Open Access: free full text available online
For the full versions of these articles see bmj.com

RESEARCH

Minimal access surgery compared with medical management for chronic gastro-oesophageal reflux disease: UK collaborative randomised trial

Adrian M Grant,[1] Samantha M Wileman,[1] Craig R Ramsay,[1] N Ashley Mowat,[2] Zygmunt H Krukowski,[2] Robert C Heading,[3] Mark R Thursz,[4] Marion K Campbell,[1] and the REFLUX Trial Group

EDITORIAL by McPherson

[1]Health Services Research Unit, Health Sciences Building, University of Aberdeen, Foresterhill, Aberdeen AB25 2ZD
[2]Aberdeen Royal Infirmary, Foresterhill, Aberdeen AB25 1LD
[3]Department of Gastroenterology, Royal Infirmary, Glasgow G4 0SF
[4]Faculty of Medicine, Imperial College, St Mary's Campus, London W2 1PG
Correspondence to: A M Grant
a.grant@abdn.ac.uk

Cite this as: *BMJ* 2008;337:a2664
doi:10.1136/bmj.a2664

This article is an abridged version of a paper that was published on bmj.com. Cite this article as: *BMJ* 2008;337:a2664

ABSTRACT

Objective To determine the relative benefits and risks of laparoscopic fundoplication surgery as an alternative to long term drug treatment for chronic gastro-oesophageal reflux disease (GORD).

Design Multicentre, pragmatic randomised trial (with parallel preference groups).

Setting 21 hospitals in the United Kingdom.

Participants 357 randomised participants (178 surgical, 179 medical) and 453 preference participants (261, 192); mean age 46; 66% men. All participants had documented evidence of GORD and symptoms for >12 months.

Intervention The type of laparoscopic fundoplication used was left to the discretion of the surgeon. Those allocated to medical treatment had their treatment reviewed and adjusted as necessary by a local gastroenterologist, and subsequent clinical management was at the discretion of the clinician responsible for care.

Main outcome measures The disease specific REFLUX quality of life score (primary outcome), SF-36, EQ-5D, and medication use, measured at time points equivalent to three and 12 months after surgery, and surgical complications.

Main results Randomised participants had received drugs for GORD for median of 32 months before trial entry. Baseline REFLUX scores were 63.6 (SD 24.1) and 66.8 (SD 24.5) in the surgical and medical randomised groups, respectively. Of those randomised to surgery, 111 (62%) actually had total or partial fundoplication. Surgical complications were uncommon with a conversion rate of 0.6% and no mortality. By 12 months, 38% (59/154) randomised to surgery (14% (14/104) among those who had fundoplication) were taking reflux medication versus 90% (147/164) randomised medical management. The REFLUX score favoured the randomised surgical group (14.0, 95% confidence interval 9.6 to 18.4; P<0.001). Differences of a third to half of 1 SD in other health status measures also favoured the randomised surgical group. Baseline scores in the preference for surgery group were the worst; by 12 months these were better than in the preference for medical treatment group.

Conclusion At least up to 12 months after surgery, laparoscopic fundoplication significantly increased measures of health status in patients with GORD.

Trial registration ISRCTN15517081.

INTRODUCTION

There is wide agreement that proton pump inhibitors are the most effective treatment for moderate to severe gastro-oesophageal reflux disease (GORD). They can, however, cause a spectrum of short term symptoms,[1] and there are concerns about long term acid suppression.[2] Surgery, especially using a minimal access laparoscopic approach, is an alternative to long term medical treatment. Although fundoplication produces resolution of reflux symptoms in up to 90% of patients,[3] exchanging symptoms associated with medical management for those of the side effects of surgery might not be advantageous for the patient nor a good use of healthcare resources.

We carried out a multicentre pragmatic randomised trial (with parallel non-randomised preference groups)[4] evaluating the clinical effectiveness, safety, and costs of a policy of relatively early laparoscopic surgery compared with optimised medical management of GORD for people judged suitable for both policies.

METHODS

Patients were eligible if they had more than 12 months' symptoms requiring maintenance treatment with a proton pump inhibitor (or alternative) for reasonable control; they had endoscopic or 24 hour pH monitoring evidence of GORD, or both; they were suitable for either policy; and the recruiting doctor was uncertain which management policy to follow. We invited any eligible patient who did not want to take part in the randomised trial because of a strong preference about treatment to join a non-randomised preference arm.

Clinical management

Participating clinical centres had partnerships between surgeons and gastroenterologists and shared the secondary care of patients with GORD. They informed

participants about the randomised trial. Participants who declined to take part in the randomised trial, because of a strong preference either for remaining on medical treatment or for undergoing surgery, were then given the opportunity to take part in the preference study.

