practices then... |</p>
<table>
<thead>
<tr>
<th>Implementation</th>
<th>10</th>
<th>Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions</th>
<th>Replace by 10a, 10b and 10c</th>
</tr>
</thead>
<tbody>
<tr>
<td>10a</td>
<td></td>
<td>Who generated the random allocation sequence, who enrolled clusters, and who assigned clusters to interventions</td>
<td>Allocation was done centrally by the evaluation team. Recruitment of clusters was done by local teams.</td>
</tr>
<tr>
<td>10b</td>
<td></td>
<td>Mechanism by which individual participants were included in clusters for the purposes of the trial (such as complete enumeration, random sampling)</td>
<td>Based on lists of registered patients.</td>
</tr>
<tr>
<td>10c</td>
<td></td>
<td>From whom consent was sought (representatives of the cluster, or individual cluster members, or both), and whether consent was sought before or after randomization</td>
<td>Both. Consent from practices was obtained prior to randomization. Consent from patients was obtained after randomization but prior to the patients being informed of their allocations.</td>
</tr>
<tr>
<td>Blinding</td>
<td>11a</td>
<td>If done, who was blinded after assignment to interventions (for example, participants, care providers, those assessing outcomes) and how</td>
<td>If blinding was not done, or was not possible, explain why.</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td>11b</td>
<td>If relevant, description of the similarity of interventions</td>
<td></td>
</tr>
<tr>
<td>Statistical methods</td>
<td>12a</td>
<td>Statistical methods used to compare groups for primary and secondary outcomes</td>
<td>How clustering was taken into account</td>
</tr>
</tbody>
</table>
### Methods for Additional Analyses

12b  Methods for additional analyses, such as subgroup analyses and adjusted analyses

---

### Adjusted Analyses

Methods for additional analyses, such as subgroup analyses and adjusted analyses are presented. We conducted exploratory subgroup analysis by insulin treatment but have not included these in the manuscript because they were hard to interpret.

---

### Results

#### Participant Flow

(a diagram is strongly recommended)

<table>
<thead>
<tr>
<th>13a</th>
<th>For each group, the number of participants who were randomly assigned, received intended treatment, and were analyzed for the primary outcome</th>
</tr>
</thead>
</table>

For each group, the number of participants or units approached to take part in the trial, the number that was eligible, and reasons for non-participation should be reported. See Figure 1. We do not present numbers who did not receive the intended treatments as these were so small. The reference cited in the note to figure 1 includes more detail.

<table>
<thead>
<tr>
<th>13b</th>
<th>For each group, losses and exclusions after randomization, together with reasons</th>
</tr>
</thead>
</table>

For each group, losses and exclusions for both clusters and individual cluster members.

See Figure 1. For clusters, losses occurred if the practice did not recruit patients for the trial, or if the patients they recruited did not meet the inclusion criteria for the current study. For patients, losses occurred if they refused to participate or if they did not meet the inclusion criteria for the current study but met the criteria for the

---
<table>
<thead>
<tr>
<th>Recruit ent</th>
<th>14a</th>
<th>Dates defining the periods of recruitment and follow-up</th>
<th>Recruit ent went from May 2008 to November 2009. Recruit ent was due to stop in September 2009 and we did not include patients recruited after that date in the analysis. Follow-up was over 12 months. D ata extraction was done retrospectively; the dates of data extraction are not reported as the schedule was complex and we believe that the general practice data was unlikely to change over the e. M one detail has already been published [8].</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline data</td>
<td>15</td>
<td>A table showing baseline demographic and clinical characteristics for each group</td>
<td>Baseline characteristics for the individual and cluster levels as applicable for each group</td>
</tr>
</tbody>
</table>
Figure 1. We were concerned about space in the text. The primary publication [1] contains some analysis of the characteristics of clusters and did not uncover differences.

<table>
<thead>
<tr>
<th>Numbers analysed</th>
<th>16 Numbers analysed</th>
<th>For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups</th>
<th>For each group, number of clusters included in each analysis</th>
<th>The primary analysis in the current study was done on 513 patients (213 control, 300 intervention) from 112 clusters (56 control, 56 intervention).</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outcomes and estimation</td>
<td>17a Outcomes and estimation</td>
<td>For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval)</td>
<td>Results at the individual or cluster level as applicable and a coefficient of intracluster correlation (ICC or k) for each primary outcome</td>
<td>We give results at the patient level as this is the relevant unit. We do not report ICCs as the study was powered based on hospitalization. However, the models include random effects at the general practice (cluster) level to take into account the observed clustering for HbA1c.</td>
</tr>
<tr>
<td></td>
<td>17b Outcomes and estimation</td>
<td>For binary outcomes, presentation of both absolute and relative effect sizes is recommended</td>
<td>The only binary outcome in the paper relates to the proportion of patients under the 7.5% (58 mmol/mol) threshold targeted by general practices. We report</td>
<td></td>
</tr>
<tr>
<td>Ancillary analyses</td>
<td>Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing pre-specified from exploratory</td>
<td>As explained above, we conducted exploratory subgroup analyses by insulin treatment but are not reporting these. We also conducted exploratory analysis of other outcomes (BMI, diastolic and systolic blood pressure, total cholesterol, and mortality). However, we are not reporting these in order to simplify the presentation of the manuscript. Effects were not significant.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Harms</td>
<td>All important harms or unintended effects in each group (for specific guidance see CONSORT for harms)</td>
<td>Harms were not assessed as part of this particular study.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discussion</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Limitations</td>
<td>Trial limitations, addressing sources of potential bias, in precision, and, if relevant, multiplicity of analyses</td>
<td>We have discussed some important limitations, including those associated with the use of routine data, threat of selection bias, and plurality of...</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| Generalisability | 21 | Generalisability (external validity, applicability) of the trial findings | Generalisability to clusters and/or individual participants (as relevant) | We have discussed the characteristics of the sample, which included more people with less well-controlled diabetes than the general population with diabetes. We also discuss the particular telehealth interventions. Telehealth may have different impacts in other settings, due to differences in context and in the decisions made about implementation. In particular, the chosen sites had a history of innovation in these areas of care and received funding and project manager support for the implementation of telehealth in this trial.

Describes key aspects of the setting that determined the trial results. Discuss possible differences in other settings where clinical traditions, health service organisation, staffing, or resources may vary from those of the trial.

Describe the key aspects of the setting that determined the trial results. Discuss possible differences in other settings where clinical traditions, health service organisation, staffing, or resources may vary from those of the trial.

We concluded that telehealth was associated with improvement in HbA1c in this trial. However, as we discuss, it is not clear whether these

| Interpretation | 22 | Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence | | |
Improvements would be sustained beyond a year, and meta-analyses have produced conflicting evidence about the long-term impact of intensive glycemic control on outcomes. Other strands of this evaluation will be considering patient-reported outcomes such as disease-specific quality of life, which are not yet available but will need to be taken into account by decision makers.

<table>
<thead>
<tr>
<th>Other information</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Registration</td>
<td>23</td>
</tr>
<tr>
<td>Protocol</td>
<td>24</td>
</tr>
<tr>
<td>Funding</td>
<td>25</td>
</tr>
</tbody>
</table>
submit this manuscript for publication.
References


