

# RESEARCH

## Effect of telehealth on quality of life and psychological outcomes over 12 months (Whole Systems Demonstrator telehealth questionnaire study): nested study of patient reported outcomes in a pragmatic, cluster randomised controlled trial

 OPEN ACCESS

Martin Cartwright *research associate in health services research*<sup>1</sup>, Shashivadan P Hirani *senior lecturer in health services research*<sup>1</sup>, Lorna Rixon *research associate in health services research*<sup>1</sup>, Michelle Beynon *research assistant in health services research*<sup>1</sup>, Helen Doll *senior research associate*<sup>2</sup>, Peter Bower *professor of health services research*<sup>6</sup>, Martin Bardsley *head of research*<sup>3</sup>, Adam Steventon *senior research analyst*<sup>3</sup>, Martin Knapp *professor of social policy*<sup>4</sup>, Catherine Henderson *research officer*<sup>4</sup>, Anne Rogers *professor of health systems implementation*<sup>5</sup>, Caroline Sanders *lecturer in medical sociology*<sup>6</sup>, Ray Fitzpatrick *professor of public health and primary care*<sup>7</sup>, James Barlow *professor of technology and innovation management (healthcare)*<sup>8</sup>, Stanton P Newman *principal investigator, professor, dean*<sup>1</sup>, for the Whole Systems Demonstrator evaluation team

<sup>1</sup>School of Health Sciences, City University London, London EC1A 7QN, UK; <sup>2</sup>University of East Anglia, Norwich, UK; <sup>3</sup>The Nuffield Trust, London, UK; <sup>4</sup>London School of Economics and Political Science, London, UK; <sup>5</sup>University of Southampton, Southampton, UK; <sup>6</sup>University of Manchester, Manchester, UK; <sup>7</sup>University of Oxford, Oxford, UK; <sup>8</sup>Imperial College Business School, London, UK

### Abstract

**Objective** To assess the effect of second generation, home based telehealth on health related quality of life, anxiety, and depressive symptoms over 12 months in patients with long term conditions.

**Design** A study of patient reported outcomes (the Whole Systems Demonstrator telehealth questionnaire study; baseline n=1573) was nested in a pragmatic, cluster randomised trial of telehealth (the Whole Systems Demonstrator telehealth trial, n=3230). General practice was the unit of randomisation, and telehealth was compared with usual care. Data were collected at baseline, four months (short term), and 12 months

(long term). Primary intention to treat analyses tested treatment effectiveness; multilevel models controlled for clustering by general practice and a range of covariates. Analyses were conducted for 759 participants who completed questionnaire measures at all three time points (complete case cohort) and 1201 who completed the baseline assessment plus at least one other assessment (available case cohort). Secondary per protocol analyses tested treatment efficacy and included 633 and 1108 participants in the complete case and available case cohorts, respectively.

Correspondence to: S P Newman [stanton.newman.1@city.ac.uk](mailto:stanton.newman.1@city.ac.uk)

Extra material supplied by the author (see <http://www.bmj.com/content/346/bmj.f653?tab=related#webextra>)

**Web appendix 1:** Whole Systems Demonstrator (WSD) terminology

**Web appendix 2:** Description of the telehealth intervention

**Web appendix 3:** Per protocol criteria

**Web figure 1:** Provision of peripheral telehealth devices to intervention participants in the WSD Telehealth Questionnaire Study

**Web figure 2:** Early removal of telehealth for reasons other than death across the 12 month trial period

**Web table 1:** Unadjusted mean change scores (Complete Case Cohort)

**Web table 2:** Unadjusted mean change scores (Available Case Cohort)

**Web table 3:** Means and estimated marginal means at each assessment point (intention-to-treat)

**Web table 4:** Means and estimated marginal means at each assessment point (per protocol)

**Web appendix 4:** Comparison of characteristics of participating and non-participating practices

**Setting** Provision of primary and secondary care via general practices, specialist nurses, and hospital clinics in three diverse regions of England (Cornwall, Kent, and Newham), with established integrated health and social care systems.

**Participants** Patients with chronic obstructive pulmonary disease (COPD), diabetes, or heart failure recruited between May 2008 and December 2009.

**Main outcome measures** Generic, health related quality of life (assessed by physical and mental health component scores of the SF-12, and the EQ-5D), anxiety (assessed by the six item Brief State-Trait Anxiety Inventory), and depressive symptoms (assessed by the 10 item Centre for Epidemiological Studies Depression Scale).

**Results** In the intention to treat analyses, differences between treatment groups were small and non-significant for all outcomes in the complete case ( $0.480 \leq P \leq 0.904$ ) or available case ( $0.181 \leq P \leq 0.905$ ) cohorts. The magnitude of differences between trial arms did not reach the trial defined, minimal clinically important difference (0.3 standardised mean difference) for any outcome in either cohort at four or 12 months. Per protocol analyses replicated the primary analyses; the main effect of trial arm (telehealth v usual care) was non-significant for any outcome (complete case cohort  $0.273 \leq P \leq 0.761$ ; available case cohort  $0.145 \leq P \leq 0.696$ ).

**Conclusions** Second generation, home based telehealth as implemented in the Whole Systems Demonstrator Evaluation was not effective or efficacious compared with usual care only. Telehealth did not improve quality of life or psychological outcomes for patients with chronic obstructive pulmonary disease, diabetes, or heart failure over 12 months. The findings suggest that concerns about potentially deleterious effect of telehealth are unfounded for most patients.

**Trial Registration** ISRCTN43002091.

## Introduction

Over the coming decades, extended life expectancy and low fertility will result in a shift in the old age dependency ratio in many countries including the United Kingdom, with a greater proportion of the population at retirement age than at working age.<sup>1-3</sup> Despite some positive changes in levels of old age disability<sup>4</sup> and years of self reported good health,<sup>5</sup> greater numbers of older people living with long term conditions are likely to present major challenges for health and social care systems in the years ahead.<sup>2 3 6 7</sup>

One response to these pressures from health systems has been the introduction of localised telehealth services. Telehealth enables the remote exchange of data between a patient and healthcare professionals to facilitate diagnosis, monitoring, and management of long term conditions.<sup>8 9</sup> Some telehealth systems incorporate an educational component aimed at improving patient knowledge<sup>10</sup> and self care (for example, treatment adherence).<sup>11 12</sup> Telehealth systems that send physiological or symptom data to a remote monitoring centre can alert healthcare professionals when disease specific clinical parameters are breached. Thus, telehealth affords the opportunity for earlier intervention, which may reduce the frequency with which expensive hospital based care is required.

Evaluations of service innovations such as telehealth need to assess the effect from the patient's perspective, using self report measures such as quality of life (QoL), psychological outcomes, and acceptability of services. This approach is in line with the developing agenda on patient reported outcomes,<sup>13-17</sup> and complements more familiar outcomes such as service use, costs, and mortality.

Generic health related QoL, anxiety, and depression are outcomes relevant to patients with the three long term conditions

that are the focus of the Whole Systems Demonstrator (WSD) Evaluation.<sup>18</sup> It is well established that health related QoL is reduced and anxiety and depression are elevated for patients with diabetes,<sup>19-21</sup> chronic obstructive pulmonary disease,<sup>22-24</sup> and heart failure.<sup>25-27</sup> Health related QoL, anxiety, and depression have been linked with poorer outcomes on endpoints including self management, disease control, health service use, costs, and mortality.<sup>28-32</sup>

However, evidence for the effect of telehealth on these outcomes is unclear. At least seven systematic reviews have examined this effect on health related QoL in heart failure,<sup>33-39</sup> and while most conclude that telehealth is beneficial, such inferences are not supported by the evidence they present. Typically, the reviews are poorly reported (for example, they report how many studies found a significant association but not how many studies looked for an association and failed to find one<sup>35</sup>), combine outcomes comprising measures that are conceptually distinct (for example, health related QoL combined with patient satisfaction and treatment adherence<sup>37</sup>), and fail to balance the evidence appropriately.<sup>39</sup>

In the two most transparent reviews, only one of three<sup>34</sup> and three of seven<sup>36</sup> studies that evaluated the monitoring of telehealth based vital signs reported any significant associations between telehealth and improvements in health related QoL. In a recent randomised controlled trial of third generation telehealth<sup>40</sup> that is not included in the cited reviews, researchers found no effect of telehealth on depression scores over 24 months, but found an overall benefit on one of eight SF-36 subscales.<sup>41</sup> Overall, claims that telehealth improves health related QoL for patients with heart failure are unsubstantiated.<sup>38</sup>

Two systematic reviews have examined the effect of telehealth on health related QoL for patients with chronic obstructive pulmonary disease. They showed ambivalent evidence; half the studies suggested a significant positive effect on health related QoL, and the other half showed no effect.<sup>42 43</sup>

One systematic review has investigated the effect of telehealth on health related QoL in diabetes.<sup>44</sup> This review confounds two patient reported outcomes with different meanings: health related QoL and patient satisfaction. Of only five studies that actually measured health related QoL, three found no difference between telephone support and usual care,<sup>45-47</sup> one failed to report differences between telehealth and usual care<sup>48</sup> and one pre-post study of telehealth found significant improvements on only three of eight SF-36 subscales (role-physical, bodily pain, and social functioning).<sup>49</sup>

Despite their relevance, few studies have examined anxiety or depression. This omission is important, given concerns about the potential detrimental effects of telehealth on patients. Such concerns include the greater burden of work on patients<sup>50</sup> and increased sense of isolation for vulnerable people by reducing face-to-face contact with healthcare professionals.<sup>8</sup>

Notwithstanding steady growth in telehealth studies over the past 20 years, robust evidence to inform policy decisions is lacking.<sup>51</sup> Systematic reviews show that although enthusiasts have written much about the promise of telehealth, most studies do not meet orthodox quality standards.<sup>42 52 53</sup> Furthermore, evidence from a few small trials of variable methodological quality is difficult to interpret.<sup>54</sup>

There is a danger of relying on even high quality systematic reviews if they pool data from low quality studies. This risk was underscored in a large scale, multicentre evaluation of automated telephone based monitoring for patients with heart failure.<sup>55</sup> By contrast with a recently updated Cochrane review,<sup>36</sup> this study found no evidence of benefit for interactive

telemonitoring on any outcome examined. These results highlight the need for rigorous, large scale, high quality independent studies to evaluate healthcare interventions before wide scale adoption.<sup>54 55</sup>

## WSD evaluation

Part of the UK government's response to this need for robust evidence was to fund the WSD Evaluation<sup>18</sup> to investigate the effects of two broad classes of technologies (telehealth and telecare) on a comprehensive range of outcomes in regions of England that had undergone the Whole Systems Redesign (web appendix 1). The design, protocol, and objectives of the WSD Evaluation have been reported in detail elsewhere.<sup>18</sup> Briefly, the evaluation comprises a pragmatic, cluster randomised controlled trial of telehealth for service users with long term conditions (chronic obstructive pulmonary disease, diabetes, heart failure; known as the WSD telehealth trial) and an equivalent trial of telecare for service users with social care needs (the WSD telecare trial).

These cluster randomised controlled trials evaluate a comprehensive range of healthcare utilisation outcomes and mortality. The evaluation was designed to avoid some of the shortcomings of previous research by conducting large, methodologically rigorous, multicentre trials across three regions of England. Each trial included a nested questionnaire study (the WSD telehealth questionnaire study and WSD telecare questionnaire study) to assess outcomes reported by patients and carers (for example, health related QoL, anxiety, depressive symptoms, functional ability, self care behaviour), and cost effectiveness based on quality adjusted life years.

Additional qualitative studies of purposive subsamples explored the experiences of patients, carers, healthcare professionals, and healthcare organisations. Collectively, the WSD Evaluation is the largest and most comprehensive investigation of telehealth and telecare so far. In the present study, we focus on the WSD telehealth questionnaire study and report on the effect of telehealth on health related QoL and two psychological outcomes (anxiety and depressive symptoms). For this part of the WSD Evaluation, we assessed the hypothesis that introduction of a broad class of home based telehealth improves quality of life, anxiety, and depressive symptoms over a 12 month period for patients with chronic obstructive pulmonary disease, diabetes, or heart failure, compared with usual care only.

## Methods

### Design and setting

The WSD telehealth trial is a pragmatic, cluster randomised controlled trial of telehealth (n=3230). This paper reports on the nested WSD telehealth questionnaire study, which was designed to include 1650 participants. Between May 2008 to December 2009, we recruited participants for the WSD telehealth trial, across three sociodemographically distinct regions in England (rural Cornwall, rural and urban Kent, and urban Newham in London) comprising four primary care trusts. Participants were also invited to take part in the WSD telehealth questionnaire study, a supplementary investigation of patient reported outcomes. We assessed participants at four and 12 months after recruitment (the last 12 month assessment occurred in December 2010). Figure 1<sup>1</sup> shows a CONSORT diagram of general practice and participant flow into the parent trial and the questionnaire study (n=1573). Table 1<sup>2</sup> compares sample characteristics at baseline across the parent trial and the nested questionnaire study.