For all participants in either the randomised or preference surgical group, surgery could be subsequently deferred or declined, by either the participant or surgeon (that is, even after trial entry). A lead surgeon (or a surgeon working under supervision) undertook the surgery. The type of fundoplication was left to the discretion of the surgeon. For the main comparisons, we considered the different fundoplication techniques as a single policy. Those allocated to medical treatment had their treatment reviewed and adjusted as necessary by a local gastroenterologist to be "best medical management." The medical protocol included the option of surgery if a clear indication developed after randomisation.

Outcome measures

Outcome measures were those judged important to patients and health services. The primary outcome was the REFLUX questionnaire score, a validated "disease specific" measure incorporating assessment of reflux and other gastrointestinal symptoms and the side effects and complications of both treatments.[5] Five symptom scores were also developed as secondary measures. Participants were followed up by postal questionnaire three and 12 months after surgery or at an equivalent time among those who did not have surgery. Issues related to long, variable length, waiting lists prohibited timing of follow-up from randomisation. Other outcome measures were health status (EQ-5D, and SF-36); serious morbidity, such as operative complications; and mortality.

Sample size

Lower recruitment than expected meant that we recalculated the power calculation and ultimately needed 196 in each group to give 80% power (α=0.05), assuming 10% attrition (that is, 176 per group with 12 month follow-up). See bmj.com.

Statistical methods

Our primary analysis in the randomised trial was by intention to treat. General linear models adjusted for the minimisation covariates (age, BMI, and sex) and, when appropriate, for baseline score and interaction between baseline score and treatment. We also included a covariate to adjust for a change in practice regarding the baseline questionnaires. Because a relatively large proportion of the randomised surgical participants did not receive surgery, we also performed per protocol analyses and analyses adjusted for treatment received to estimate efficacy of treatment. For the preference groups, we analysed statistically only the REFLUX score. See bmj.com.

RESULTS

Recruitment took place in 21 centres from March 2001 to June 2004: 357 participants were in the randomised arm (178 allocated to surgery and 179 to medical management) and 453 in the preference arm (261 chose surgery and 192 chose medical management). See figure on bmj.com. Three and 12 month follow-up questionnaires were received from 86% and 89%, respectively. Three participants died, two in the preference for surgery group and one in the randomised medical group; none had surgery. All deaths were unrelated to trial participation.

The characteristics of the randomised participants were similar and lay between those in the preference groups; participants in the preference for surgery group were younger and had been prescribed medication for GORD for longer; participants in the preference for medical treatment group were older, more likely to be women, and had been prescribed medication for a shorter time. See bmj.com.

Surgical management—In total, 111 (62%) of those randomised to surgery and 218 (84%) participants in the preference for surgery group actually received fundoplication. See bmj.com.

Antireflux medication—By 12 months after surgery, 38% (59/154) of the randomised surgical participants were taking medication compared with 90% (147/164) of the randomised medical participants (nearly all of whom were taking proton pump inhibitors). Among

Table 1 | Health status at baseline and at three and 12 months after surgery. Figures are means (SD)

	Randomised participants				Preference participants			
	Surgical		Medical		Surgical		Medical	
	Randomised (n=178)	Per protocol (n=111)	Randomised (n=179)	Per protocol (n=169)	Preference (n=261)	Per protocol (n=218)	Preference (n=192)	Per protocol (n=189)
Primary outcome (REFLUX QoL)								
Baseline	63.6 (24.1)	61.9 (24.5)	66.8 (24.5)	68.2 (24.2)	55.8 (23.2)	55.9 (23.2)	77.5 (19.7)	78.0 (19.1)
3 months	83.9 (19.4)	85.9 (19.0)	70.6 (24.6)	70.8 (24.4)	80.4 (21.6)	82.5 (20.3)	80.2 (18.2)	80.6 (17.7)
12 months	84.6 (17.9)	88.3 (15.6)	73.4 (23.3)	73.1 (23.7)	83.3 (20.7)	86.0 (17.9)	79.2 (19.2)	79.4 (19.0)
Secondary outcome (EQ-5D index)								
Baseline	0.71 (0.26)	0.72 (0.24)	0.72 (0.25)	0.73 (0.25)	0.68 (0.26)	0.68 (0.26)	0.75 (0.22)	0.75 (0.22)
3 months	0.79 (0.23)	0.81 (0.24)	0.69 (0.30)	0.70 (0.30)	0.81 (0.25)	0.82 (0.24)	0.76 (0.23)	0.77 (0.23)
12 months	0.75 (0.25)	0.78 (0.23)	0.71 (0.27)	0.71 (0.27)	0.79 (0.26)	0.80 (0.25)	0.74 (0.24)	0.74 (0.24)

Table 2 | Difference* (95% confidence interval) in REFLUX quality of life score† at 12 months after surgery in randomised participants

