Results of a feasibility randomised controlled trial (RCT) for WATCH IT: a programme for obese children and adolescents

Maria Bryant[a], Amanda Farrin[a], Deborah Christie[b], Susan A Jebb[c], Ashley R Cooper[d] and Mary Rudolf[e]

Background   In the evaluation of childhood obesity interventions, few researchers undertake a rigorous feasibility stage in which the design and procedures of the evaluation process are examined. Consequently, phase III studies often demonstrate methodological weaknesses.

Purpose Our aim was to conduct a feasibility trial of the evaluation of WATCH IT, a community obesity intervention for children and adolescents. We sought to determine an achievable recruitment rate; acceptability of randomisation, assessment procedures, and dropout rate; optimal outcome measures for the definitive trial; and a robust sample size calculation.

Method Our goal was to recruit 70 participants over 6 months, randomise them to intervention or control group, and retain participation for 12 months. Assessments were taken prior to randomisation and after 6 and 12 months. Procedures mirrored those intended for a full-scale trial, but multiple measures of similar outcomes were included as a means to determine those most appropriate for future research. Acceptability of the research and impact of the research on the programme were ascertained through interviewing participants and staff.

Results We recruited 70 participants and found that randomisation and data collection procedures were acceptable. Self-referral (*via* media promotion) was more effective than professional referral. Blinding of assessors was sustained to a reasonable degree, and optimal outcome measures for a full-scale trial were identified. Estimated sample size was significantly greater than sample sized reported in published trials. There was some negative impact on the existing programme as a result of the research, a lesson for designers of future trials.

Limitations We successfully recruited socially disadvantaged families, but the majority of families were of White British nationality. The composition of the participants was an added valuable lesson, suggesting that recruitment strategies to obtain a more heterogeneous ethnic sample warrant consideration in future research.

Conclusions This study provided us with confidence that we can run a phase III multi-centre trial to test the effectiveness of WATCH IT. Importantly, it was invaluable in informing the design not only of that trial but also of future evaluations of childhood obesity treatment interventions. *Clinical Trials* 2011; **8**: 755–764. http://ctj.sagepub.com

[a]Clinical Trials Research Unit, University of Leeds, Leeds, UK, [b]Child and Adolescent Psychological Services, University College London, London, UK, [c]Human Nutrition Research, Medical Research Council, Cambridge, UK, [d]Exercise, Nutrition and Health Sciences, University of Bristol, Bristol, UK, [e]Department of Paediatrics, University of Leeds, Leeds, UK
Author for correspondence: Maria Bryant, Clinical Trials Research Unit, University of Leeds, Leeds LS2 9JT, UK
Email: m.j.bryant@leeds.ac.uk

Introduction

The need to intervene to reduce the high prevalence of childhood obesity has prompted policy makers and researchers to identify effective options for its treatment. There remains an uncertainty in the research literature as to the best treatment approach, and there is still a lack of well-designed, robust trials that evaluate the effectiveness of interventions [1]. In the most recent Cochrane review for treatment interventions, there was some agreement that multi-component lifestyle interventions were optimal [1], but there were only a limited number of high-quality studies and few met the criteria for inclusion in the meta-analysis. Most of the 64 identified studies demonstrated methodological weakness in one or more areas, including inadequate group allocation, lack of blinding, and poorly reported follow-up and dropout rates. Thus, in their bid to find answers, researchers must conduct rigorous evaluations of such complex interventions as set out in the UK Medical Research Council (MRC) framework for complex interventions [2]. Within this model, feasibility trials are emphasised as essential to examine the viability of conducting definitive trials and to ensure that necessary refinements can be made to interventions and study procedures. Such protocols should mirror the procedures intended in the full trial [2], enabling researchers to identify and rectify shortcomings, determine appropriate outcome measures, and calculate an appropriate sample size.

WATCH IT is a programme for obese children and adolescents, set up in 2002 and delivered in a community setting. It was formally piloted in an uncontrolled study in 2004 [3–6], showing proof of concept and leading to funding to conduct a feasibility randomised controlled trial (RCT) to support development of a full-scale multi-centre trial to test its effectiveness. The programme was developed in Leeds, has been extended to Birmingham, and is now included in the UK Department of Health's National Framework for Weight Management Services [7]. In this study, we aimed to conduct preliminary work needed to inform the design and conduct of a multi-centre trial and to identify and rectify any shortcomings. We sought to determine the best way to recruit young people at an achievable recruitment rate, estimate loss to follow-up, determine the acceptability of randomisation and assessment procedures, assess blinding integrity, determine optimal primary and secondary outcome measures, and calculate a robust sample size.

Methods

Design

With the ultimate goal of conducting a future large multi-centre individually RCT, a feasibility study was developed to test this design in one centre. An assessor-blinded randomised controlled feasibility trial was performed in which participants were allocated to WATCH IT or to a waiting list control group for 12 months. The trial was designed to address the specific objectives relating to study design for a full trial and not to determine effectiveness. Data were collected at baseline, 6, and 12 months.