## Cluster level recruitment and randomisation

Allocation was conducted at the cluster (general practice) level. All 365 general practices in the four primary care trusts were invited to participate. To maximise participation, the evaluation adopted a pragmatic approach: each practice provided intervention participants for one technology (telehealth or telecare) in one trial and control participants for the other technology (telecare or telehealth) in the other trial, ensuring equity of access to advanced assistive technology at the level of the practice population.<sup>18</sup> Consenting practices were allocated to the intervention and control groups by the trial statistician (HD), using a centrally administered minimisation algorithm that ensured comparability across trial arms in terms of practice size; deprivation; proportion of patients from non-white ethnic groups; prevalence of diabetes, chronic obstructive pulmonary disease, and heart failure; and WSD site.

There was no blinding for practices, participants, or assessors, although most measures in the WSD telehealth and telecare questionnaires studies were self reported. Participants allocated to the control arm were informed that they would be offered the appropriate technology at the end of the 12 month trial period, subject to a further needs assessment.

## Participant level recruitment

In participating practices, patients with chronic obstructive pulmonary disease, diabetes, or heart failure were deemed eligible on the basis of one of the following:

- Inclusion on the relevant Quality Outcomes Framework register in primary care
- A confirmed medical diagnosis in primary or secondary care medical records, as indicated by general practice Read codes or ICD-10 (international classification of diseases, 10th revision) codes
- Confirmation of disease status by a local clinician (such as general practitioner or community matron) or hospital consultant.

Patients were not excluded on the basis of additional physical comorbidities. However, participants were required to have a telephone landline for broadband internet connection (at all WSD sites), and a digital television (in Newham). Other financial costs (including phone calls and data transmission to the monitoring centres) were paid for by the local WSD project teams. Since the telehealth system used in the trial required participants to read and respond to textual information presented via a base unit or television screen, sufficient English language literacy was required, as determined by the local WSD project team. Cognitive impairment was not an exclusion criterion for the WSD telehealth trial, provided that an informal carer was available to assist with use of the telehealth system. However, cognitive impairment was an exclusion criterion for the questionnaire study because we aimed to collect self reported data without third party influence. Participants with physical impairments could receive practical assistance with completing the questionnaire battery from an independent trained researcher.

All 15 171 potentially eligible patients in participating practices were contacted about the study. To meet ethical obligations, these patients were initially sent and asked to complete a data sharing consent form if they were interested in the study and willing to allow their medical and social care data to be shared with the WSD research team. Follow-up letters and telephone calls encouraged responses.

Once data sharing consent was received eligibility was confirmed by the local WSD project team and eligible patients

were contacted to arrange a home visit to discuss the research in more detail. At this visit, the suitability of the participant's home infrastructure was checked, and a participant information sheet and consent form were provided. Participants provided written consent and indicated whether they would be willing to take part in the supplementary questionnaire study. Those willing were contacted by trained interviewers to arrange a baseline interview in the participant's home. At baseline interview, patients received a second information sheet relating specifically to the questionnaire study, and signed a second consent form for this part of the evaluation.

To minimise participant burden and create mutually exclusive subgroups for subsequent disease specific analyses (not reported here), participants with at least two of the three long term conditions were allocated to a single index condition using simple randomisation. Based on a prospective power calculation (see below), we aimed to recruit 1650 participants into the questionnaire study with an approximately even split between the three long term conditions (fig 2). Recruitment ended in December 2009.

### Telehealth treatment (intervention arm)

To facilitate comparisons between clinical studies, four classes (or "generations") of telehealth have recently been proposed on the basis of the type of data transfer, decision making ability of the care provider reviewing the data, and level of integration of all systems with the patient's primary care structure.<sup>40</sup>

First generation telehealth comprises non-reactive data collection and analysis systems. Measurements of interest are collected and transferred to the care provider asynchronously (that is, by store and forward protocols). There is no full telemedical system, and the provider cannot respond immediately to patient data. Second generation systems have a non-immediate analytical or decision making structure. Data transfer is synchronous—that is, there is some real time processing of patient data using, for example, automated algorithms to interpret the data. Care providers can recognise important changes in essential measurements, but delays can occur if the systems are only active during office hours.

Third generation systems provide constant analytical and decision making support. Monitoring centres are physician led, staffed by specialist nurses, and have full therapeutic authority 24 h per day, seven days per week. Fourth generation systems are an extension of third generation systems, comprising invasive (such as with surgical implantation) and non-invasive telemedical devices for data collection. The complexity of incoming information and subsequent therapeutic decisions requires the continuous presence of a physician.

WSD sites delivered variations of telehealth, but all systems focused on monitoring vital signs, symptoms, and self management behaviour. They provided general and disease specific health education, with non-immediate review by specialist nurses and other care providers. This configuration most closely approximates second generation telehealth.

Web appendix 2 and figure 3 describe the WSD telehealth intervention. Web figure 1 shows the provision of peripheral telehealth devices to intervention participants according to diagnosis of long term condition in each WSD site. Sites differed in the number of peripheral devices installed per participant, with a mode (across all long term conditions) of two in Cornwall and three in Kent and Newham. Web figure 2 shows the early removal of telehealth from participants for reasons other than death, by site. Differences in functionality of the telehealth equipment supplied, type and number of peripheral devices

provided, transfer of data to the monitoring centres, triage or risk stratification, and response pathways reflected variations that would occur if telehealth was implemented across the UK's entire health system.

### Usual care treatment (control arm)

Participants allocated to the control arm continued to receive their existing healthcare and social services, in line with local protocols, for the 12 months of the trial. Across the three WSD sites, healthcare was provided by a combination of community matrons, district nurses, specialist nurses, general practitioners, and hospital services based on clinical need. Patients had pre-established, tailored care plans that included routine assessments at a frequency appropriate for their disease severity—typically ranging from once per week to once or twice per year. Control participants had no telehealth or telecare equipment installed their homes for the duration of the study. A Lifeline pendant (a personal alarm) plus a smoke alarm linked to a monitoring centre were not, on their own, sufficient to classify as telecare for current purposes. We planned to reassess control participants at the end of the trial and, if still eligible, offer them telehealth.

All participants (intervention and control) were beneficiaries of the Whole Systems Redesign, which was a precondition of sites' participation in the trial. Putative benefits for patients included a better understanding of their condition and how to look after themselves through the development of self care behaviours and the continued support of services such as community matrons (web appendix 1).

### Trial assessment procedures

Outcomes were assessed at the level of the patient. At baseline, questions on outcome measures were answered by participants with a trained researcher on hand to explain or clarify the meaning of particular questions or assist with completing the questionnaire if participants were physically unable to do so. After the baseline interview, two further assessments were conducted. A short term assessment was conducted at about four months (median duration 127 days (interquartile range 37); 132 days (40) for control group, 126 days (35) for intervention group), and a long term assessment at around 12 months (347 days (49); 358 days (48), 342 days (47)). Duration at both assessments was similar across trial arms.

The questionnaire battery was the same at baseline and at short term; long term assessment included two additional scales measuring functional status<sup>56</sup> and impact of illness<sup>57</sup> (not reported here). At short term assessment, the survey battery was primarily administered as a postal survey with one reminder letter for non-responders; some participants also received telephone reminders. At long term assessment, the survey was posted to participants and non-responders were contacted to arrange a home interview with a trained researcher, in line with the baseline protocol. Participants who did not complete a questionnaire at short term were still invited to complete a questionnaire at long term. However, participants who withdrew from the trial, including intervention participants who asked for the telehealth equipment to be removed before the end of the 12 month trial period, were not sent further questionnaires after their withdrawal date.

### Patient reported outcomes

Findings in the current report are based on instruments assessing different domains of generic health related QoL (SF-12, EQ-5D), anxiety (Brief State-Trait Anxiety Inventory (STAI)), and

depressive symptoms (Center for Epidemiologic Studies Depression Scale (CESD-10)).

The SF-12<sup>58</sup> is a 12 item measure of general health status and health related QoL that uses norm based scoring for the general population in the United States in 1998. The instrument was scored in two subscales, the physical component summary score and the mental component summary score; higher scores represent better health related QoL. The SF-12 has shown good test-retest reliability, validity, and responsiveness, and is recommended for patients with heart failure.<sup>59</sup>

The EQ-5D<sup>60</sup> assesses five domains of generic health related QoL (mobility, self care, usual activities, pain and discomfort, anxiety and depression) and can generate either a health state (of 243 different states) or a single summary score (higher scores reflect better health related QoL). The EQ-5D has shown good validity and responsiveness and has been recommended for patients with diabetes<sup>61</sup> and, more cautiously, for patients with chronic obstructive pulmonary disease<sup>62</sup> and heart failure.<sup>59</sup> For current purposes, the summary score was used.

The Brief STAI<sup>63</sup> is a six item measure of state anxiety that has shown acceptable reliability and validity.<sup>63 64</sup> It is widely used in clinical research, notably in studies of patients with diabetes.<sup>65</sup> The state version, rather than the trait version, of the Brief STAI was used (higher scores reflect greater state anxiety).

The CESD-10<sup>66</sup> is a 10 item measure of depressive symptoms covering cognitive, emotional, and behavioural domains. It has acceptable validity and reliability,<sup>66</sup> and sensitivity and specificity.<sup>67</sup> The original 20 item version has been used widely with clinical populations, including chronic obstructive pulmonary disease<sup>68</sup> and heart failure,<sup>69</sup> although both versions of the scale include items that confound symptoms of physical illness with symptoms of depression (for example, "I felt that everything I did was an effort"; "My sleep was restless").<sup>70</sup> Scores range from 0 to 30, with higher scores indicating more depressive symptoms.

Minimal clinically important differences (MCIDs) have not been established for these patient reported outcomes. To evaluate the magnitude of any treatment effect, we regarded a trial defined MCID as an effect size equivalent to Cohen's  $d=0.3$ . This magnitude represents a "small" effect in the behavioural sciences.<sup>71</sup>

## Covariates in the analyses

Data were collected on a range of sociodemographic and trial related characteristics that could plausibly be related to the study outcomes. These variables were used as covariates in the main analyses. Date of birth and sex were extracted from general practice records. Ethnicity was assessed by self report, using 16 response categories based on standard UK categories from the Office of National Statistics<sup>72</sup>; missing responses were subsequently completed using data from medical records, where available. Education was assessed by self report using five response categories ranging from no formal education to graduate or professional level. We used participants' postcodes to allocate an index of multiple deprivation score.<sup>73</sup>

Comorbidity was assessed by a count of diagnosed conditions in hospital episode statistics over the three years before the trial began. The WSD project teams provided data for participants' WSD site; the presence or absence of a diagnosis of chronic obstructive pulmonary disease, diabetes, and heart failure; and the number and type of telehealth peripheral devices installed. The WSD evaluation team held data for participants' allocation (to telehealth or usual care) and calculated the duration of exposure to telehealth (in days) at the time each assessment

questionnaire was completed. Owing to the variability in telehealth duration for intervention participants at short term and long term assessments, this variable was included as a covariate.

## Sample size

For the telehealth questionnaire study, a power calculation was conducted on the basis of detecting a small effect size (Cohen's  $d=0.3$ ), allowing for an intracluster correlation coefficient of 0.05, power of 80%, and  $P<0.05$ . This calculation indicated that about 500 patients would be required to allow sufficient power to detect this small difference, ranging from 420 participants (five from each of 84 practices) to 520 participants (10 from each of 52 practices). These numbers were inflated by 10% to allow for the maximum possible increase in sample size due to variable cluster size.<sup>74</sup> The required sample size thus increased to 550. For sufficient power in our secondary subgroup analyses (not reported here), we aimed to recruit 550 patients per long term condition, or 1650 overall. All analyses reported here exceed the required sample size (550) and are therefore adequately powered.

## Statistical methods

Missing self reported data could occur at the questionnaire level or at the item or scale level. A participant who completes the questionnaire battery at baseline could fail to complete the questionnaire at short term or long term. Alternatively, a participant who largely completes a questionnaire could nevertheless fail to provide responses to certain items or may miss out whole scales within the battery.

For the outcomes reported, missing values at the questionnaire level were not imputed. We imputed missing values at the item or scale level using two methods. If a missing value belonged to a scale and at least 50% of responses were available for the scale (for a particular participant), we used the series mean for that scale (for that participant) to fill in missing values. If a missing value for an item did not belong to a scale (for example, index of multiple deprivation score) or if fewer than 50% of scale items were completed, missing values (either for items or scale totals) were multiply imputed ( $m=10$ ), on the basis of available data from several scales and items across all participants. We did multiple imputation using the Markov chain Monte Carlo function (SPSS).

We repeated analyses on each of the ten imputed datasets, and thereafter used standard multiple imputation procedures to combine the multiple scalar and multivariate estimates<sup>75-77</sup> with SPSS (version 19) and NORM.<sup>78</sup> We explored the influence of missing data at the questionnaire level by conducting complete case analyses (participants with data for all variables at all time points) and available case analyses (participants with data for all variables at baseline and at least one other time point). Depending on the reasons for missingness, both these approaches can generate biased results, but they are used here as sensitivity analyses to assess the robustness of the findings.