Model	Intention to treat	Per protocol	Adjusted for treatment received
I: adjusted for minimisation variables	11.2 (6.4 to 16.0)	15.4 (10.0 to 20.9)	16.7 (9.7 to 23.6)
II: adjusted for minimisation variables and baseline REFLUX QoL score	14.1 (9.6 to18.6)	19.1 (14.0 to 24.1)	20.3 (13.8 to 26.8)
III: adjusted for minimisation variables, baseline score, and interaction between treatment and baseline REFLUX QoL score	14.0 (9.6 to18.4)	18.4 (13.6 to 23.2)	19.4 (13.0 to 25.8)

*Surgery group minus medical group, all P<0.001.
†Higher scores mean patient felt better (range 0-100).

those who had surgery, use of antireflux medication dropped to 9% at three months and 14% (14/104) at 12 months after surgery.

Health status—There were substantial differences between the randomised intention to treat groups in the REFLUX score with the surgery group having better scores than the medical group (table 1). See bmj.com. Statistical analyses for the primary outcome showed strong evidence of increases in scores favouring surgery (table 2). This reflected improvements across all symptom domains within the measure. There was strong evidence of an interaction effect between randomised group and baseline REFLUX score (interaction term: −0.35, −0.53 to −0.17; P<0.001); as baseline REFLUX score decreased (baseline symptoms were more severe) the treatment effect of surgery increased. Similar patterns in the randomised groups were seen in the SF-36 and EQ-5D scores although with some evidence of attenuation at 12 months.

Possible side effects of surgery—No differences were detected between the trial groups in their questionnaire responses at 12 months regarding "difficulty swallowing" and "bloatedness/trapped wind," but there was some evidence of more frequent "wind from the lower bowel" after surgery.

Preference groups—The participants in the preference for surgery group had lower mean REFLUX scores at baseline than those in the preference for medical treatment group (55.8 *v* 77.5). Despite this, at follow-up at 12 months, according to intention to treat analysis (difference 3.9, −0.2 to 8.0; P=0.064) and per protocol analysis (6.3, 2.4 to 10.2; P=0.002) the REFLUX score favoured the preference surgical group. For participants in the preference group, other quality of life scores also tended to favour the surgical group.

WHAT IS ALREADY KNOWN ON THIS TOPIC

Many people require regular proton pump inhibitors to control symptoms of gastro-oesophageal reflux disease, and there are concerns regarding the impact of long term use

Laparoscopic fundoplication can relieve symptoms of reflux, but side effects of surgery might be worse than symptoms associated with the best medical management

WHAT THIS STUDY ADDS

At one year follow-up reflux specific symptoms and general quality of life were better in patients who underwent laparoscopic fundoplication compared with those receiving optimised medical treatment

DISCUSSION

Principal findings

In patients with GORD, laparoscopic fundoplication results in better symptom relief and improved quality of life compared with optimised medical treatment. The worse the symptoms at entry the larger were the improvements after surgery. Similar differences were seen in most of the other measures of health status. There was some evidence of a narrowing of the differences between three and 12 months. There were small improvements in scores in the medical groups; this might reflect their having specialist review at trial entry to optimise their drug treatment.

The results in the preference groups were consistent with the randomised comparison. None of the three participants who died had surgery and complications were uncommon; but confidence intervals around estimated frequencies were wide despite inclusion of all surgical participants, leaving important uncertainty about the magnitude of surgical risk.

Strengths and weaknesses

We used a pragmatic trial design, with many patients, centres and experienced surgeons, thus allowing the results to be interpreted within a "real life" NHS context. The addition of the preference groups gives an indication of probable behaviour if surgery were to become more freely available.

We explored the impact of a third of those randomised to surgery not having fundoplication: firstly, through per protocol analyses limited to those randomised who received their allocated management, and, secondly, through an adjusted approach in an attempt to circumvent the probable selection bias of per protocol analyses. In the event, these two approaches gave similar results.

Our study was limited to patients who were on long term acid suppression with proton pump inhibitors, who had symptoms that were reasonably controlled, and who were clinically suitable for either policy; it is to these sorts of patients with GORD that the results are generalisable. Pressure on surgical services in the NHS meant the average time between trial entry and surgery was eight to nine months, an important factor in some

patients' eventual decision to choose not to have surgery after all. About a third of those who did not have fundoplication after allocation to surgery were refused surgery for clinical reasons.

Comparison with other studies

We identified two other randomised trials comparing laparoscopic Nissen fundoplication with continued medical management.[6 7] The results of the two trials were consistent with ours.

Conclusions

Laparoscopic fundoplication significantly increased health status at least to 12 months after surgery. The narrowing of differences in health status between three and 12 months could reflect a postoperative placebo effect or could indicate decreasing effectiveness of surgery over time. We have therefore instituted annual follow-up using similar questionnaires and plan to report long term effectiveness after five years of follow-up.