Intervention

Full details of the WATCH IT intervention are described elsewhere [6]. In short, WATCH IT is a programme for obese children and adolescents, delivered in the community setting by non-professional health trainers. Obesity is most prevalent in the socially disadvantaged, who are notoriously most difficult to reach and treat. Thus, WATCH IT was purposely located within socially disadvantaged areas of Leeds. Use of health trainers, rather than medical professionals, is part of an National Health Service (NHS) initiative to include a cost-effective addition to services aimed at tackling inequalities in health and targeting resources on individuals and areas in greatest need. WATCH IT health trainers are appointed for their personal qualities and communication skills. They do not have to have any formal health qualifications, but they receive 2 weeks of training in the intervention delivery and have ongoing support and supervision from a team leader (an educationalist), a children's nurse, dietician, psychologist, and a paediatrician. The WATCH IT programme takes a motivation-enhancing, solution-focussed approach and is embedded within Primary Care Trust Services in Leeds. The core programme length is 4 months, after which families can opt to continue for a further 4 or 8 months. Child and parent/carer receive weekly individual appointments structured on the Healthy Eating Lifestyle Programme [6] and group physical activity sessions. Individual weekly appointments are provided for young people and their parents, allowing any emotional or social issues that may be affecting the young person's ability to achieve healthy behaviours also to be addressed.

Recruitment and consent

In order to determine whether we could recruit the desired number of participants and ascertain which recruitment methods were most successful, the number of enquiries and the consent rate were analysed. We sought to recruit 70 children over a 6-month period beginning September 2006. This target was chosen pragmatically based on participation in the programme in previous years. Obese children and adolescents considered for enrolment were

aged 8–16 years and had a body mass index (BMI) > 98th percentile value and a parent or carer with fluent spoken English. Children with a medical cause for obesity, severe learning difficulties, significant medical or psychiatric problems, or siblings already enrolled as participants were excluded. Candidate participants were recruited through health professionals or self-referral; publicity was amplified during the recruitment period through media coverage. In addition, the WATCH IT database was searched for children who had been referred previously or enquired about the programme but had not taken part. During the study period, only families agreeing to be part of the research were able to enrol in the WATCH IT treatment programme.

After informed consent (parental consent and child assent) and baseline assessment, participants were randomised to either WATCH IT or a waiting list control for 12 months using a remote automated telephone randomisation system. Randomisation was stratified by BMI standard deviation score (SDS; ≤3.0 vs. >3.0), age (≤12 years vs. >12 years), gender, and maternal level of education (less than General Certificate of Secondary Education (GCSE) or equivalent (attainment reached at the age of 16 years) vs. higher). Full ethical approval was obtained from the Leeds West NHS Research Ethics Committee.

Acceptability of randomisation and assessment procedures

Acceptability of randomisation and procedures was determined by measuring loss to follow-up and by exploring children's and parents' views about their participation in the research. Ten percent of trial participants (including families who had withdrawn from the intervention) were randomly selected to answer structured questions at the 12-month assessment to ascertain their views on the randomisation procedure and outcome measures. Open response questions were used to get a general overview of acceptability (e.g., what has been your overall experience of the WATCH IT assessments? Please try to think about both positive and negative things); prompts were provided (e.g., what about the way that it was set up, so that you had a 50/50 chance of starting WATCH IT straight away, or going onto a waiting list?). Focussed questions concerning the assessments also were asked (e.g., do you have any comments about some of the types of tests and assessments that you or your child were asked to do during each visit?) with prompts (what about the blood tests that your child had? How about the questionnaires; you or your child?). Parents and children were interviewed separately; thus, a total of 14 interviews were conducted. Detailed contemporaneous notes were taken and transcribed at the end of the study by the project manager (M.B.).

Blinding

Baseline assessments were performed before randomisation to avoid bias. Follow-up assessments performed after randomisation were conducted by assessors who were blinded to the treatment allocation for each family. A letter was mailed to participants prior to each assessment appointment explaining the need and importance of concealing their group allocations during assessments. In order to explore how effective our blinding strategy was, we asked our assessors to try to guess the treatment allocation of each family at the end of each assessment. They were also asked to make a note of the point at which they thought that they had guessed the treatment allocation and whether the child or parent revealed the group identity.

Determination of primary outcome measure for the definitive trial

We examined the relationship between BMI, waist circumference, and bioimpedance (BIA) with dual energy X-ray absorptiometry (DXA) to determine the ability of these measures to predict change in adiposity as measured by DXA, so that the most feasible and accurate measure could be included in future studies. Trained researchers measured weight, height, waist circumference, and BIA (HYDRA ECF-ICF model 4200; Xitron technologies, San Diego, CA) and performed a DXA (Lunar Prodigy; GE Medical Systems, Madison, WI) scan at baseline and 6 and 12 months. Height was measured to within 0.1 cm using a wall-mounted Seca stadiometer (Vogel and Halke, Hamburg, Germany). To ensure consistency, two measurements were taken, and an average was used. Whenever they differed by >0.5 cm, a third measurement was taken, and an average of the closest two was used. Weight was measured in light clothing with no shoes (to within 0.1 kg) using a calibrated Seca digital weighing scale (Vogel and Halke). Waist was measured twice at 4 cm above the umbilicus. Whenever measurements were >1.0 cm apart, a third measurement was taken, and an average of the closest two was used.