General practices were the unit of randomisation and were directly involved in the delivery of care to all participants, which could result in participants within practices being more similar than participants between practices. Causes of similarity within practices include pre-existing case mix differences between practice populations, and both general and specific practice effects (for example, factors that facilitate or inhibit access, general practitioner case load, the extent to which care is centred around the patient). To account for practice differences, multilevel modelling was used with observations (at different

time points) nested within participants, and participants nested within practices. The model included random intercepts and random slopes at the practice level.

Repeated measures for each outcome over the trial period were analysed with the linear mixed model procedures in SPSS. We used restricted maximum likelihood to estimate model parameters, with an ante-dependent (first order) variance-covariance matrix structure. A separate analysis was conducted for each of the five outcome variables, and the main effect of trial arm (telehealth v usual care) was estimated to answer the principal research question. We estimated the main effect of time to determine whether the outcome measures were different in the short and long term. The interaction between trial arm and time (trial arm×time) was also estimated to determine whether the trial arm had differential effects at short term and long term.

In each model, the baseline measure of a respective outcome variable was treated as a covariate, with the measures at short term and long term treated as the outcome. Covariates included in the model adjusted for baseline distributional differences between trial arms on sociodemographic and trial related variables that could be related to the outcomes.

Sociodemographic covariates included age, sex, ethnicity (white or non-white); education (ordinal, five levels); deprivation (continuous data); diagnosis of chronic obstructive pulmonary disease, diabetes, or heart failure; and total number of comorbidities (ordinal, nine levels). Trial related covariates included WSD site, number of peripheral telehealth devices installed (ordinal, five levels), and duration of exposure to telehealth (in days) at short term and long term assessment (continuous data). For all parameter tests, the  $\alpha$  level was set to 0.05.

We did intention to treat analyses to assess treatment effectiveness, as the most appropriate strategy for analysing pragmatic randomised controlled trials. However, this approach is conservative and risks underestimating treatment effects.<sup>79 80</sup> Complex healthcare interventions administered as part of a pragmatic trial risk being administered suboptimally, compared with being administered in routine care.

Obtaining an estimate of treatment efficacy would require a heavily resourced explanatory randomised controlled trial. However, an approximate evaluation of efficacy in pragmatic randomised controlled trials can be achieved via per protocol analyses. Thus, we conducted secondary per protocol analyses that analysed patients according to the treatment received rather than the treatment allocated (web appendix 3). Per protocol analyses risk overestimating the potential benefits that would be observed in routine practice. Considering primary (intention to treat) and secondary (per protocol) analyses together can help to disentangle treatments effects from implementation effects. Sensitivity analyses assessed the robustness of the findings to decisions taken at the analysis stage. Primary and secondary analyses were conducted for complete and available case cohorts. Here, data in the results section are taken from the primary analyses unless specified as being from secondary analyses.

## Results

### Descriptive statistics

At baseline, 1573 participants provided data in the questionnaire study (845 allocated telehealth, 728 usual care). The overall response rate was 62.7% (986/1573) at four months and 61.9% (974/1573) at 12 months, with a higher rate for telehealth at

both assessments. Overall, 48.3% (759/1573) of questionnaire participants were included in the complete case cohort, and 76.4% (1201/1573) in the available case cohort; again, with higher rates for telehealth (fig 1). Participants receiving telehealth in the WSD telehealth trial were thus more likely than those receiving usual care to opt in to the questionnaire study, more likely to provide data at both follow-up assessments, and consequently more likely to be included in complete and available case cohorts.

Table 1 presents sample characteristics for the 3230 participants in the parent trial, the 1573 participants in the nested questionnaire study at baseline, and those retained in the available case (n=1201) and complete case (n=759) cohorts. Compared to the parent trial, the complete case cohort in the questionnaire study had proportionally more participants from Kent, fewer non-white participants, fewer cases of chronic obstructive pulmonary disease, more cases of heart failure, a lower level of deprivation, and a higher level of education.

The trial arms were closely balanced in sample size in the parent trial cohort (telehealth=49.7% (1605/3230), usual care=50.3% (1625/3230)) but showed a marked discrepancy in the questionnaire study's complete case cohort (telehealth=56.8% (431/759), usual care=43.2% (328/759)). This discrepancy reflected the differences in response rates already described. Relative to the parent trial cohort, the questionnaire study's complete case cohort also had differences between trial arms in the proportion of participants in Cornwall and Newham, the proportion of participants with each long term condition, and the level of deprivation. The observed differences do not always show predictable patterns when comparing the parent trial cohort with the questionnaire study's baseline, available case, and complete case cohorts. Overall, table 1 shows that the composition of the questionnaire study samples were subject to potential bias—both in terms of patients who agreed to take part in the questionnaire study (baseline) and those who completed follow-up assessments—which underlines the need for case mix adjustment in the analyses.

### Preliminary unadjusted analyses

Figures 4 and 5 present unadjusted means and 95% confidence intervals for all outcomes by trial arm at all time points for the two analysis cohorts. The overlapping confidence intervals suggest that differences between arms at each assessment point were non-significant. Web tables 1 and 2 present similar analyses of unadjusted mean change scores. Some significant differences suggested that the telehealth arm had a slower rate of deterioration over time than the usual care arm in physical component score, anxiety, and depressive symptoms. However, the magnitude of all mean differences in change scores failed to reach the trial defined MCID (web tables 1 and 2).

### Primary analyses: treatment effectiveness

In intention to treat analyses, we used multilevel modelling to control for the baseline score of the respective outcome measures, key covariates, and the intracluster correlation (see methods). Key covariates included age; sex; ethnicity; education; deprivation; diagnoses of chronic obstructive pulmonary disease, diabetes, and heart failure; number of comorbidities; WSD site; number of peripheral telehealth devices installed; and duration of exposure to telehealth at each assessment. Parameter estimates, analogous to regression coefficients, and significance level are shown for the main effects of trial arm (telehealth v usual care) and time (short term v long term assessment) and

their interaction for each of the outcome measures (table 2). Tests of the effects of the covariates are not presented. Table 2 shows that trial arm, time, and their interaction were not significant for any outcome measure in either cohort.

To assist with the interpretation of table 2, web table 3 presents unadjusted means at baseline and estimated marginal means (EMMs) at short term and long term for all outcomes. EMMs were derived from a model that accounted for the intracluster correlation, all continuous or ordinal covariates, and baseline outcome measure; but not categorical covariates (known as “factors” within SPSS). For both cohorts, the pattern of means and EMMs closely mirror the parameter estimates in table 2 and reaffirm that differences between trial arms are clinically insubstantial. Minor differences of interpretation for some outcomes between table 2 and web table 3 are explained by the underlying differences in the statistical models used to generate the values reported.

Figures 6 and 7 show the adjusted effect sizes for trial arm (telehealth v usual care). Outcomes for both cohorts at short term and long term failed to reach the trial defined MCID. Further, all confidence intervals crossed zero, suggesting that estimates of the true treatment effect in the population could favour either telehealth or usual care. The true direction of the effect is uncertain and the magnitude of the effect is clinically insignificant.

## Secondary analyses: treatment efficacy

In per protocol analyses, multilevel modelling generated no significant main effects for trial arm or time for any outcome measure in either cohort (table 3). The two significant interaction terms for the SF-12 mental component score and CESD-10 in the complete case cohort reflect deteriorations for telehealth at short term, whereas scores remained stable over time for usual care (web table 4). The interaction findings were not replicated in the available case cohort (table 3). Differences between trial arms were unlikely to be clinically significant (figs 8 and 9).

No adverse events or side effects related to any of the telehealth devices were reported in the intervention group throughout the trial.

## Discussion

This large cluster randomised trial of second generation, home based telehealth for patients with chronic obstructive pulmonary disease, diabetes, or heart failure found no main effect of telehealth on generic health related QoL, anxiety, or depressive symptoms over 12 months. These null findings were consistent across a series of sensitivity analyses for the five validated outcome measures (tables 2 and 3). The null findings for the primary intention to treat analyses show that telehealth is not effective, while the null findings for the secondary per protocol analyses show that telehealth is not efficacious. Assessed against the trial defined MCID (equal to Cohen's  $d=0.3$ ), population estimates showed that the small, non-significant differences between trial arms in the primary analyses did not reach clinically significant levels for any outcome, in any cohort, at any time point (figs 6 and 7).

Exploratory investigations of trial arm×time interactions showed two significant effects for the mental component of health related QoL and depressive symptoms (table 3). At face value, these findings suggest that telehealth participants deteriorated at short term assessment before recovering to levels closer to baseline (and closer to usual care scores) at long term. However,

these findings were not robust across sensitivity analyses, and point estimates of effect size for these outcomes did not reach clinical significance at either assessment point (figs 8 and 9; table 3).

The overall consistency of results demonstrates that the findings are robust to variations in attrition (complete case v available case analyses), protocol fidelity (intention to treat v per protocol analyses), and choice of outcome measure. The similarity of the patient reported outcomes across trial arms suggests that concerns about the potentially deleterious effect of telehealth<sup>8 50</sup> are unfounded for most patients, since deterioration on any of the five outcome measures over the assessment period was not significant, compared with usual care. For the purposes of service planning, the current findings should be considered with other evidence from the WSD Evaluation on the effect of telehealth on hospital use and mortality<sup>81</sup> and cost effectiveness.<sup>82</sup>

## Comparison with other studies

When comparing our findings to existing research, it is important to distinguish between the statistical results reported in the extant literature and the conclusions drawn by authors. When considering only studies evaluating broadly equivalent forms of telehealth (that is, home based, vital signs monitoring using store and forward technology), systematic reviews have shown that fewer than half found any significant benefits to health related QoL,<sup>34 36 42 44</sup> and those that did only found effects on only a minority of the QoL measures used.<sup>49</sup>

Despite methodological variation across studies, these findings suggest that the effect of telehealth on health related QoL is weak or non-existent. To this extent, the available literature concurs with the current findings. However, some authors have observed that the conclusions drawn in many telehealth studies are often unduly positive.<sup>42 53</sup> With some notable exceptions,<sup>42</sup> the current study's conclusions therefore differ markedly from most extant studies and reviews examining the effect on telehealth on health related QoL, which are typically interpreted as showing benefits despite presenting equivocal evidence.

The scope for inappropriate inferences is increased when small and methodologically weak studies generate inconclusive results. The current findings underline the importance of using data from adequately powered, high quality trials to make decisions about telehealth implementation and caution against reliance on meta-analyses based on small, poor quality studies.<sup>40 55</sup> Our findings for second generation telehealth over 12 months mirror the recent null finding for third generation telehealth over 24 months.<sup>41</sup> Few studies have examined the effect of telehealth on anxiety or depressive symptoms, and the current findings extend our understanding of these outcomes.

## Strengths and limitations

The WSD telehealth trial is one of the largest randomised studies to evaluate the effect of telehealth on patient reported outcomes. A total of 1573 participants from 154 general practices across four primary care trusts (regional health authorities) provided questionnaire data at baseline. For the intention to treat analysis, 1201 participants from 150 practices were included in the available case cohort, and 759 from 131 practices in the complete case cohort (fig 1). By including participants with any of three long term conditions (chronic obstructive pulmonary disease, diabetes, or heart failure), imposing minimal exclusion criteria, and assessing participants over 12 months, the generalisability of the findings is maximised. The inclusion of three assessment points, multiple outcome measures, and robust statistical methods affords greater confidence in the findings.

Notwithstanding these strengths, some potential caveats should be acknowledged. All practices in the four primary care trusts were invited to participate in the WSD Evaluation, and 61% (224/365) agreed and identified at least one patient who met the eligibility criteria and participated in either the WSD telehealth trial (179 practices) or WSD telecare trial (217 practices). Data for pretrial practice characteristics show that participating practices differed in practice size, deprivation, ethnic composition, diabetes prevalence, and WSD site, but had no differences in the prevalence of chronic obstructive pulmonary disease or heart failure (web appendix 4). Despite these differences, recruited practices were not highly selected and heterogeneity was preserved (web appendix 4). For example, the percentages of non-participating practices categorised as having a low, medium, or high proportion of non-white patients were 32%, 43%, and 25% respectively, while corresponding percentages for participating practices were 25%, 34%, and 41%, respectively.

Similar concerns could be raised about the representativeness of participants in the questionnaire study. For practical and ethical reasons, we were unable to collect data on all patients who refused to participate in the WSD Evaluation at each stage of recruitment. Nevertheless, table 1 shows selection bias in those patients who agreed to participate in the nested questionnaire study relative to the parent trial, and attrition bias in those who were retained at follow-up. Participants allocated to telehealth (in the parent trial) were more likely than those allocated to usual care to agree to participate in the questionnaire study and to complete one or both assessments. Reasons for the relative advantage of the telehealth arm in recruitment and retention are unclear, though it is consistent with the principle of reciprocity; people receiving a notional benefit (such as telehealth) are more likely to comply with subsequent requests.<sup>83</sup> Potential threats to external validity from self selection or attrition bias underline the need to take care when generalising the results beyond the context of the trial. However, the relatively high level of practice participation and the large and heterogeneous participant sample support our assertion that any effect on the external validity of the trial is likely to be minor.