Details of the trial coordination team, the trial steering group, the data monitoring committee, and the REFLUX trial group responsible for recruitment in the clinical centres are on bmj.com.
Contributors: See bmj.com.
Funding: This study was funded by the NIHR Health Technology Assessment Programme (as part of project no. 97/10/99) and the full project report is published in *Health Technology Assessment* 2008;12:1/181. The Health Services Research Unit is funded by the Chief Scientist Office of the Scottish Government Health Directorates. The funders of this study (NCCHTA), other than the initial peer review process before funding and six monthly progress reviews, did not have any involvement in study design; in the collection, analysis, and interpretation of data; in the writing of the report; nor in the decision to submit the paper for publication. The views and opinions expressed in this paper are those of the authors and do not necessarily reflect those of the Department of Health or the funders that provide institutional support for the authors.
Competing interests: None declared.
Ethical approval: This study was approved by the Scottish multicentre research ethics committee and the appropriate local research ethics committees. All participants gave informed consent.
Provenance and peer review: Not commissioned; externally peer reviewed.

1 Leufkens H, Claessens A, Heerdink E, Van Eijk J, Lamers CBHW. A prospective follow-up study of 5669 users of lansoprazole in daily practice. *Aliment Pharmacol Ther* 1997;11:887-97.
2 Laine L, Ahnen D, McClain C, Solcia E, Walsh JH. Potential gastrointestinal effects of long-term acid suppression with proton pump inhibitors. *Aliment Pharmacol Ther* 2000;14:651-68.
3 Zacharoulis D, O'Boyle CJ, Sedman PC, Brough WA, Royston CMS. Laparoscopic fundoplication: a 10-year learning curve. *Surg Endosc* 2006;20:1662-70.
4 Torgerson D, Sibbald B. Understanding controlled trials: What is a patient preference trial? *BMJ* 1998;316:360.
5 Macran S, Wileman S, Barton G, Russell I. The development of a new measure of quality of life in the management of gastro-oesophageal reflux disease: the Reflux questionnaire. *Qual Life Res* 2007;16:331-43.
6 Mahon D, Rhodes M, Decadt B, Hindmarsh A, Lowndes R, Beckingham I, et al. Randomized clinical trial of laparoscopic Nissen fundoplication compared with proton-pump inhibitors for treatment of chronic gastro-oesophageal reflux. *Br J Surg* 2005;92:695-9.
7 Anvari M, Allen C, Marshall J, Armstrong D, Goeree R, Ungar W, et al. A randomized controlled trial of laparoscopic Nissen fundoplication versus proton pump inhibitors for treatment of patients with chronic gastroesophageal reflux disease: one-year follow-up. *Surg Innov* 2006;13:238-49.

Accepted: 24 September 2008

Commentary: Randomised trials of surgical and non-surgical treatment: a role model for the future

Jane M Blazeby,[1, 2] C Paul Barham,[2] Jenny L Donovan[1]

[1]Department of Social Medicine, University of Bristol, Bristol BS9 2PS
[2]Division of Surgery, Head and Neck, University Hospitals Bristol NHS Foundation Trust, Bristol BS2 8HW
Correspondence to: J M Blazeby
j.m.blazeby@bris.ac.uk

Cite this as: *BMJ* 2008;337:a2747
doi:10.1136/bmj.a2747

Symptoms of acid reflux and heartburn are increasingly common, and doctors need to understand the benefits and risks of surgical and medical treatment from the patients' perspective when offering treatment. Obtaining reliable information from randomised controlled trials to fully inform patients is difficult, however, because of the many recognised obstacles to undertaking such trials of diverse treatments, including difficulties with recruitment and standardising treatments and the inability to provide blinding. In the linked study, Grant et al carried out a multicentre pragmatic randomised trial of minimal access surgery (laparoscopic fundoplication) compared with optimised medical management (standard proton pump inhibition) for chronic gastro-oesophageal reflux disease (GORD).[1] The results were clear: up to 12 months after starting treatment, surgery was more effective than medical treatment at relieving symptoms related to reflux and improving general health.

This is an excellent model of how to design and conduct a robust and pragmatic randomised trial within the complexity of the NHS.[1] For example, the comprehensive cohort design—including eligible participants who chose their treatment as well as those randomised—ensured a large and wide ranging group of participants and analyses were performed by intention to treat, per protocol, and preference groups. To facilitate participation of clinicians and centres, clinical judgment was permitted in relation to inclusion and exclusion criteria and the type of surgical procedure or best medical management. The design was flexible, the sample size large enough to cope with variable organisational delays in surgery, and patients could decide to defer or decline their allocated or chosen treatment. Outcome was assessed with a validated patient reported outcome measure of symptoms and health, rather than outcomes measured and interpreted by an observer.[2 3] All these factors ensure that the findings are relevant to patients and clinicians by reflecting real world experiences of clinical practice.