Selection of secondary outcome measures for the definitive trial.

The following measurements were taken in the children at baseline, 6, and 12 months to determine the feasibility and acceptability of their inclusion for a large multi-centre trial rather than treatment effect: 2-h oral glucose tolerance, lipid level, liver function assay, blood pressure, fitness (step test), and physical activity over a 7-day period measured by accelerometry (Actigraph™). In addition, parental height

and weight were measured. The time points of 6 and 12 months (rather than at completion of the WATCH IT programme's four monthly stages) were chosen for reasons of cost and to allow comparison with other trials. Questionnaires included the WATCH IT Diet questionnaire and Home Food Availability checklist (two questionnaires designed specifically to examine foods aligned to the dietary goals promoted as part of the intervention), Dutch Eating Behaviour Questionnaire [8], Physical Activity Questionnaire for Children (PAC-Q) [9], Robinson School-Based Sedentary Behaviour Questionnaire [10], Paediatric Quality of Life (PedsQoL) [11], Strengths and Difficulties Questionnaire (SDQ) [12], and the Harter Scale of Perceived Social and Cognitive Competence [13]. These measures were chosen because they address key components of the intervention (i.e., diet, eating behaviour, physical activity, sedentary behaviour, and psychological well-being). Multiple measures of similar constructs were administered where possible (e.g., psychological well-being) to determine which measure to include in the definitive trial according to feasibility, acceptability, and sensitivity.

Sample size estimation for full trial

Percent body fat from DXA was used as the primary outcome to estimate sample size for a multi-centre trial. We examined the variability of DXA measurements and estimated the difference between control and intervention measurements over time, as these were unavailable from other studies. The study also provided a more accurate estimate of the likely eventual dropout rate in a larger trial.

Identification of shortcomings

Issues related to the research were monitored throughout the study period in order to identify and rectify any emerging shortcomings that were not previously considered in the study protocol. The project manager maintained a constant dialogue with the data collection staff, clinical staff, and study participants throughout the study period. In addition, clinical staff were asked to report the impact of the research on the WATCH IT programme at regular steering committee meetings.

Data analysis

Feasibility studies are not designed to detect a treatment effect and analysis should be mainly descriptive or concentrate on confidence interval (CI) estimation [14]. Data (mean, CIs) are presented below for interest even though the study was not powered to detect an effect of the intervention. Questionnaire feasibility was determined *via* participant acceptability, mean duration for completion (minutes), and standardised response means (SRM), a measure of the sensitivity of each questionnaire to detect change. SRM was calculated as mean change in scores or values divided by the standard deviation in change scores [15]. The SRM provides the relative magnitude of mean change compared with the variability of change, with higher SRM values indicating greater sensitivity. BMI and waist circumference measurements were converted to SDSs using UK 1990 growth references [16]. For selection of a primary outcome for the multi-centre trial (based on the ability of BMI, waist circumference, and BIA to predict DXA), multiple linear regression modelling was used, starting with a full model and using backwards selection, to identify which variables were predictors of DXA fat mass after adjusting for age and gender. Model assumptions were checked (normally distributed residuals and constant variance) and, when necessary, variables were log transformed to improve model fit. The models that best predicted baseline DXA (those with the most significant predictors and the highest R^2 value) were then applied separately to the 6- and 12-month DXA results in order to see how well the models predicted fat mass. Analyses were performed using SAS software (version 9.1; SAS Institute, Inc., Cary, NC).

Results

Sample

Characteristics of participants at baseline are shown in Table 1. Mean BMI SDS was greater in the control group, and there were more severely obese participants assigned to this group (BMI SDS $\geq$ 3.5). The majority of families were Caucasian and economically disadvantaged. Just more than 50% had an annual household income below £15,000, with 14% earning less than £5000 per year. The majority of mothers were not educated beyond GCSE or equivalent.

Recruitment and consent

Ninety-six children were newly referred during the study recruitment period from September 2006 to March 2007. Sixty-three (66%) of these children were recruited and consented to the trial from these de novo referrals. A further seven families were recruited by identifying and contacting 84 families who had been referred to WATCH IT previously but had not responded to correspondence from the

Table 1. Baseline participant characteristics

	Intervention (N = 35)		Control (N = 35)	
	Mean	SD	Mean	SD
Child age (years)	11.5	1.8	11.3	2.2
Guardian age (years)	40.5	10.2	39.5	6.7
Child BMI SDS	2.86	0.45	3.11	0.47
	N	%	N	%
Gender – boys	13	37	12	34
Gender – female carer	31	89	34	97
Highest level of maternal education				
None	6	17	5	14
GCSE or equivalent[a]	15	43	17	49
A-level or equivalent[a]	5	14	6	17
Degree or higher	9	26	7	20
Child ethnicity				
White	32	91	29	83
South Asian	0	0	3	9
Black	1	3	2	6
Mixed ethnicity	1	3	1	3
Annual household income				
Less than £5000	3	9	5	14
£5000–£14,999	14	40	13	37
£15,000–£35,000	11	31	11	35
More than £35,000	7	20	6	17