Participants allocated to telehealth did not receive equivalent treatments in all sites. Provision of peripheral telehealth devices (web figure 1) and response to biometric readings (web appendix 2) varied substantially by site and long term condition, as did the likelihood of having equipment removed prematurely for reasons other than death (web figure 2). In a pragmatic trial, this heterogeneity reflects the variability of implementation that would be observed in a wider rollout of telehealth, thereby increasing the generalisability of the findings.

In line with the original trial protocol, the analysis sought to draw conclusions about a general class of technology (telehealth) rather than about the effect of specific peripheral devices (pulse oximeter, glucometer, weight scales, blood pressure monitor) for specific long term conditions. Pooling patients with different profiles of long term conditions could mask differential treatment effects; therefore, planned analyses will examine the effect of telehealth on health related QoL, anxiety, and depressive symptoms for three subgroups of participants indexed to one long term condition (chronic obstructive pulmonary disease, diabetes, or heart failure).

We measured health related QoL using three generic scales (SF-12 physical component score, SF-12 mental component score, and EQ-5D) to assess different dimensions of the construct. It is recommended that assessment of health related QoL includes both generic and disease specific measures to capture the full range of effect of illness on health related

QoL,<sup>84 85</sup> and some evidence indicates that disease specific measures are more sensitive to clinical change.<sup>86</sup> Forthcoming analyses will examine the effect of telehealth on these disease specific measures. Although patient reported outcomes were the *a priori* primary endpoint, disease specific clinical markers would have afforded a more comprehensive description of the sample. Such markers include forced expiratory volume ratio for chronic obstructive pulmonary disease, HbA<sub>1c</sub> for diabetes, and the New York Heart Association classification for heart failure. Unfortunately, there were logistical barriers to the timely acquisition of clinical biomarkers. The planned analyses of disease specific measures of health related QoL (that is the chronic respiratory questionnaire, diabetes health profile, and Minnesota Living with Heart Failure questionnaire) will go some way to describing the clinical severity of the long term condition samples.

Despite providing an extensive description of the implemented telehealth treatment across sites and conditions (web appendix 2; fig 1, web fig 1), there were inevitably some aspects of the treatment where detailed data were unavailable. We do not have detailed information about changes to treatment or other clinical decisions initiated in response to telehealth data. Detailed information was not available about the degree to which telehealth participants adhered to their behavioural regimens (such as monitoring schedules, treatment adherence). We also do not have detailed information on the degree to which the telehealth technology encountered technical problems that interfered with measurement or the exchange of messages between participants and the monitoring centres. However, although about 20% of telehealth participants in the questionnaire study had their equipment removed prematurely during the trial (web fig 2), 80% retained telehealth services for the full 12 months; alternatively, around 10% of telehealth participants had their equipment removed in the first six months of the trial while 90% retained it for longer. In web appendix 2, protocols used by the monitoring centres show that any missing measurement sessions (whether from technical failure or participant non-adherence) were responded to within 72 h. Therefore, equipment failure or non-adherence are not plausible explanations for the observed null findings.

A further issue concerns the particular version of telehealth evaluated in the parent trial. Telehealth was implemented as monitoring of vital signs done daily (up to five days per week), supplemented by questions assessing health status and symptom severity, plus an educational component. The educational component consisted of brief textual information delivered through a static telehealth base unit with a small liquid crystal display (LCD) screen (Cornwall and Kent) or a dedicated interactive television channel (Newham). Telehealth participants in Newham could also watch short educational videos with disease specific information. Physiological and symptom report data were transferred to a monitoring centre using store and forward technology. In terms of a recently proposed classification of telehealth,<sup>40</sup> the telehealth system evaluated here most closely approximates a second generation system. The current findings therefore cannot be generalised to third or fourth generation systems that involve both invasive and non-invasive physiological monitoring with real time analytical and decision making support by physicians or physician led specialist nurses. Telehealth can only be studied as technology in use, and research evidence will always lag behind the latest technological advances.<sup>87</sup> However, most systems that have been tested so far represent first or second generations; third or fourth generations should be recognised a distinct class of

intensive interventions for select clinical populations at high risk.

The parent trial was set in the context of Whole Systems Redesign, and the three WSD sites were selected on the basis of having achieved substantial integration of health and social care. Assuming that this integration improves outcomes, the trial sought to identify any added benefit of telehealth services beyond those accrued from enhanced integration. In contexts with less integrated health and social care, telehealth benefits may be more likely to emerge. This argument assumes that integrated health and social care generates ceiling or floor effects for health related QoL, anxiety, and depressive symptoms. However, the baseline means show that our sample had similar health related QoL to other comparable clinical samples<sup>88</sup> with scope for either improvement or deterioration (figs 4 and 5). It is therefore unlikely that the lack of observed telehealth benefits can be attributed to the integrated care context or to recruitment of atypical clinical samples.

## Should we expect telehealth to improve health related QoL or psychological outcomes?

If telehealth delivers tailored healthcare that is acceptable to patients and facilitates more responsive interventions from professionals, resulting in better disease control with fewer exacerbations and admissions, we might expect corresponding improvements in health related QoL and psychological outcomes over time. Similarly, if telehealth leads to improved self care behaviour and efficacy, we might expect increases in health related QoL and reductions in negative affect. It remains unclear whether improvements in these patient reported outcomes are driven primarily by objective improvements in physical health, or by subjective improvements in perceptions of agency or control. Alternatively, telehealth could reduce health related QoL and psychological wellbeing owing to the increased burden of self monitoring, concerns about intrusive surveillance, a perceived lack of user friendliness, or the undermining of the traditional (face-to-face) therapeutic relationship. More research is required to understand the many potential beneficial and harmful mechanisms by which telehealth could affect patient reported outcomes. However, our findings strongly suggest no net benefit from telehealth; therefore, it should not be used as a tool to improve health related QoL or psychological outcomes.

## Future directions

The current findings point to other avenues of enquiry. Planned subgroup analyses for chronic obstructive pulmonary disease, diabetes, and heart failure using disease specific outcomes of health related QoL represent the next step in this process.

Alternative measures of health gain (such as self care, perceived impact of illness, and activities of daily living) are available for participants in the WSD telehealth questionnaire study and may offer a different perspective on the potential effect of telehealth. Effects of telehealth might not be uniform across all patients, and analyses may suggest subgroups of patients for whom telehealth is either particularly beneficial or harmful.

Building on existing qualitative work from the WSD Evaluation<sup>89</sup> and the data in web figure 2, we aim to identify predictors of early removal of telehealth. Quantitative analyses will also explore intervention participants' and carers' perceptions of telehealth over the course of the study, and ask whether these perceptions moderate outcomes. Other research questions in the WSD Evaluation are identified in the protocol paper.<sup>18</sup>

We thank Alan Glanz (Department of Health) and Chris Ham (The King's Fund) for their support throughout the study; all the individual participants for their time and interest in the study; and all the managers and professionals in Cornwall, Kent, and Newham in the health and social services and in the participating case study organisations for their help.

Whole System Demonstrator evaluation team members: Stanton P Newman (principal investigator), Martin Bardsley (The Nuffield Trust), James Barlow (Imperial College London), Jennifer Beecham (London School of Economics), Michelle Beynon (City University London/University College London), John Billings (The Nuffield Trust), Andy Bowen (University of Manchester), Pete Bower (University of Manchester), Martin Cartwright (City University London/University College London), Theopisti Chrysanthaki (Imperial College London), Jennifer Dixon (The Nuffield Trust), Helen Doll (University of East Anglia), Jose-Luis Fernandez (London School of Economics), Ray Fitzpatrick (Oxford University), Catherine Henderson (London School of Economics), Jane Hendy (Imperial College London), Shashivadan P Hirani (City University London/University College London), Martin Knapp (London School of Economics), Virginia MacNeill (Oxford University), Lorna Rixon (City University London/University College London), Anne Rogers (University of Southampton), Caroline Sanders (University of Manchester), Luis A Silva (City University London/University College London), Adam Steventon (The Nuffield Trust).

Contributors: MC and SPH are joint first authors. MC, SH, LR, and MBe conducted preliminary analyses under HD's supervision. MC conducted the final analyses and drafted the manuscript. HD, PB, MBa, MK, CH, AR, CS, RF, JB, SH, and SN contributed to development of the overall WSD study protocol. SN is the principal investigator for the WSD Evaluation. HD is the trial statistician and guarantor of statistical quality for the WSD Evaluation. MC, SH, LR, MBe, and SN contributed to the planning of the questionnaire data collection; MC, SH, LR, and MBe coordinated the daily implementation of the questionnaire assessment protocol and maintained trial data. HD, SH, MC, LR, MBe, AS, and CH contributed to planning of the analyses. All the authors reviewed the manuscript. The evaluation team met regularly during the trial period, reviewed interim documents and preliminary analyses, and contributed as a whole to discussions of the analytical strategy.

Funding: This is an independent report commissioned and funded by the Policy Research Programme in the UK Department of Health.

Competing interests: All authors have completed the ICMJE uniform disclosure form at [www.icmje.org/doi\\_disclosure.pdf](http://www.icmje.org/doi_disclosure.pdf) and declare: support from the Department of Health and the University College London Hospitals and University College London; several authors have undertaken evaluative work funded by government or public agencies but these have not created competing interests; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

Ethical approval: The study was approved by Liverpool research ethics committee (ref 08/H1005/4).

Data sharing: No additional data available.

- Christensen K, Doblhammer G, Rau R, Vaupel JW. Ageing populations: the challenges ahead. *Lancet* 2009;374:1196-208.
- Office for National Statistics. The interactive population pyramid. 2011. [www.neighbourhood.statistics.gov.uk/HTMLDocs/dvc1/UKPyramid.html](http://www.neighbourhood.statistics.gov.uk/HTMLDocs/dvc1/UKPyramid.html).
- Spillman BC, Lubitz J. The effect of longevity on spending for acute and long-term care. *N Engl J Med* 2000;342:1409-15.
- Robine J-M, Romieu I, Michel J. Trends in health expectancies. In: Robine J-M, Jagger C, Mathers C, Crimmins EM, Suzman R, eds. *Determining health expectancies*. John Wiley and Sons, 2003:75-104.
- Doblhammer G, Kytr J. Compression or expansion of morbidity? Trends in healthy-life expectancy in the elderly Austrian population between 1978 and 1998. *Soc Sci Med* 2001;52:385-91.
- Rice DP. Economic implications of increased longevity in the United States. *Annu Rev Public Health* 2004;25:457-73.
- Department of Health. Our health, our care, our say: A new direction for community services. 2006. [www.official-documents.gov.uk/document/cm67/6737/6737.pdf](http://www.official-documents.gov.uk/document/cm67/6737/6737.pdf).
- McLean S, Protti D, Sheikh A. Telehealthcare for long term conditions. *BMJ* 2011;342:d120.

**What is already known on this topic**

For long term conditions, telehealth has been promoted to reduce healthcare costs while improving health related quality of life (HRQoL), by facilitating self monitoring with remote surveillance by healthcare professionals

Evidence for the benefits of telehealth is ambivalent, with little empirical evidence on benefits on psychological outcomes

Methodologically rigorous trials of telehealth in relation to health related quality of life and psychological outcomes are required

**What this study adds**

Compared with usual care, second generation telehealth had no effect on HRQoL, anxiety, or depressive symptoms for patients with chronic obstructive pulmonary disease, diabetes, or heart failure over 12 months

The findings suggest that claims for potentially salutary or deleterious effects of telehealth are unfounded for most patients

Telehealth should not be introduced with the aim of improving quality of life or psychological outcomes