By 12 months, patients randomised to surgery reported better outcomes than patients randomised to

medical management, and in all analyses patients undergoing fundoplication reported fewer symptoms of general discomfort, wind, nausea, and vomiting and less limitation in activity than the medical group. Dysphagia scores were similar in both groups. The use of antireflux medication after 12 months was 38% in those randomised to surgery (14% among those who had fundoplication) compared with 90% of medical participants.

Although these robust results support the use of laparoscopic surgery for the control of acid reflux in the short term, there are three further issues that might affect whether patients opt for surgery in future. The study followed up participants for 12 months, but longer term outcomes will be needed. Whether surgery is cost effective will be critical to providers of health care. Surgical complications were rare, but wide confidence intervals mean uncertainty about the true magnitude of risk, and other studies have shown that, when experienced, complications from surgery can be severe. Even with these robust results, patients and clinicians will still need to weigh up the clear symptomatic benefit likely to result from fundoplication with the (albeit rare) risk of complication and unknown longer term outcome.

Competing interests: JMB is an upper GI surgeon who does not undertake laparoscopic fundoplication. CPB is an upper GI surgeon who regularly undertakes laparoscopic fundoplication. Both authors have gastro-oesophageal reflux but neither has had surgery. CPB has received payments from Ethicon Endo-Surgery for organising educational training sessions in minimally invasive oesophagectomy.
Provenance and peer review: Commissioned; not externally peer reviewed.

1 Grant AM, Wileman SM, Ramsay CR, Mowat A, Krukowski ZH, Heading RC, et al.Minimal access surgery compared with medical management for chronic gastro-oesophageal reflux disease: UK collaborative randomised trial. *BMJ* 2008;337:a2664.
2 US Food and Drug Administration. *Guidance for industry. Patient reported outcome measures: use in medical product development to support labelling claims*. 2006. www.fda.gov/cber/gdlns/prolbl.htm.
3 Macran S, Wileman S, Barton G, Russell I. The development of a new measure of quality of life in the management of gastro-oesophageal reflux disease: the Reflux questionnaire. *Qual Life Res* 2007;16:331-43.

Patients' preferences within randomised trials: systematic review and patient level meta-analysis

Preference Collaborative Review Group

EDITORIAL by McPherson

Correspondence to: H Tilbrook, Research Fellow, York Trials Unit, Department of Health Sciences, University of York, York YO10 5DD
het2@york.ac.uk

Cite this as: *BMJ* 2008;337:a1864
doi:10.1136/bmj.a1864

ABSTRACT

Objective To systematically review fully randomised patient preference trials and to explore the impact of preferences on attrition and outcome by meta-analysis of patient level data.

Data sources Citation search using Science Citation Index and Google Scholar and search of the main electronic databases (Medline, CINAHL, Embase, and AMED) with a combination of key words.

Study selection Fully randomised patient preference trials that compared treatments for any clinical condition were included. Other types of preference trials and cross over trials were excluded. Other inclusion criteria: participants aged 16 years and over; primary, self-reported outcomes measured on a continuous numerical scale. From 167 studies identified and screened, 17 were identified as fully randomised patient preference trials.

Data synthesis Of the 17 trials identified, 11 authors provided raw data for the meta-analysis. Data collected were baseline and follow-up data for the main outcome, randomised allocation data, preference data, and demographic data. Baseline and first post-intervention follow-up data for the main outcome were standardised. To improve homogeneity, data for only the eight musculoskeletal trials (n=1594) were combined. To estimate the effects of preferences on outcomes and attrition, three groups were compared: patients who had a preference and were randomly allocated to their preferred treatment; patients who had a preference and were randomly allocated to the treatment they did not prefer; and patients who had no preference.

Results Patients who were randomised to their preferred treatment had a standardised effect size greater than that of those who were indifferent to the treatment assignment (effect size 0.162, 95% confidence interval 0.011 to 0.314; P=0.04). Participants who received their preferred treatment also did better than participants who did not receive their preferred treatment (effect size 0.152, −0.035 to 0.339), although this was not statistically significant (P=0.11). Participants allocated to their undesired treatment had outcomes that were no different from those who were indifferent. Participants who were allocated to their undesired treatment were less likely to be lost to first follow-up compared with indifferent participants (odds ratio 1.70, 1.076 to 2.693; P=0.02). No difference was found in attrition between patients allocated to their preference and those who were indifferent.

Conclusions Preferences among patients in musculoskeletal trials are associated with treatment effects. In open randomised trials, preferences should be ascertained before randomisation.