[a]General Certificate of Secondary Education (GCSE) and A-levels are national examinations taken at the age of 16 and 18 years, respectively.

clinical staff and who had not joined the programme. We therefore attained our target of 70 participants randomised in a 7-month time frame, that is, 39% of 180 contacts or referrals. Thirty participants (31%) were professionally referred and 66 (69%) were self-referred. More self-referred families consented to participate in the research than professionally referred families (55% and 29%, respectively). The most effective form of promotion was television compared with methods such as newspaper, poster, radio, and email advertisements, with television accounting for 41% of referrals. General practitioners made 79% of professional referrals; the remainder were from school nurses and other professionals.

Acceptability of randomisation and assessment procedures

Figure 1 shows the flow of participants through the study, including those lost to follow-up. Only 22 eligible referrals that could be contacted declined to take part. Of these, one refusal was due to the research process (i.e., the possibility of taking the child out of school for the research assessments); one was due to

a family bereavement; and the remainder were not interested in attending the WATCH IT programme.

Of the 10% of families who were randomly selected to provide feedback at the end of the study, all except one parent reported a positive experience. Parents allocated to the control arm reported that they were disappointed that they had to wait but said they would do so again for the well-being of their child. Despite individual consent provided at the start of the study, one parent voiced ethical concerns about the randomisation procedure. The majority of parents said that the assessments were interesting and that they were happy to be occupied completing questionnaires while their child was engaged in assessments. Children and adolescents were also positive about the assessments and enjoyed doing some of the tests; most stated that the baseline blood test was the worst part of participation. When asked about individual components, few negative comments were made by parents. Many noted that they did not like their child having a blood test but understood the necessity. One parent complained that the child was asked to attend on a school morning, resulting in the need to reschedule appointments to holiday periods. We also asked the children and adolescents about their understanding of the randomisation procedure and all responded that they understood and that they did not mind the risk of having to wait for 12 months.

Blinding

Assessors recorded estimations of participant group allocation for 86 assessments. Overall, they assumed that the majority of assessments were with control participants and correctly guessed the group allocation for 76% of 6-month assessments and 51% of 12-month assessments. Six participants inadvertently revealed their group allocation at 6 months (four from the intervention group) and five did so during the 12-month assessment (one from the intervention group). Of those who accidentally revealed their group allocation at 6 months, allocations were guessed correctly for three participants at 12 months. The most common point at which the assessor guessed the group allocation was during the fitness test, which was the final test. This test is also administered as part of the WATCH IT programme.

Determination of primary outcome measure for the multi-centre trial

Results indicate that BMI ($r = 0.93$; 95% CI 0.88, 0.95; $p < 0.0001$), waist circumference ($r = 0.83$; 95% CI 0.73, 0.89; $p < 0.0001$), and BIA ($r = 0.94$; 95% CI 0.92, 0.97; $p < 0.0001$) strongly correlated with DXA
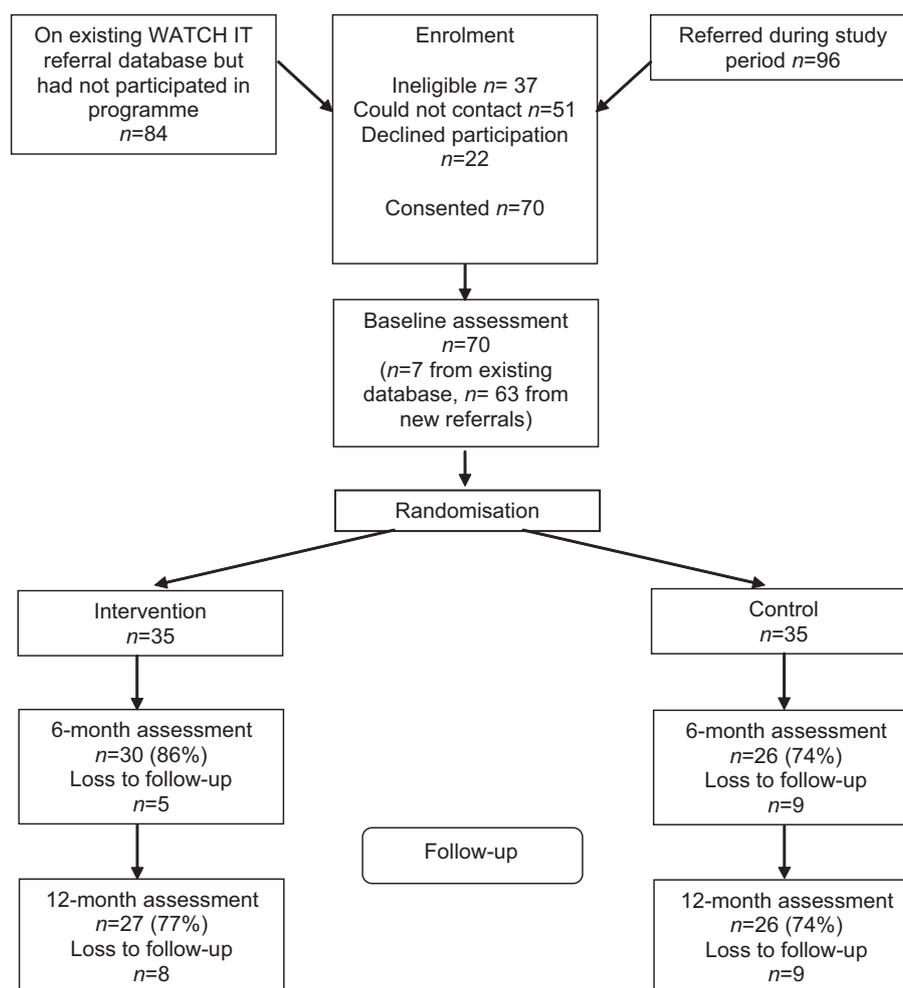
Figure 1. Flow of participants in the feasibility trial of WATCH IT.