- 9 Sood S, Mbarika V, Jugoo S, Dookhy R, Doarn CR, Prakash N, et al. What is telemedicine? A collection of 104 peer-reviewed perspectives and theoretical underpinnings. *Telemed J E Health* 2007;13:573-90.
- 10 Ramaekers BLT, Janssen-Boyne JJ, Gorgels APM, Vrijhoef HJM. Adherence among telemonitored patients with heart failure to pharmacological and nonpharmacological recommendations. *Telemed J E Health* 2009;15:517-24.
- 11 Antonicelli R, Testarmata P, Spazzafumo L, Gagliardi C, Bilo G, Valentini M, et al. Impact of telemonitoring at home on the management of elderly patients with congestive heart failure. *J Telemed Telecare* 2008;14:300-5.
- 12 Seibert PS, Whitmore TA, Patterson C, Parker PD, Otto C, Basom J, et al. Telemedicine facilitates CHF home health care for those with systolic dysfunction. *Int J Telemed Appl* 2008. doi:10.1155/2008/235031.
- 13 US Food and Drug Administration. Draft guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims. 2009. www.fda.gov/downloads/Drugs/Guidances/UCM193282.pdf.
- 14 European Medicines Agency. Reflection paper on the regulatory guidance for the use of health-related quality of life (HRQL) measures in the evaluation of medicinal products. 2005. www.emea.europa.eu/docs/en\_GB/document\_library/Scientific\_guideline/2009/09/WC500003637.pdf.
- 15 Department of Health. Guidance on the routine collection of patient reported outcome measures (PROMs). 2009. www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH\_092647.
- 16 Dawson J, Doll H, Fitzpatrick R, Jenkinson C, Carr AJ. The routine use of patient reported outcome measures in healthcare settings. *BMJ* 2010;340:c186.
- 17 Greenhalgh J, Long AF, Flynn R. The use of patient reported outcome measures in routine clinical practice: lack of impact or lack of theory? *Soc Sci Med* 2005;60:833-43.
- 18 Bower P, Cartwright M, Hirani S, Barlow J, Hendy J, Knapp M, et al. A comprehensive evaluation of the impact of telemonitoring in patients with long-term conditions and social care needs: protocol for the Whole Systems Demonstrator cluster randomised trial. *BMC Health Serv Res* 2011;11:184.
- 19 Glasgow RE, Ruggiero L, Eakin EG, Dryfoos J, Chobanian L. Quality of life and associated characteristics in a large national sample of adults with diabetes. *Diabetes Care* 1997;20:562-7.
- 20 Lloyd CE, Dyer PH, Barnett AH. Prevalence of symptoms of depression and anxiety in a diabetes clinic population. *Diabetic Med* 2000;17:198-202.
- 21 Peyrot M, Rubin RR. Levels and risks of depression and anxiety symptomatology among diabetic adults. *Diabetes Care* 1997;20:585-90.
- 22 Ferrer M, Alonso J, Morera J, Marrades RM, Khalaf A, Aguar MC, et al. Chronic obstructive pulmonary disease stage and health-related quality of life. The Quality of Life of Chronic Obstructive Pulmonary Disease Study Group. *Ann Intern Med* 1997;127:1072-9.
- 23 Mikkelsen RL, Middelboe T, Pisinger C, Stage KB. Anxiety and depression in patients with chronic obstructive pulmonary disease (COPD). A review. *Nord J Psychiatry* 2004;58:65-70.
- 24 Putman-Casdorff H, McCrone S. Chronic obstructive pulmonary disease, anxiety, and depression: state of the science. *Heart Lung* 2009;38:34-47.
- 25 Rutledge T, Reis VA, Linke SE, Greenberg BH, Mills PJ. Depression in heart failure—a meta-analytic review of prevalence, intervention effects, and associations with clinical outcomes. *J Am Coll Cardiol* 2006;48:1527-37.
- 26 Konstam V, Moser DK, De Jong MJ. Depression and anxiety in heart failure. *J Card Fail* 2005;11:455-63.
- 27 Heo S, Moser DK, Lennie TA, Zarnbroski CH, Chung ML. A comparison of health-related quality of life between older adults with heart failure and healthy older adults. *Heart Lung* 2007;36:16-24.
- 28 Ciechanowski PS, Katon WJ, Russo JE. Depression and diabetes—impact of depression symptoms on adherence, function, costs. *Arch Intern Med* 2000;160:3278-85.
- 29 Moussavi S, Chatterji S, Verdes E, Tandon A, Patel V, Ustun B. Depression, chronic diseases, and decrements in health: results from the World Health Surveys. *Lancet* 2007;370:851-8.
- 30 Yohannes AM, Willgoss TG, Baldwin RC, Connolly MJ. Depression and anxiety in chronic heart failure and chronic obstructive pulmonary disease: prevalence, relevance, clinical implications and management principles. *Int J Geriatr Psychiatry* 2010;25:1209-21.
- 31 Maurer J, Rebbapragada V, Borson S, Goldstein R, Kunik ME, Yohannes AM, et al. Anxiety and depression in COPD current understanding, unanswered questions, and research needs. *Chest* 2008;134:43-56S.
- 32 Rodríguez-Artalejo F, Guallar-Castillon P, Pascual CR, Otero CM, Montes AO, García AN, et al. Health-related quality of life as a predictor of hospital readmission and death among patients with heart failure. *Arch Intern Med* 2005;165:1274-9.
- 33 Martínez A, Everss E, Luis Rojo-Alvarez J, Pascual Figal D, García-Alberola A. A systematic review of the literature on home monitoring for patients with heart failure. *J Telemed Telecare* 2006;12:234-41.
- 34 Clark RA, Inglis SC, McAlister FA, Cleland JGF, Stewart S. Telemonitoring or structured telephone support programmes for patients with chronic heart failure: systematic review and meta-analysis. *BMJ* 2007;334:942-5.
- 35 Maric B, Kaan A, Ignaszewski A, Lear SA. A systematic review of telemonitoring technologies in heart failure. *Eur J Heart Fail* 2009;11:506-17.
- 36 Inglis SC, Clark RA, McAlister FA, Ball J, Lewinter C, Cullington D, et al. Structured telephone support or telemonitoring programmes for patients with chronic heart failure. *Cochrane Database Syst Rev* 2010:CD007228.
- 37 Polisen J, Tran K, Cimon K, Hutton B, McGill S, Palmer K, et al. Home telemonitoring for congestive heart failure: a systematic review and meta-analysis. *J Telemed Telecare* 2010;16:68-76.
- 38 Schmidt S, Schuchert A, Krieg T, Oeff M. Home telemonitoring in patients with chronic heart failure: a chance to improve patient care? *Dtsch Arztebl Int* 2010;107:131-8.
- 39 Clarke M, Shah A, Sharma U. Systematic review of studies on telemonitoring of patients with congestive heart failure: a meta-analysis. *J Telemed Telecare* 2011;17:7-14.
- 40 Anker SD, Koehler F, Abraham WT. Telemedicine and remote management of patients with heart failure. *Lancet* 2011;378:731-9.
- 41 Koehler F, Winkler S, Schieber M, Sechtem U, Stangl K, Böhm M, et al. Impact of remote telemedical management on mortality and hospitalizations in ambulatory patients with chronic heart failure: the telemedical interventional monitoring in heart failure study. *Circulation* 2011;123:1873-80.
- 42 Bolton CE, Waters CS, Peirce S, Elwyn G. Insufficient evidence of benefit: a systematic review of home telemonitoring for COPD. *J Eval Clin Pract* 2011;17:1216-22.
- 43 Polisen J, Tran K, Cimon K, Hutton B, McGill S, Palmer K, et al. Home telehealth for chronic obstructive pulmonary disease: a systematic review and meta-analysis. *J Telemed Telecare* 2010;16:120-7.
- 44 Polisen J, Tran K, Cimon K, Hutton B, McGill S, Palmer K. Home telehealth for diabetes management: a systematic review and meta-analysis. *Diabetes Obes Metab* 2009;11:913-30.
- 45 Maljanian R, Grey N, Staff I, Conroy L. Intensive telephone follow-up to a hospital-based disease management model for patients with diabetes mellitus. *Dis Manag* 2005;8:15-25.
- 46 Piette JD, Weinberger M, McPhee SJ. The effect of automated calls with telephone nurse follow-up on patient-centered outcomes of diabetes care: a randomized, controlled trial. *Med Care* 2000;38:218-30.
- 47 Whitlock WL, Brown A, Moore K, Pavlisca H, Dingbaum A, Laceyfield D, et al. Telemedicine improved diabetic management. *Mil Med* 2000;165:579-84.
- 48 Jansa M, Vidal M, Viaplana J, Levy I, Conget I, Gomis R, et al. Telecare in a structured therapeutic education programme addressed to patients with type 1 diabetes and poor metabolic control. *Diabetes Res Clin Pract* 2006;74:26-32.
- 49 Chumbar NR, Neugaard B, Kobb R, Ryan P, Qin HJ, Joo YS. Evaluation of a care coordination/home-telehealth program for veterans with diabetes: health services utilization and health-related quality of life. *Eval Health Prof* 2005;28:464-78.
- 50 Rogers A, Kirk S, Gately C, May CR, Finch T. Established users and the making of telecare work in long term condition management: implications for health policy. *Soc Sci Med* 2011;72:1077-84.
- 51 Ekeland AG, Bowes A, Flottorp S. Effectiveness of telemedicine: a systematic review of reviews. *Int J Med Inform* 2010;79:736-71.
- 52 Barlow J, Singh D, Bayer S, Curry R. A systematic review of the benefits of home telecare for frail elderly people and those with long-term conditions. *J Telemed Telecare* 2007;13:172-9.
- 53 Bensink M, Hailey D, Wootton R. A systematic review of successes and failures in home telehealth. Part 2: final quality rating results. *J Telemed Telecare* 2007;13:10-4.
- 54 Chaudhry SI, Matterna JA, Krumholz HM. Telemonitoring in patients with heart failure: authors' reply. *N Engl J Med* 2011;364:1078-80.
- 55 Chaudhry SI, Matterna JA, Curtis JP, Spertus JA, Herrin J, Lin ZQ, et al. Telemonitoring in patients with heart failure. *N Engl J Med* 2010;363:2301-9.
- 56 Bowling A, Farquhar M, Browne P. Life satisfaction and associations with social network and support variables in 3 samples of elderly people. *Int J Geriatr Psychiatry* 1991;6:549-66.
- 57 Klimidis S, Minas IH, Yamamoto K. Impact of illness scale: reliability, validity, and cross-cultural utility. *Compr Psychiatry* 2001;42:416-23.
- 58 Ware JE Jr, Kosinski M, Turner-Bowker DM, Gandek B. User's manual for the SF-12v2 health survey with a supplement documenting SF-12 health survey. QualityMetric, 2002.
- 59 Mackintosh A, Gibbons E, Fitzpatrick R. A structured review of patient-reported outcome measures (PROMs) for heart failure. University of Oxford, 2009. [http://phi.uhce.ox.ac.uk/pdf/PROMs\\_Oxford\\_HeartFailure\\_17092010.pdf](http://phi.uhce.ox.ac.uk/pdf/PROMs_Oxford_HeartFailure_17092010.pdf).
- 60 Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol Group. *Ann Med* 2001;33:337-43.
- 61 Gibbons E, Fitzpatrick R. A structured review of patient-reported outcome measures (PROMs) for diabetes. University of Oxford, 2009. [http://phi.uhce.ox.ac.uk/pdf/Diabetes\\_2009FINAL.pdf](http://phi.uhce.ox.ac.uk/pdf/Diabetes_2009FINAL.pdf).
- 62 Davies N, Gibbons E, Fitzpatrick R. A structured review of patient-reported outcome measures (PROMs) for chronic obstructive pulmonary disease (COPD). University of Oxford, 2009. [http://phi.uhce.ox.ac.uk/pdf/COPD2009\\_FINAL.pdf](http://phi.uhce.ox.ac.uk/pdf/COPD2009_FINAL.pdf).
- 63 Marteau TM, Bekker H. The development of a 6-item short-form of the state scale of the Spielberger state trait anxiety inventory (STAI). *Br J Clin Psychol* 1992;31:301-6.
- 64 Perpina-Galvan J, Richart-Martinez M, Jose Cabanero-Martinez M. Reliability and validity of a short version of the STAI anxiety measurement scale in respiratory patients. *Arch Bronconeumol* 2011;47:184-9.