INTRODUCTION

Although random allocation in a controlled trial is intended to evenly distribute characteristics of participants that may affect outcome and remove selection bias, it may not deal with other potential biases. One of these is patients' preferences. If patients with

This article is an abridged version of a paper that was published on bmj.com. Cite this article as: *BMJ* 2008;337:a1864

preferences consent to be randomised then some patients will get their preferred treatment and others will not. Those who receive their preferred treatment might be better motivated and comply better with the treatment programmes and report better outcomes.[1] Patients who do not receive their preferred treatment may experience "resentful demoralisation,"[2] may be less motivated, may not comply with the treatment programme, may not report accurately during follow-up, and may even drop out of the trial and thereby introduce bias that affects the internal validity of the trial.[1] Effects of preference are likely to be more apparent in unblinded trials in which the outcome measure is subjective and self reported by the participant. A "therapeutic effect" of patients' preferences can also occur.[3] These are the psychological effects that influence outcomes and are similar to the placebo effect.[3]

One approach to dealing with patients' preferences is the partially randomised preference design.[1 4] Patients without strong preferences for a treatment are randomised, and those with strong preferences are given their choice, resulting in a four armed trial.[4] Such a design enables comparisons between patients with and without a preference and an exploration of patients' characteristics associated with preference. An important disadvantage of this design is that the outcome may be affected by uncontrolled confounders in the non-randomised groups,[5] which may bias the results.[6] Whether the randomised group should be compared with the non-randomised group is therefore debatable.[2 5]

An alternative approach is a standard randomised controlled trial in which, after participants have given consent in the usual way and before randomisation, patients' preferences are recorded: this design would be a fully randomised preference trial.[2] These preferences are then taken into account in the analysis of the trial.

King and colleagues systematically reviewed the effects of participants' preferences in randomised controlled trials.[1 7] They investigated the effects of preferences on outcomes and recruitment to trials and the influence of preferences on attrition. Their review identified 34 randomised controlled trials, but only two were fully randomised patient preference trials.[8 w1] The review, which used evidence from partially randomised preference trials, found no evidence that preferences influenced attrition. The trials included in the review were very heterogeneous, and the authors reported that they were therefore unable to "reach definitive conclusions for any particular clinical field."[1]

Here we report the results of a systematic review and an individual patient data analysis of fully randomised patient preference trials. The aim was to assess whether preferences had an interaction with effectiveness of treatment and whether attrition from the study was different between patients who received their preferred treatment, those who did not, and those who had no preference.

METHODS

We built on the extensive systematic review by King and colleagues (2005). To find additional papers we did a citation search using Science Citation Index and Google Scholar and using Torgerson et al (1996) and Torgerson and Sibbald (1998),[2 9] which were the first papers to describe the fully randomised patient preference trial. We also made searches in Medline, CINAHL, Embase, and AMED and included trials that were known to us through personal knowledge.

We included only fully randomised preference trials.[2] Other inclusion criteria were participants 16 years and over and primary self reported outcomes measured on a continuous scale. Additionally, we chose to focus on only one self reported primary outcome and exclude secondary outcome measures to avoid type I error (see bmj.com).

We approached the authors of the identified trials to release their data for the study. Because the studies used different outcome measures, we calculated standardised baseline and outcome scores for all patients by dividing their scores by the baseline standard deviation for the sample population for that trial. We used the first follow-up point after treatment had finished, as this would tend to minimise loss of data and be most likely to show an effect of preference if one existed.

Analysis

To analyse if patients' preferences influence outcomes, we used multiple regression with dummy variables for categorical predictors. The dependent variable was the standardised first follow-up score, and the predicting variable was patient's preference for treatment. We adjusted the analysis for baseline scores and categorical variables of trial and treatment allocation, which adjusted for treatment effects of individual interventions. In the regression analyses, we treated study as a fixed effect. For calculating the odds ratios for response to first follow-up questionnaires, we used logistic regression analysis. The dependent variable was missing data at first follow-up, and in the analysis we adjusted for age, sex, trial, and severity of condition at baseline. We compared the outcomes of three preference groups: patients with a preference who got what they wanted, those with a preference who did not get what they wanted, and those who had no preference (that is, they were indifferent). We first compared the two preference groups with the indifferent group; then we compared patients with a preference who were randomised to their desired treatment with those who were randomised to their undesired treatment and the indifferent group.

RESULTS

Our search identified 167 possibly relevant studies, of which 17 met the inclusion criteria. We obtained patient level data for 11 studies (see bmj.com).[w1-w12]

The proportion of patients who had a preference in the 11 trials ranged from 16% for a trial of a solution finding approach to back pain to 85% for a trial comparing general practitioner's care with hospital

care for non-urgent skin problems. The median preference rate was 56% (interquartile range 43-63%). The data show, therefore, that for this sample of fully randomised preference trials, most trial participants were not indifferent to the treatment to which they were going to be allocated.