results, with a lower (but moderate) correlation of DXA measurements with BMI SDS ($r = 0.55$; 95% CI 0.35, 0.7; $p < 0.0001$) and waist circumference SDS ($r = 0.62$; 95% CI 0.45, 0.75; $p < 0.0001$). Results were similar for cross-sectional and longitudinal analyses, with BIA consistently demonstrating the strongest relationship with DXA and greatest ability to predict future adiposity as measured by DXA.

Selection of secondary outcome measures for the multi-centre trial

No negative comments were made during feedback interviews for any individual assessment. However, data collection staff reported that children often complained or exhibited apathy during completion of the Harter scale and the WATCH IT diet questionnaire, which took longer to complete than the other

questionnaires (Table 2). The Harter scale took twice as long to complete as the other psychological measures, including the SDQ and the PedsQoL. The WATCH IT diet questionnaire had many questions asking about foods and drinks throughout the day, with five simple summary questions aimed at estimating consumption of some key items over the whole day. On examination, there was a moderate correlation between the summary questions and the sum of the detailed individual items (R^2 range 0.5–0.7; $p < 0.001$), indicating that it would be possible and desirable to change the format to focus on the summary questions.

SRMs were calculated to estimate the sensitivity of each questionnaire to detect change. While insufficient data were available to determine a clinically sufficient SRM for each of the tools, we were able to make comparisons between questionnaires. Comparisons were especially important for psychological questionnaires in which more than one outcome was measured. SRM

Table 2. Questionnaire feasibility

	Domains	Administration duration, minutes (range)	SRM[a]
Harter Scale [13]	Self-esteem	9.1 (1–29)	0.01
PedsQoL	Social functioning		
Child		3.5 (1–15)	0.55
Parent		—[b]	0.36
SDQ [12]	Behavioural screening		
Child		4.3 (1–14)	−0.37
Parent		—[b]	−0.51
DEBQ [8]	Eating behaviour	6.3 (1–23)	0.33
WATCH IT[c] diet	Dietary intake	12.5 (1–43)	NA[d]
PAC-Q [9]	Physical activity	9.1 (1–45)	0.14
Robinson screen time	Duration and frequency of screen time	10.2 (1–33)	0.13

[a]Standardised response mean (SRM; sensitivity, calculated by mean change divided by the standard deviation in change scores [15]).
[b]Parent completion was not timed.
[c]Developed specifically for WATCH IT.
[d]Questionnaire does not generate a single score.

values for change between baseline and 12 months for the psychological well-being scales are shown in Table 2, with the PedsQoL demonstrating the greatest sensitivity. Thus, overall, the PedsQoL was considered the most feasible psychological well-being questionnaire, with a relatively low time for completion and the greatest sensitivity.

Valid accelerometer data in which there was at least 500 min of recording for at least 3 days [17] were recorded for 51, 27, and 28 children at baseline, 6, and 12 months, respectively, translating to 73% (51/70), 48% (27/56), and 53% (28/53) of complete accelerometry data for those participants who remained in the trial.

Sample size estimation for full-scale trial

To calculate the target sample size, we assumed that the primary outcome variable for the full-scale trial will be percent body fat at 12 months based on DXA measurement. A standard deviation of 2.3 in change in percent body fat was observed in the feasibility study. In the absence of clinical information regarding what constitutes a clinically important difference in change in percent body fat, we assumed that a difference of 0.75 is important, which translates to a standardised effect size of 0.326, usually considered moderate. Thus, 199 participants per group would be required to provide 90% power at the 5% significance level. However, as the sample size calculations should account for the natural clustering of outcomes by trainer (we estimated sample cluster size to be at most 16 children in the full-scale trial), the sample size was inflated by a design effect of 1.75 (assuming an intraclass correlation coefficient (ICC) of 0.05) to yield a revised total sample size of 69 participants. Allowing for 25% loss to follow-up (as observed in the feasibility study), we estimated

that the multi-centre trial would require 930 participants to provide a definitive indication of clinical effectiveness and to allow a moderate effect size to be detected in other outcomes.