- 65 Eborall HC, Griffin SJ, Prevost AT, Kinmonth A-L, French DP, Sutton S. Psychological impact of screening for type 2 diabetes: controlled trial and comparative study embedded in the ADDITION (Cambridge) randomised controlled trial. *BMJ* 2007;335:486-9.
- 66 Andresen EM, Malmgren JA, Carter WB, Patrick DL. Screening for depression in well older adults: evaluation of a short-form of the CES-D. *Am J Prev Med* 1994;10:77-84.
- 67 Boey KW. Cross-validation of a short form of the CES-D in Chinese elderly. *Int J Geriatr Psychiatry* 1999;14:608-17.
- 68 van Manen JG, Bindels PJE, Dekker FW, Ijzermans CJ, van der Zee JS, Schade E. Risk of depression in patients with chronic obstructive pulmonary disease and its determinants. *Thorax* 2002;57:412-6.
- 69 Lesman-Leegte I, Jaarsma T, Sanderma R, Linssen G, van Veldhuisen DJ. Depressive symptoms are prominent among elderly hospitalised heart failure patients. *Eur J Heart Fail* 2006;8:634-40.
- 70 Mancuso CA, Westermann H, Choi TN, Wenderoth S, Briggs WM, Charlson ME. Psychological and somatic symptoms in screening for depression in asthma patients. *J Asthma* 2008;45:221-5.
- 71 Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. Lawrence Erlbaum Associates, 1988.
- 72 Office of National Statistics. Ethnic group statistics: a guide for the collection and classification of ethnicity data. Her Majesty's Stationery Office, 2003:39.
- 73 Department of Health. Indices of deprivation. 2007. <http://webarchive.nationalarchives.gov.uk/+http://www.communities.gov.uk/communities/neighbourhoodrenewal/deprivation/deprivation07/>.
- 74 Eldridge SM, Ashby D, Kerry S. Sample size for cluster randomized trials: effect of coefficient of variation of cluster size and analysis method. *Int J Epidemiol* 2006;35:1292-300.
- 75 Allison PD. Missing data. Sage Publications, 2001.
- 76 Little RJA, Rubin DB. Statistical analysis with missing data. 2nd ed. Wiley-Interscience, 2002.
- 77 Schafer JL, Olsen MK. Multiple imputation for multivariate missing-data problems: a data analyst's perspective. *Multivariate Behav Res* 1998;33:545-71.
- 78 Schafer JL. NORM: multiple imputation of incomplete multivariate data under a normal model (version 2.03). Pennsylvania State University, 1999.
- 79 Higgins JPT, Deeks JJ, Altman DG. Special topics in statistics: intention-to-treat issues In: Higgins JPT, Green S, eds. Cochrane handbook for systematic reviews of interventions (version 5.1.0). Cochrane Collaboration, 2011.
- 80 Heritier SR, Gebiski VJ, Keech AC. Inclusion of patients in clinical trial analysis: the intention-to-treat principle. *Med J Aus* 2003;179:438-40.
- 81 Steventon A, Bardsley M, Billings J, Dixon J, Doll H, Hirani S, et al. Effect of telehealth on use of secondary care and mortality: findings from the Whole System Demonstrator cluster randomised trial. *BMJ* 2012;344:e3874.
- 82 Henderson C, Knapp M, Fernández J, Beecham J, Hirani S, Cartwright M, et al. Cost-effectiveness of telehealth in patients with long-term conditions: the Whole Systems Demonstrator cluster randomised trial. *BMJ* [forthcoming].
- 83 Regan DT. Effects of a favor and liking on compliance. *J Exp Soc Psychol* 1971;7:627-39.
- 84 Engstrom CP, Persson LO, Larsson S, Sullivan M. Health-related quality of life in COPD: why both disease-specific and generic measures should be used. *Eur Respir J* 2001;18:69-76.
- 85 Katsura H, Yamada K, Kida K. Both generic and disease specific health-related quality of life are deteriorated in patients with underweight COPD. *Respir Med* 2005;99:624-30.
- 86 Pickard AS, Yang Y, Lee TA. Comparison of health-related quality of life measures in chronic obstructive pulmonary disease. *Health Qual Life Outcomes* 2011;9:26.
- 87 Bartoli L, Zanaboni P, Masella C, Ursini N. Systematic review of telemedicine services for patients affected by chronic obstructive pulmonary disease (COPD). *Telemed J E Health* 2009;15:877-83.
- 88 Jenkinson C, Layte R, Jenkinson D, Lawrence K, Petersen S, Paice C, et al. A shorter form health survey: Can the SF-12 replicate results from the SF-36 in longitudinal studies? *J Public Health Med* 1997;19:179-86.
- 89 Sanders C, Rogers A, Bowen R, Bower P, Hirani S, Cartwright M, et al. Exploring barriers to participation and adoption of telehealth and telecare within the Whole System Demonstrator trial: a qualitative study. *BMC Health Serv Res* 2012;12.

**Accepted:** 16 January 2013

**Cite this as:** *BMJ* 2013;346:f653

This is an open-access article distributed under the terms of the Creative Commons Attribution Non-commercial License, which permits use, distribution, and reproduction in any medium, provided the original work is properly cited, the use is non commercial and is otherwise in compliance with the license. See: <http://creativecommons.org/licenses/by-nc/2.0/> and <http://creativecommons.org/licenses/by-nc/2.0/legalcode>.

## Tables

**Table 1 | Sample characteristics at baseline. Data are no (%) unless stated otherwise**

	WSD telehealth trial (parent trial)			WSD telehealth questionnaire study (nested study)								
				Baseline cohort			Available case cohort			Complete case cohort		
	TH (n=1605; 49.7%)	UC (n=1625; 50.3%)	Total (n=3230)	TH (n=845; 53.7%)	UC (n=728; 46.3%)	Total (n=1573)	TH (n=670; 55.8%)	UC (n=531; 44.2%)	Total (n=1201)	TH (n=431; 56.8%)	UC (n=328; 43.2%)	Total (n=759)
<b>WSD site</b>												
Cornwall	566 (35.3)	625 (38.5)	1191 (36.9)	256 (30.3)	234 (32.1)	490 (31.2)	221 (33.0)	168 (31.6)	389 (32.4)	149 (34.6)	107 (32.6)	256 (33.7)
Kent	583 (36.3)	595 (36.6)	1178 (36.5)	343 (40.6)	283 (38.9)	626 (39.8)	285 (42.5)	216 (40.7)	501 (41.7)	202 (46.9)	153 (46.6)	355 (46.8)
Newham	456 (28.4)	405 (24.9)	861 (26.7)	246 (29.1)	211 (29.0)	457 (29.1)	164 (24.5)	147 (27.7)	311 (25.9)	80 (18.6)	68 (20.7)	148 (19.5)
<b>Sex</b>												
Female	664 (41.4)	658 (40.5)	1322 (40.9)	350 (41.4)	290 (39.8)	640 (40.7)	271 (40.4)	207 (39.0)	478 (39.8)	171 (39.7)	124 (37.8)	295 (38.9)
Male	941 (58.6)	967 (59.5)	1908 (59.1)	495 (58.6)	438 (60.2)	933 (59.3)	399 (59.6)	324 (61.0)	723 (60.2)	260 (60.3)	204 (62.2)	464 (61.1)
<b>Ethnicity*</b>												
Non-white	187 (11.7)†	180 (11.1)†	367 (11.4)†	111.5 (13.2)*	99.9 (13.7)*	211.4 (13.4)*	75 (11.2)	68 (12.8)	143 (11.9)	30 (7.0)	23 (7.0)	53 (7.0)
White	1251 (77.9)†	1264 (77.8)†	2515 (77.9)†	733.5 (86.8)*	628.1 (86.3)*	1361.6 (86.6)*	595 (88.8)	463 (87.2)	1058 (88.1)	401 (93.0)	305 (93.0)	706 (93.0)
Data missing	167 (10.4)†	181 (11.1)†	348 (10.8)†	—	—	—	—	—	—	—	—	—
<b>Long term condition diagnoses‡</b>												
COPD	902 (56.2)	962 (59.2)	1864 (57.7)	416 (49.2)	327 (44.9)	743 (47.2)	339 (50.6)	237 (44.6)	576 (48.0)	238 (55.2)	158 (48.2)	396 (52.2)
Diabetes	570 (35.5)	485 (29.8)	1055 (32.7)	339 (40.1)	283 (38.9)	622 (39.5)	243 (36.3)	205 (38.6)	448 (37.3)	128 (29.7)	108 (32.9)	236 (31.1)
Heart failure	566 (35.3)	581 (35.8)	1147 (35.5)	336 (39.8)	327 (44.9)	663 (42.1)	284 (42.4)	248 (46.7)	532 (44.3)	178 (41.3)	162 (49.4)	340 (44.8)
<b>Other (mean (SD))</b>												
Age (years)	69.68 (11.55)	70.80 (11.64)	70.24 (11.61)	70.11 (11.81)	70.61 (11.78)	70.34 (11.79)	70.14 (10.98)	70.46 (11.61)	70.28 (11.26)	70.61 (9.72)	71.21 (11.00)	70.87 (10.29)
Deprivation (IMD)*	28.34 (14.82)†	27.93 (13.53)†	28.13 (14.18)†	27.71 (15.01)*	28.55 (13.79)*	28.10 (14.46)*	26.15 (14.24)*	28.10 (13.77)*	27.01 (14.07)*	24.44 (13.62)*	25.76 (12.96)*	25.01 (13.35)*
No of comorbidities	1.75 (1.79)	1.80 (1.82)	1.77 (1.81)	1.84 (1.80)	2.02 (1.86)	1.92 (1.83)	1.85 (1.78)	2.00 (1.88)	1.92 (1.83)	1.72 (1.76)	1.93 (1.84)	1.81 (1.80)
No of TH devices	2.68 (0.70)	0.04 (0.36)	1.35 (1.43)	2.77 (0.68)	0.07 (0.49)	1.52 (1.47)	2.74 (0.69)	0.07 (0.47)	1.56 (1.46)	2.69 (0.65)	0.04 (0.32)	1.54 (1.42)

TH=telehealth; UC=usual care; COPD=chronic obstructive pulmonary disease; IMD=index of multiple deprivation score; available case cohort=includes baseline assessments and short term or long term assessment; complete case cohort=includes assessments at all three time points (baseline, four months (short term), 12 months (long term)).

\*Participants ( $\leq 12$ ) had missing data for ethnicity or deprivation in some cohorts of the questionnaire study. Missing values were multiply imputed ( $m=10$ ), and the values reported are imputed averages.

†The parent trial cohort had more participants than the questionnaire study cohort with missing data for ethnicity ( $n=348$ ) and deprivation ( $n=12$ ). These missing values were not imputed; instead, missing values are reported explicitly for ethnicity, while for deprivation we reported the means (and standard deviations) based on 3218 (99.6%) participants with data available. All other values are based on complete (non-imputed) data.

‡Participants could have had more than one diagnosis; therefore, the percentages do not add up to 100%.

Table 2| Parameter estimates for trial arm, time, and their interaction, intention to treat analysis

Outcome Measure	Complete case cohort (n=759)						Available case cohort (n=1201)					
	Trial arm*		Time†		Time×trial arm		Trial arm*		Time†		Time×trial arm	
	Estimate (SE)	P	Estimate (SE)	P	Estimate (SE)	P	Estimate (SE)	P	Estimate (SE)	P	Estimate (SE)	P
PCS scale (US 1998 NBS)	+1.59 (2.26)	0.480	+0.12 (0.40)	0.771	+0.38 (1.47)	0.795	+2.32 (1.73)	0.181	+0.08 (0.38)	0.843	−0.13 (1.36)	0.923
MCS scale (US 1998 NBS)	−1.55 (2.93)	0.597	−0.83 (0.50)	0.097	−2.37 (1.85)	0.200	−1.74 (2.25)	0.440	−0.70 (0.49)	0.147	−1.02 (1.73)	0.557
EQ-5D scale	−0.01 (0.07)	0.904	+0.01 (0.01)	0.360	+0.01 (0.05)	0.807	+0.05 (0.06)	0.383	+0.01 (0.01)	0.570	−0.01 (0.05)	0.820
Brief STAI scale	+0.18 (1.07)	0.866	−0.01 (0.19)	0.953	+0.21 (0.72)	0.765	+0.10 (0.84)	0.905	+0.05 (0.18)	0.791	+0.09 (0.67)	0.897
CESD-10 scale	−0.35 (1.46)	0.812	+0.04 (0.25)	0.865	+1.34 (0.95)	0.157	−1.20 (1.14)	0.293	−0.01 (0.24)	0.974	+0.99 (0.88)	0.262

PCS=physical component score; MCS=mental component score; NBS=norms based scoring; SE=standard error.

Data are based on multilevel models controlling for the intraclass correlation, all covariates, and the relevant baseline outcome measure. Parameter estimates can be interpreted as the observed difference in an outcome measure (for example, PCS) between levels of a predictor variable (for example, telehealth v usual care) when the intraclass correlation and all covariates are taken into account. For example, parameter estimate +1.59 for trial arm on the PCS scale indicates that patients receiving telehealth had a score 1.59 units higher than patients receiving usual care (reference category) when the intraclass correlation, all covariates, and the baseline PCS score are taken into account.

\*Telehealth=0; usual care=1 (reference category).

†Baseline assessment=1, short term assessment (at four months)=2, long term assessment (at 12 months)=3 (reference category). For the time variable, the main effect tests the hypothesis that outcome measure differs between short and long term assessments while controlling for baseline scores and other covariates, including trial arm (that is, testing the effect of time on the outcome measure, while the effect of trial arm and all other covariates held constant).