Of the 11 studies with patient level data, eight (n=1594) were interventions for musculoskeletal conditions. We restricted our pooled analysis to these studies; consequently, clinical heterogeneity would be reduced. We found no evidence of statistical heterogeneity between studies (χ^2=6.02, df=8; P=0.64).

We found no differences in standardised baseline scores (mean 0.261 for preference group; 0.267 for indifferent group; 95% confidence interval for mean difference −0.068 to 0.177; P=0.38) but a difference in age (mean 48.57 for preference group; 46.99 for indifferent groups; 0.21 to 2.96; P=0.024). Women were more likely to have a preference, although this was not significantly different (46.3% (n=439) for women; 42.3% (n=271) for men; −1% to 9%; P=0.12).

After adjustment for baseline score, trial, and treatment allocation, in terms of treatment effect overall, patients who received their preference had significantly greater improvements than did those who were indifferent (mean effect size 0.162, 95% confidence interval 0.011 to 0.314; P=0.036). In contrast, contrary to expectations, those patients with a preference who did not receive their preferred treatment had only a slight, non-significant difference favouring an improvement relative to the indifferent group (effect size 0.011, −0.142 to 0.164; P=0.89). When we compared patients randomised to their preferred treatment with those randomised to their unpreferred treatment, we found an effect size of 0.152 (−0.035 to 0.339), which was not statistically significant (P=0.11). The total number of participants included in the analysis to compare the effects of preferences on outcomes was 1398.

After adjustment for severity of condition at baseline, age, and sex, those who did not receive their preference were more likely to return their questionnaire at first follow-up than were those who were indifferent (odds ratio 1.70, 95% confidence interval 1.08 to 2.69; P=0.02); those who did receive their preference also had an increased response rate, but this difference was not significantly different from the indifferent group (odds ratio 1.26, 0.82 to 1.94; P=0.29). The comparison of those who did not get their preference with those who did yielded an odds ratio of 1.35, but this was not significant (95% confidence interval 0.78 to 2.33; P=0.29). The total number of participants included in the analysis to compare the effects of preferences on attrition was 1583.

DISCUSSION

In contrast with an earlier review,[1] we have found some evidence to suggest that preferences can modify treatment outcomes, although these effects were not all in the expected direction. In terms of increased study attrition, which is widely hypothesised to be affected by preference, we found the converse of what was expected: participants with a preference who are randomised to the opposite treatment are actually more likely to return follow-up data than those who were indifferent.

WHAT IS ALREADY KNOWN ON THIS TOPIC

The effect of patients' preferences on treatment outcomes in randomised controlled trials is uncertain

Alternative, partially randomised, trials have been designed to overcome the potential problem of patients' preferences in trials

An existing systematic review found no evidence that preferences influenced attrition and some evidence of an effect on outcomes in partially randomised trials

WHAT THIS STUDY ADDS

Patients' preferences do affect treatment outcomes in randomised controlled trials in musculoskeletal medicine

"Resentful demoralisation" did not occur in participants who did not get their preferred treatment

Use of the standard randomised controlled trial, which collects preference data before randomisation, provides the opportunity to take preference effects into account in the analysis

Strengths and limitations

By pooling the data from the eight musculoskeletal trials, we obtained a relatively large sample size. The design of this study—in which we made comparisons between patients who received their preference, patients who did not receive their preference, and those with no preference—is, as far as we can determine, novel for detecting effects of preference. Also, in most of the trials included in the analysis, the questionnaires had been completed "alone" by the participants and were therefore unlikely to have been influenced by any treatment preference held by health professionals.

Our study does have some limitations. Firstly, we restricted our meta-analysis to musculoskeletal trials, so our results may not be applicable to other areas of health care. Secondly, within each of the individual component trials participants with a very strong preference for usual care will tend not to be recruited into the studies. However, several studies did recruit participants who had a preference for treatments that were available outside of the trial setting. Thirdly, we interpreted missing data at the first follow-up after the intervention as attrition. Fourthly, eight of the 17 studies were identified through the personal knowledge of one of the authors. Four of these studies were unpublished, suggesting that other similar trials exist that we did not identify. Finally, confounding could be present.

Implications

In terms of treatment effects, we did see an increased treatment effect size among participants who were

randomised to their preferred treatment compared with those who were indifferent to the treatment allocation. This increased effect is not necessarily a bias—it might be a genuine effect in the sense that the treatment works better among patients who desire and receive their preferred treatment. It could, however, change the cost effectiveness of a treatment. A treatment may not be cost effective overall but could be among a subgroup of participants who prefer that treatment.[10] Consequently, we would conclude that in trials—particularly when a novel therapy is available only within the context of the study and outcomes are subjective—pre-randomised preferences should be identified and recorded so that they can be accounted for in any analysis. Our results provide no evidence that "resentful demoralisation" leads to a reduction in effect size if people are allocated to their undesired treatment.