Identification of shortcomings

The research process was found to impact negatively on the WATCH IT programme in a number of ways. Attendance data for the programme were comparable with those of the pilot study [6], with 74% (n = 26) completing the core 4-month phase, and 63% (n = 22) and 46% (n = 16) opting to continue for a further 4 or 8 months, respectively. In order to maximise participation in the feasibility study, entry into WATCH IT was only *via* the trial during the study period and usual enrolment was halted immediately prior to its start. As a result of randomisation of families to the waiting list, the programme ran at reduced capacity. Health trainers reported a disruption to group sessions (especially physical activity sessions), in which they believed that the children were less motivated. The health trainers also complained that the requirement to adhere to protocol constrained them and prevented them from working in a more flexible manner to suit the individual needs of children. Another major influence of the research was that at the end of the trial, capacity had to double to allow the waiting list control participants immediate access to the programme as they had been guaranteed during recruitment.

Change in BMI and adiposity

The trial was not powered to demonstrate effectiveness of the intervention with only 35 children per

treatment arm. Mean change in BMI SDS was 0.03 (95% CI –0.05, 0.11) in the intervention group and –0.03 (95% CI –0.12, 0.06) in the control group. Change in percent body fat was 1.40 (95% CI 0.31, 2.38) for the intervention group and 0.20 (95% CI –1.41, 1.72) for the control group. Mean change in waist circumference SDS was –0.08 (95% CI –0.24, 0.07) in the intervention group and –0.03 (95% CI –0.16, 0.11) in the control group.

Discussion

The findings from our feasibility trial can be used to inform the design and conduct of trials in childhood obesity research. We were able to recruit the desired number of participants within a desirable time frame and showed acceptable loss to follow-up for the research that was comparable with other similar research protocols [1]. Furthermore, our recruitment strategy resulted in adequate numbers of children from socially disadvantaged families, a group that are more likely to be obese, yet can be difficult to access and recruit into research. However, the majority of families were of White British nationality; we therefore learnt that recruitment of a more heterogeneous ethnic sample warrants consideration in future research. The research methodology was acceptable to participants, and the recruitment strategies were successful. However, returning to an old database to recruit families who had been referred previously to an ongoing programme but had failed to participate in it was not productive.

Implementation of the feasibility trial has enabled us to develop a robust protocol for a definitive multi-centre trial. However, since the feasibility trial was only conducted at one centre, we recognise that there may be other feasibility issues with multiple centres that were not measurable or observed at a single centre. Key findings did not necessarily form part of the formal hypotheses of the trial. However, the feasibility trial enables us to address issues that were not been considered prior to the conduct of large multi-centre trials. Potential shortcomings, related to pragmatic issues associated with data collection, were identified. For example, initiation of an intervention was dependent upon the needs and availability of both the family and the programme, and the actual intervention start date could not always immediately follow baseline assessment. Thus, baseline measurements may have changed prior to the actual start of the intervention. Designers of future protocols should consider adopting a flexible start date for the intervention. Similarly, follow-up assessments may not be made exactly as specified in the research timeline because of holidays or sickness. Other key lessons for the design of an evaluation of a complex intervention in obese children are

that blinding of assessors is important to reduce bias; automated electronic randomisation improves efficiency and separates responsibility from the researcher; children should be interviewed without their parents; and careful consideration should be placed on the ordering of assessments. We discovered that assessment blinding was most likely to become compromised during a test that also was administered as part of the intervention programme. Thus, future study designs should ensure tests which can reveal group allocation (i.e., those most likely to stimulate conversation related to the intervention) should be administered late in the assessment. Our initial protocol did not consider interviewing children separately from their parents. It soon became apparent, however, that separation would be necessary in order to get responses to questionnaires directly from the child, rather than *via* the parent. In the first two interviews, it was obvious that parents found it difficult to avoid answering the questions for their children.

Retention strategies were not formalised within the protocol, but we had an acceptable level of dropout (24% withdrawal overall). On reflection, we believe that establishing a good rapport with both children and their parents was essential, along with providing full details of the level of involvement required as part of participation in both the intervention and research. Recruitment and consent occurred within a clinical setting and was conducted by a childhood obesity expert, helping to formalise the value and prestige of participation.

The feasibility trial also provided an excellent test bed for potential outcome measures for a definitive trial. We used DXA to measure fat mass as a primary outcome, which incurred a cost of £14,700. We estimated that its use in the definitive trial (based on the suggested sample size of 930) would cost £195,300. Financial constraints and availability of a DXA apparatus therefore may preclude it use. Since DXA data were well correlated with BMI, BIA, and waist circumference, which primary outcome to use may be a pragmatic choice. Secondary outcomes in childhood obesity research are notoriously difficult to measure. Self-report methods often are biased systematically and objective measures are usually expensive and often infeasible. There is a lack of consistency in the use of outcome measures, in part, due to a lack of valid tools. Importantly, through this work, we have confirmed that more research is required to have confidence in the secondary outcome measures. While we found the measures to be feasible, we believe that more research is necessary to ascertain the validity and reliability of outcomes within a paediatric obese population. The UK National Obesity Observatory has published a standard evaluation framework to guide proficient research for assessing childhood

obesity interventions, but it is evident that one of the greatest barriers at present is the lack of valid and reliable outcome measures [18].