Table 3| Parameter estimates for trial arm, time, and their interaction, per protocol analysis

Outcome measure	Complete case cohort (n=633)						Available case cohort (n=1108)					
	Trial arm*		Time†		Time×trial arm		Trial arm*		Time†		Time×trial arm	
	Estimate (SE)	P	Estimate (SE)	P	Estimate (SE)	P	Estimate (SE)	P	Estimate (SE)	P	Estimate (SE)	P
PCS (US 1998 NBS) scale	-1.17 (3.82)	0.761	+0.01 (0.43)	0.975	+2.55 (3.58)	0.476	+1.04 (2.66)	0.696	-0.04 (0.40)	0.926	-0.02 (2.96)	0.994
MCS (US 1998 NBS) scale	+1.74 (4.91)	0.724	-0.87 (0.53)	0.105	-10.88 (4.37)	0.013	+4.54 (3.38)	0.179	-0.63 (0.51)	0.217	-7.07 (3.73)	0.058
EQ-5D scale	-0.12 (0.12)	0.316	+0.02 (0.02)	0.323	+0.14 (0.12)	0.264	-0.05 (0.09)	0.573	+0.01 (0.01)	0.357	+0.08 (0.10)	0.445
Brief STAI scale	-1.93 (1.76)	0.273	-0.15 (0.21)	0.463	+3.10 (1.70)	0.129	-1.29 (1.28)	0.315	-0.11 (0.19)	0.564	+1.62 (1.43)	0.258
CESD-10 scale	-2.63 (2.51)	0.294	-0.02 (0.27)	0.945	+6.41 (2.37)	0.007	-2.53 (1.73)	0.145	-0.12 (0.26)	0.636	+3.65 (1.96)	0.062

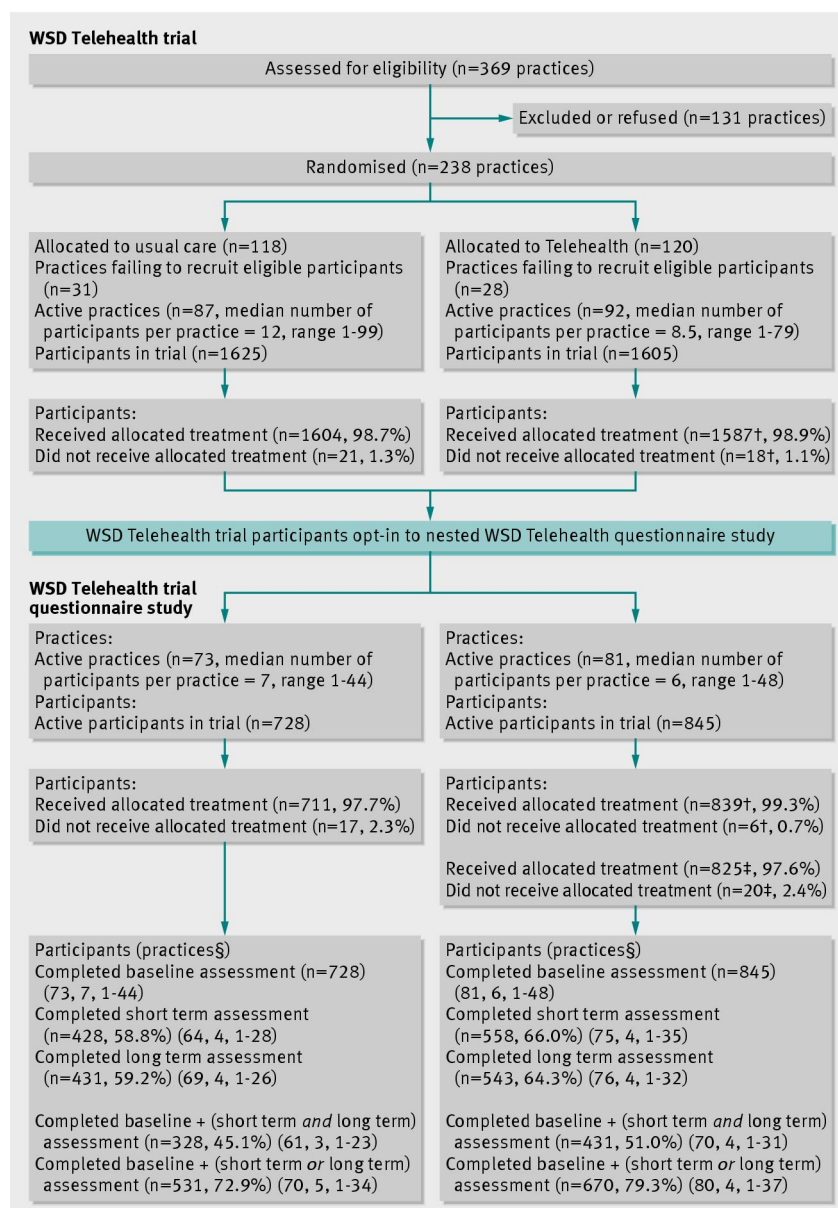
PCS=physical component score; MCS=mental component score; NBS=norms based scoring; SE=standard error.

Data are based on multilevel models controlling for baseline outcome score, all covariates and intraclass correlation. No specific hypotheses were made about the effect of telehealth on particular outcomes at particular time points; therefore, any investigation of time×trial arm interaction terms must be considered exploratory (hypothesis generating) rather than confirmatory (hypothesis testing). The value afforded to such findings when drawing inferences must be weighted accordingly. Moreover, sensitivity analyses across multiple outcomes, cohorts, analytical approaches (intention to treat v per protocol), and parameters (trial arm, time, trial arm×time) leads to the reporting of 60 significance tests (tables 2 and 3). At the stated  $\alpha$  level of 0.05, we would expect three of these to be significant by chance alone, while reducing  $\alpha$  to 0.01 would render one of the two significant interaction term in table 3 (complete case cohort) non-significant. The lack of significant interaction terms in the primary analyses (for both cohorts) and secondary analyses (available case cohort) highlights the general lack of robustness. Furthermore, trial arm×time interaction terms were not significant for PCS, EQ-5D, or CESD-10 in table 3 despite ostensibly measuring closely related constructs. When a trial produces overwhelmingly null results, there is a danger of overemphasising any significant findings, but consideration of the salient factors shows that the two significant interaction terms are not robust, with reasonable likelihood that they reflect chance effects resulting from the additional inclusion criteria applied in the secondary analyses. They should be interpreted with caution.

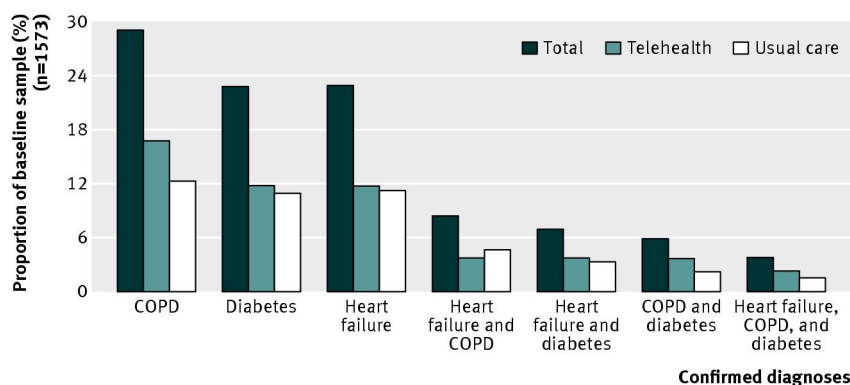
\*Telehealth=0; usual care=1 (reference category).

†Short term assessment (at four months)=2, long term assessment (at 12 months)=3 (reference category). The only a priori hypothesis made about telehealth was that it would improve health related QoL and psychological outcomes relative to usual care.

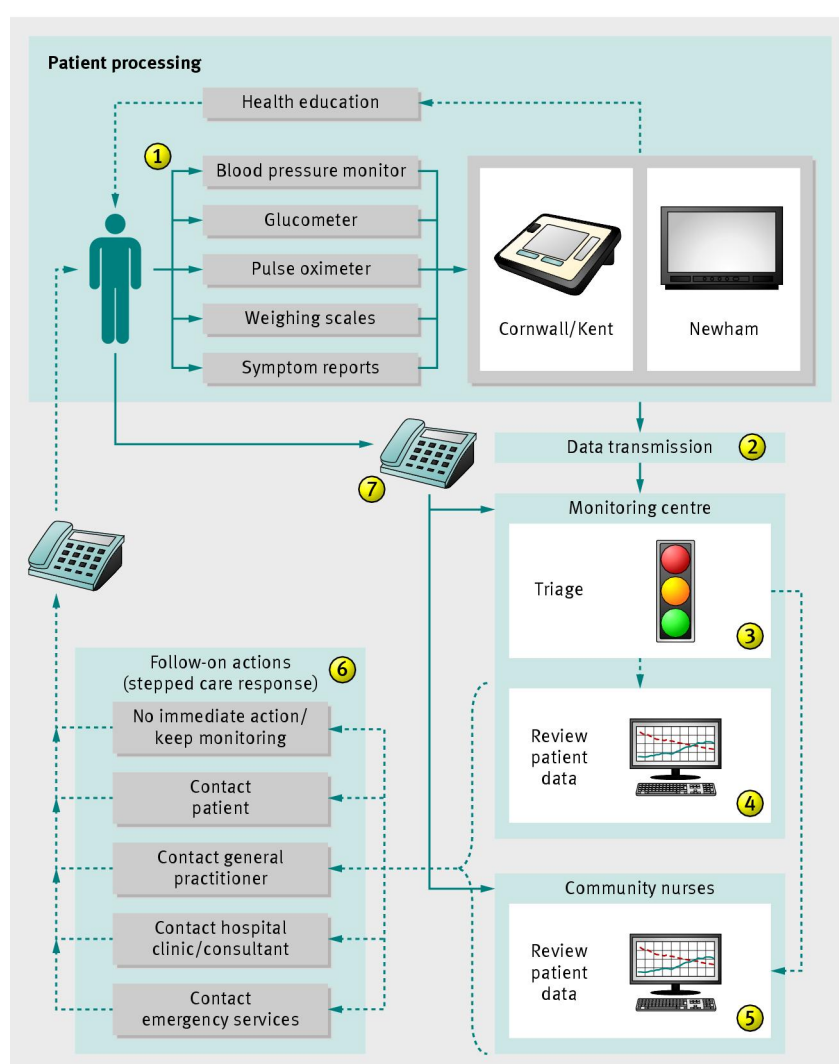
## Figures



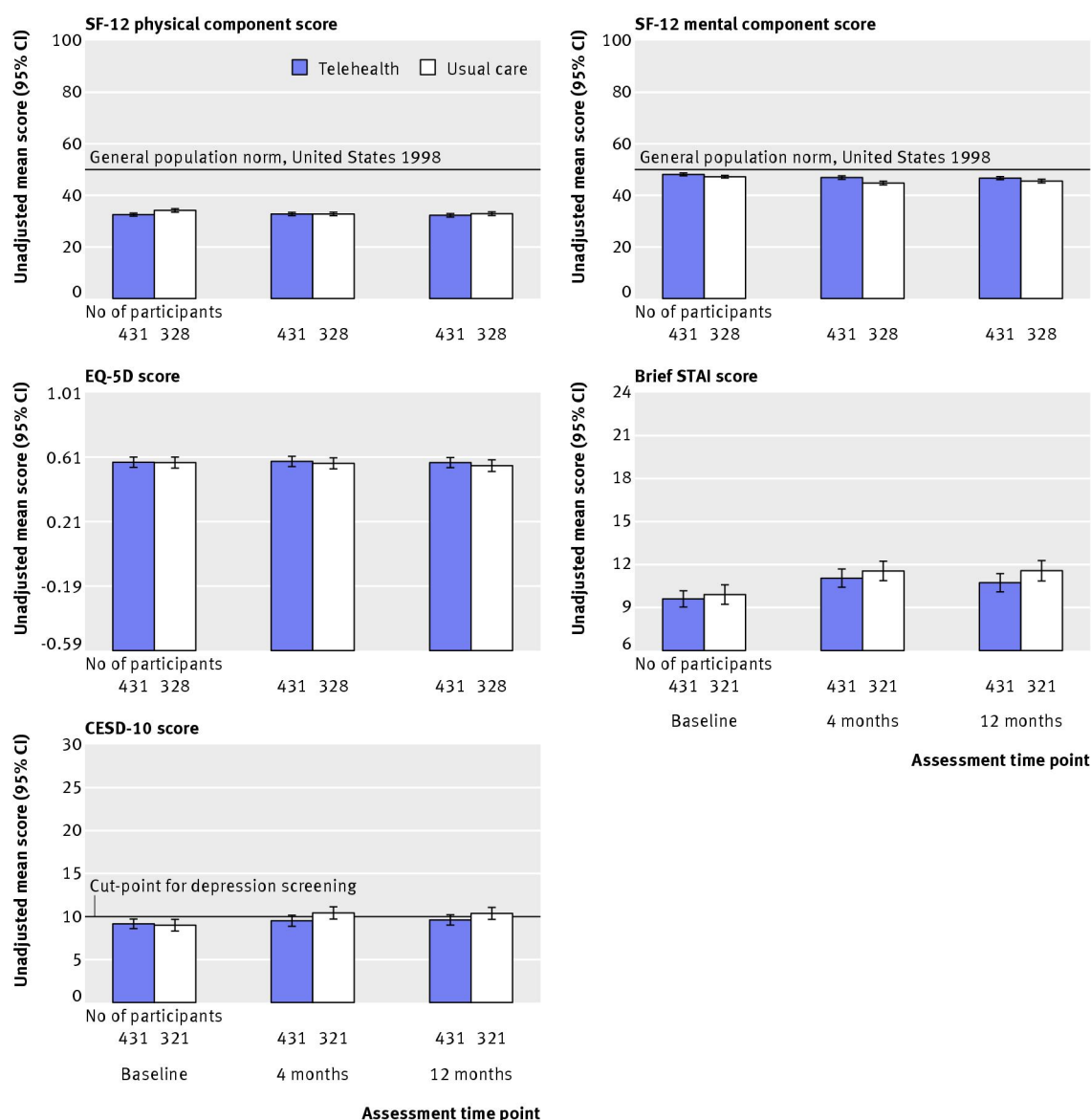
**Fig 1** CONSORT diagram of the WSD telehealth trial and WSD telehealth questionnaire study. \*Recruitment into the questionnaire study was implemented at the patient level, but descriptive data at the cluster level (general practice) are presented for comparison with the parent trial. †Allocated treatment based on the installation of any telehealth device (pulse oximeter, glucometer, weighing scales, or blood pressure monitor) regardless of participant's diagnosed condition. ‡Allocated treatment based on the installation of at least one "critical" telehealth device for a diagnosed condition (web appendix 2). §Second set of brackets show number of active practices, and median number and range of participants per practice