Accelerometry is accepted as a useful and objective method to measure habitual physical activity [19–21]. However, in a community-based paediatric sample, we had many monitors that were returned without valid data. Using relatively relaxed criteria for data quality of at least 500 min of recording for at least 3 days [17], we received only 53% of accelerometers with valid data at 12 months. Given the high degree of resources required from researchers and participants to collect accelerometry data, future researchers with this type of population should consider carefully whether their collection is cost-effective. Lower criteria might have to be accepted; for example, rather than choosing a pre-defined inclusion criterion (e.g., minimum of 10 h a day), Alhassan *et al.* [22] calculated the number of minutes required to obtain a correlation coefficient of 0.80 for the average count per minute of randomly selected blocks of 30 min correlated with average counts for the full day.

Importantly, this feasibility trial enabled us to estimate sample size for a future multi-centre RCT. Based on our calculations, we estimated that 930 children would be required to detect a statistically significant moderate difference between the intervention and control group, a larger sample size than those of studies in the recent Cochrane report on interventions for treating obesity in children [1], in which, out of 64 trials reported, the average number of randomised participants was 85 with a range of 16–539 participants. Because of expected clustering by trainer, we inflated the nominal sample size by assuming an ICC of 0.05. Sample size was calculated assuming individual level randomisation, in addition to the inflation factor. Data from the feasibility study enabled us to realistically estimate variation, in addition to likely intraclass correlation of outcomes.

Assessing the influence of interventions according to efficacy (i.e., assessment in people who received the assigned intervention and adhered to the protocol) and effectiveness (i.e., assessment in all people assigned to the intervention, regardless of protocol adherence) [23] is challenging since the difference between efficacy and effectiveness is not clear in the evaluation of complex interventions and they may not be mutually distinct. Rather, owing to the complexities of behavioural change interventions, they may be viewed as a continuum. For example, we are confident that our study design meets scientific rigour in terms of optimising independence of the treatment effect *via* implementation of carefully allocated active and control conditions, appropriately stratified randomisation, use of standardised treatment and evaluation protocols, generation of a homogeneous comparison group, and optimal blinding.

Demonstration of such study characteristics lends itself to provision of evidence of efficacy. However, the nature of the intervention and its delivery within the NHS results in uncertainty of the level of adherence to the interventions regardless of provider fidelity to the intervention. Because it is not possible to consider all potential confounding factors within a community-based intervention for which participants are expected to make lifestyle changes, efficacy cannot be guaranteed. Ultimately, the WATCH IT programme aims to provide services across the UK. Thus, it is imperative that it demonstrate effectiveness (rather than efficacy) under 'real-world' conditions or in 'natural' settings [24]. Furthermore, it has been argued that trials demonstrating efficacy are not essential before effectiveness trials are conducted, provided that the effectiveness trial meets all of the standards of efficacy trials.

The feasibility trial allowed us to identify the impact of research on the normal operation of the intervention programme. With initiation of the feasibility study, the programme initially was constrained to operate at lower than usual capacity due to randomisation. It then had to increase capacity at the end of the trial to allow entry of the waiting list control group participants. Recognition of these issues has informed the design of future trials, which must address the impact of randomisation on programmes during the trials and at their conclusion. Surprisingly, we also discovered through the health trainers that families who had started WATCH IT before the trial and were not involved in the research (but who were integrated with the trial participants) were disappointed that their clinical assessment lacked the depth that was given in the trial assessments. Such information has enabled us to consider the importance of working with the service providers and service users in the design of future trials; it is possible that future trials will be conducted only in areas where there is no established service. A constant dialogue was maintained with the service providers throughout the study period to help the researchers to understand the impact of the study on the intervention. In addition, we elicited formal feedback from at least one service provider during regular steering committee meetings. While this feedback ensured that we met the feasibility study objectives, we recommend continued service provider involvement in the design of future RCTs in order to gain greater understanding of the processes involved in generating a behavioural change.

To conclude, this feasibility trial has been invaluable in informing the design of future research for WATCH IT, and importantly, it has emphasised the necessity for any childhood obesity intervention research to follow guidelines for the evaluation of complex interventions. We hope that lessons learned from the conduct of our feasibility RCT will help

future researchers. However, we recommend that all researchers evaluating the effect of obesity interventions conduct an exploratory feasibility or pilot study prior to designing definitive trials of effectiveness. Feasibility testing of obesity interventions is common in the literature, but there remains a lack of studies that test the feasibility of conducting the research. Those that do provide valuable assistance to future researchers [25] but cannot substitute the conduct of feasibility trials related to the evaluation of specific interventions. Our results have highlighted the appropriateness of following a cyclical sequence of research pathways, as suggested by the MRC [2], so that additional formative research can be considered prior to large-scale testing if necessary. In doing so, researchers and clinical staff can avoid wasting time and/or resources on running definitive trials before they are 100% confident in both the research and the intervention.

Funding

References

1. **Oude Luttikhuis H, Baur L, Jansen H**, *et al*. *Interventions for Treating Obesity in Children (Cochrane Review)*. Wiley & Sons, Chichester, UK, 2009.

2. **Craig P, Dieppe P, Macintyre S**, *et al*. Developing and evaluating complex interventions: new Medical Research Council guidance. *BMJ* 2008; **337**: a1655.

3. **Dixey R, Rudolf M, Murtagh J**. WATCH IT: obesity management for children: a qualitative exploration of the views of parents. *Int J Health Promot Educ* 2006; **44**: 131–7.

4. **McElhone S, Rudolf MCJ**. What sort of quality of life do obese children and adolescents in the UK have? *Arch Dis Child* 2005; **90**(suppl. 11): A49.

5. **Murtagh J, Dixey R, Rudolf M**. A qualitative investigation into the levers and barriers to weight loss in children: opinions of obese children. *Arch Dis Child* 2006; **91**: 920–3.

6. **Rudolf M, Christie D, McElhone S**, *et al*. Watch It: a community based programme for obese children and adolescents. *Arch Dis Child* 2006; **91**: 736–9.

7. **Cross Government Obesity Unit.** Healthy weight, healthy lives: one year on. COI for the Department of Health, 2009.

8. **VanStrien T, Frijlers J, Bergers G, Defares P**. The Dutch Eating Behaviours Questionnaire (DEBQ) for assessment of restrained, emotional and external eating. *Int J Eat Disord* 1986; **5**: 747–55.

9. **Crocker PRE, Bailey DA, Faulkner RA, Kowalski KC, McGrath R.** Measuring general levels of physical activity: preliminary evidence for the Physical Activity Questionnaire for children. *Med Sci Sports Exerc* 1997; **29**: 1344–9.

10. **Robinson TN.** Reducing children's television viewing to prevent obesity. A randomized controlled trial. *JAMA* 1999; **282**: 1561–7.

11. **Varni J, Seid M, Kurtin, P.** PedsQL™4.0: Reliability and Validity of the Pediatric Quality of Life Inventory™ Version 4.0 Generic Core Scales in Healthy and Patient Populations. *Medical Care 2001*; **3**(8):800–812.

12. **Goodman R.** The Strengths and Difficulties Questionnaire: a research note. *J Child Psychol Psychiatry* 1997; **38**: 581–6.

13. **Harter S.** The perceived competence scale for children. *Child Development* 1982; **53**: 87–97.

14. **Lancaster G, Dodd S, Williamson P.** Design and analysis of pilot studies: recommendations for good practice. *J Eval Clin Pract* 2002; **10**: 307–12.

15. **Stratford PW, Binkley JM, Solomon P**, *et al*. Defining the minimum level of detectable change for the Roland Morris Questionnaire. *Phys Ther* 1996; **76**: 359–65.

16. **Cole TJ, Freeman JV, Preece MA.** Body mass index reference curves for the UK. *Arch Dis Child* 1990; **1995**(73): 25–9.

17. **Janz KF, Burns TL, Torner JC**, *et al*. Physical activity and bone measures in young children: the Iowa Bone Development Study. *Pediatrics* 2001; **107**: 1387–93.

18. **Roberts K, Cavill N, Rutter H.** *Standard Evaluation Framework for Weight Management Interventions*. National Obesity Observatory, Oxford, 2009.

19. **Cliff DP, Reilly JJ, Okely AD.** Methodological considerations in using accelerometers to assess habitual physical activity in children aged 0-5 years. *J Sci Med Sport* 2009; **12**: 557–67.

20. **Elberg J, McDuffie JR, Sebring NG**, *et al*. Comparison of methods to assess change in children's body composition. *Am J Clin Nutr* 2004; **80**: 64–9.

21. **Stevens J, Murray DM, Catellier DJ**, *et al*. Design of the Trial of Activity in Adolescent Girls (TAAG). *Contemp Clin Trials* 2005; **26**: 223–33.

22. **Alhassan S, Sirard JR, Spencer TR, Varady A, Robinson TN.** Estimating physical activity from incomplete accelerometer data in field studies. *J Phys Act Health* 2008; **5**: S112–S25.

23. **Jadad AR, Enkin MW.** *Randomized Controlled Trials: Questions, Answers, and Musings*. Blackwell Publishing: Oxford, UK, 2007.

24. **Flay B, Biglan A, Boruch R**, *et al*. Standards of evidence: criteria for efficacy, effectiveness and dissemination. *Prev Sci* 2005; **6**: 151–75.

25. **Warren JM, Golley RK, Collins CE**, *et al*. Randomised controlled trials in overweight children: practicalities and realities. *Int J Pediatr Obes* 2007; **2**: 73–85.