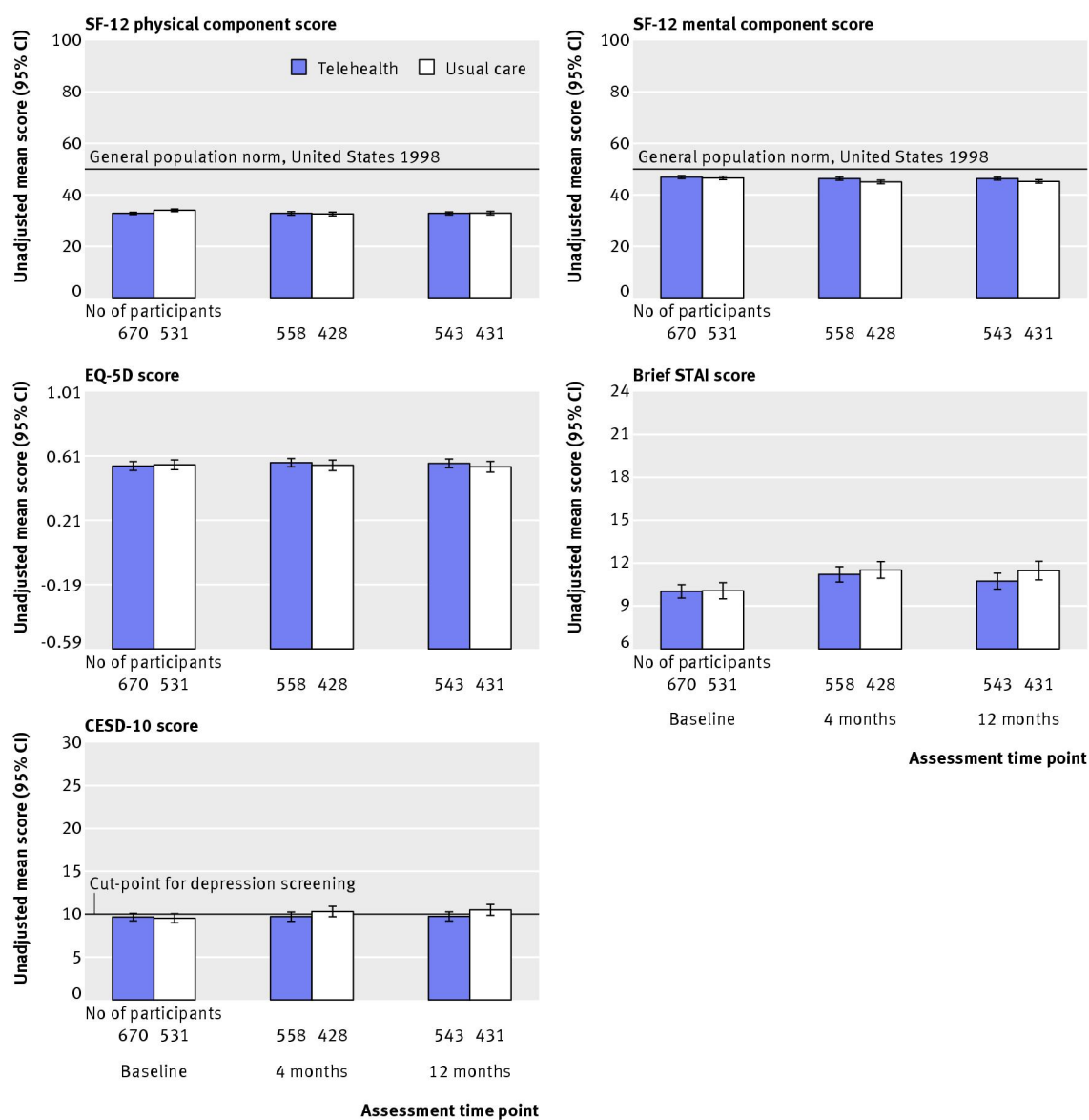
**Fig 2** Composition of baseline sample, by diagnosis of long term condition. COPD=chronic obstructive pulmonary disease



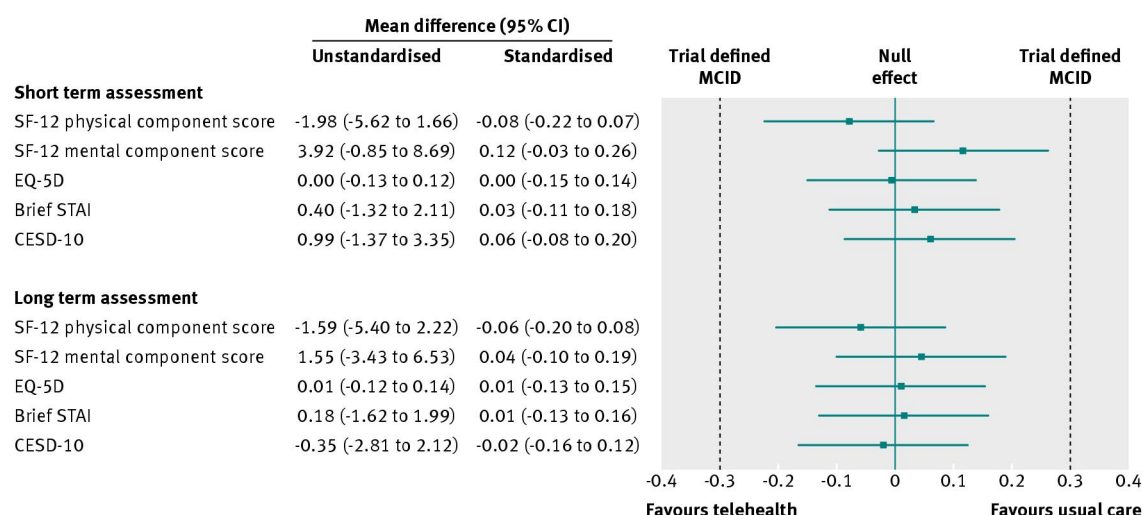
**Fig 3** Overview of WSD telehealth intervention. Numbers indicate stages described in web appendix 2



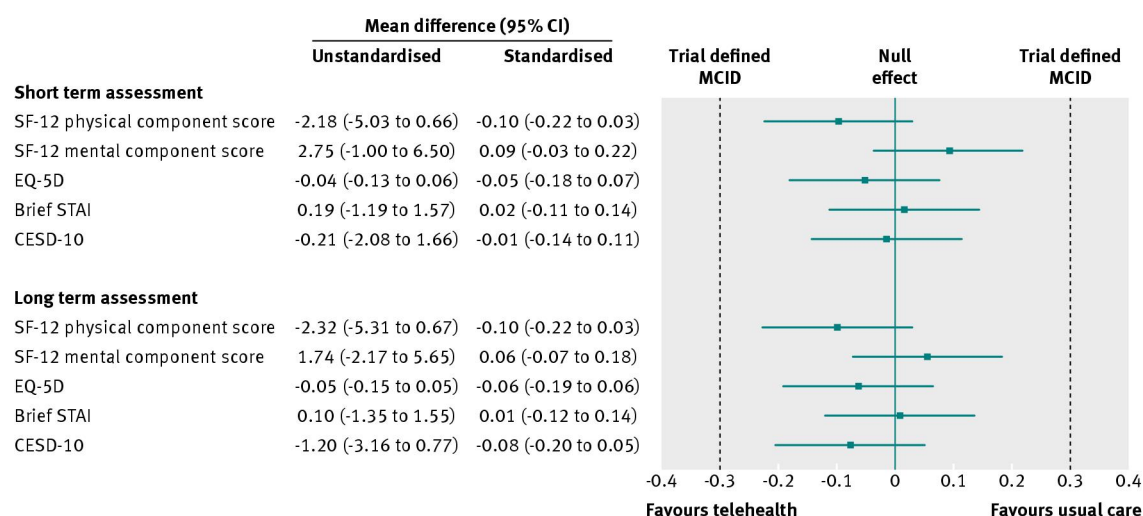
**Fig 4** Outcomes for complete case cohort (n=759). Short term assessment at four months, long term assessment at 12 months



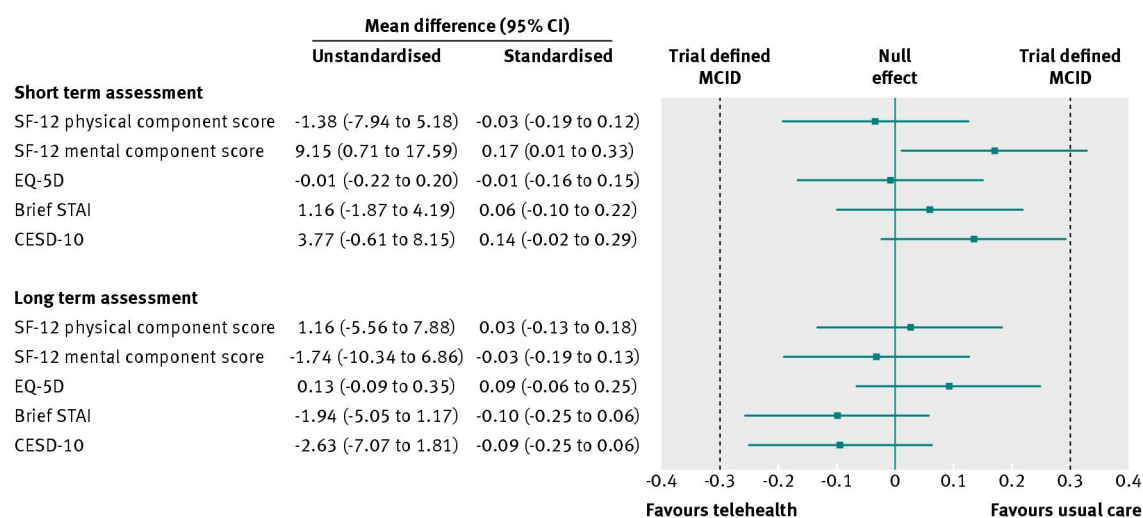
**Fig 5** Outcomes for available case cohort (n=1201). Short term assessment at four months, long term assessment at 12 months



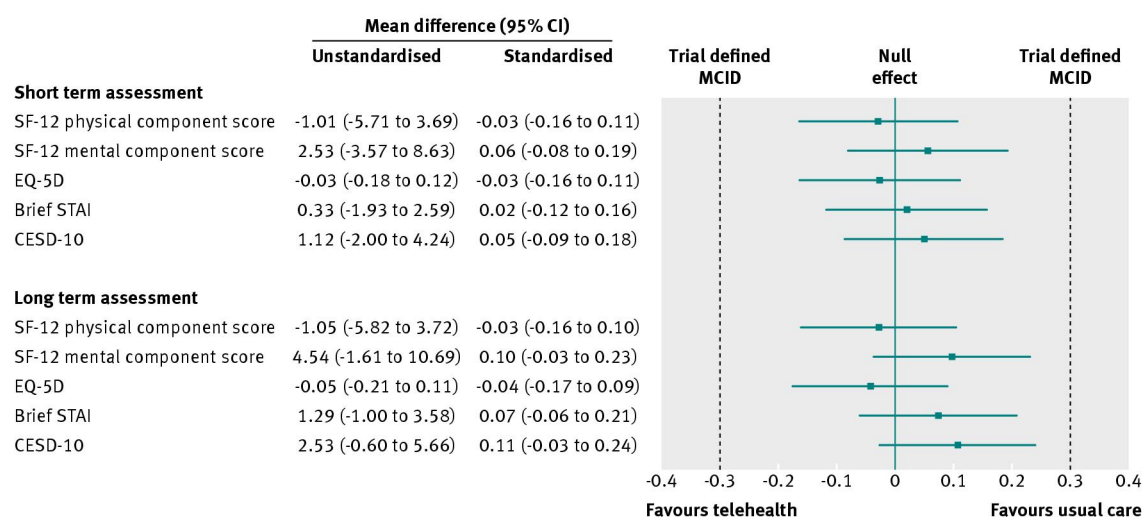
**Fig 6** Standardised adjusted effect sizes for intention to treat analysis, complete case cohort. Effect sizes (mean differences) were calculated at short term and long term, on the basis of EMMs. Web table 3 shows numbers for each arm at each assessment point. Unstandardised mean differences represent the estimated (adjusted) magnitude of difference between arms in the original scale metric. Standardised mean differences allow for direct comparisons across different outcomes. These standardised mean differences were calculated with a correction for sample size (Hedge's g), and are interpreted in the same way as Cohen's d



**Fig 7** Standardised adjusted effect sizes for intention to treat analysis, available case cohort. Web table 3 shows numbers for each arm at each assessment point



**Fig 8** Standardised adjusted effect sizes for per protocol analysis, complete case cohort. Web table 4 shows numbers for each arm at each assessment point



**Fig 9** Standardised adjusted effect sizes for per protocol analysis, available case cohort. Web table 4 shows numbers for each arm at each assessment point