Stratifying by preferences would help the power of a trial if either the trial was small or the proportion of patients with a preference was small. In these cases, trialists might consider using preference as a stratification variable.

Conclusions

This review shows that treatment preferences affect outcomes in a sample of musculoskeletal trials but that they are not detrimental to attrition rates. The fully randomised preference trial seems to be more widely used now, which would enable further work to determine the treatment effects of patients' preferences in different clinical conditions.

This paper was a joint collaboration with the members of the Preference Collaborative Group: Simon J Adamson, J Martin Bland, Elaine M Hay, Ruth E Johnson, Gareth T Jones, Henry Kitchener, Jennifer A Klaber Moffett, Gary J Macfarlane, Hugh MacPherson, Sionnadh McLean, Linsey Nelson, Chris Salisbury, Elaine Thomas, Helen E Tilbrook, and David J Torgerson.
Contributors: See bmj.com.
Funding: None.
Competing interests: None declared.
Ethical approval: Not needed.
Provenance and peer review: Not commissioned; externally peer reviewed.

1 King M, Nazareth I, Lampe F, Bower P, Chandler M, Morou M, et al. Impact of participants and physician intervention preferences on randomised trials. a systematic review. *JAMA* 2005;293:1089-99.
2 Torgerson DJ, Klaber-Moffett J, Russell IT. Patient preferences in randomised trials: threat or opportunity? *J Health Serv Res Policy* 1996;1:194-7.
3 McPherson K, Britton AR, Wennberg JE. Are randomised trials controlled? Patient preferences and unblind trials. *J Roy Soc Med* 1997;90:652-6.
4 Brewin CR, Bradley C. Patient preferences and randomised clinical trials. *BMJ* 1989;299:313-5.
5 Silverman WA, Altman DG. Patients' preferences and randomised trials. *Lancet* 1996;347:171-4.
6 Halpern SD. Evaluating preference effects in partially unblinded, randomized clinical trials. *J Clin Epidemiol* 2003;56:109-15.
7 King M, Nazareth I, Lampe F, Bower P, Chandler M, Morou M, et al. Conceptual framework and systematic review of the effects of participants' and professionals' preferences in randomised controlled trials. *Health Technol Assess* 2005;9(35):1-186, iii-iv.
8 Hardy GE, Barkham M, Shapiro D, Reynolds S, Rees A. Credibility and outcome of cognitive-behavioural and psychodynamic-interpersonal psychotherapy. *Br J Clin Psychol* 1995;34:555-69.
9 Torgerson D, Sibbald B. Understanding controlled trials: what is a patient preference trial? *BMJ* 1998;316:360.
10 Manca A, Epstein DM, Torgerson DJ, Klaber-Moffett JA, Coulton S, Farrin AJ, et al. Randomized trial of a brief physiotherapy intervention compared with usual physiotherapy for neck pain patients: cost-effectiveness analysis. *Int J Technol Assess Health Care* 2006;22:67-75.

Accepted: 27 August 2008

Exercise on prescription for women aged 40-74 recruited through primary care: two year randomised controlled trial

Beverley A Lawton,[1] Sally B Rose,[1] C Raina Elley,[2] Anthony C Dowell,[3] Anna Fenton,[4] Simon A Moyes[2]

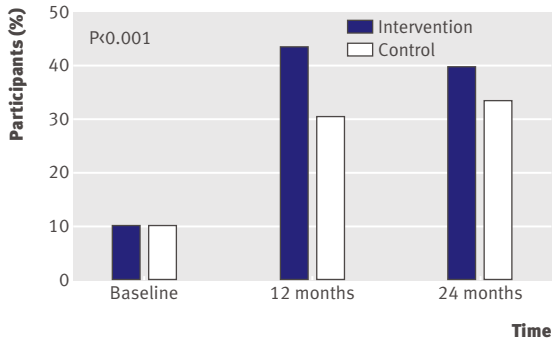
EDITORIAL by Iliffe et al

[1]Women's Health Research Centre, Department of Primary Health Care and General Practice, PO Box 7343, University of Otago, Wellington, New Zealand
[2]Department of General Practice and Primary Health Care, School of Population Health, University of Auckland, Private Bag 92019, Auckland, New Zealand
[3]Department of Primary Health Care and General Practice, PO Box 7343, University of Otago, Wellington, New Zealand
[4]Christchurch Women's Hospital, Christchurch, New Zealand
Correspondence to: B Lawton Bev.Lawton@otago.ac.nz

Cite this as: *BMJ* 2008;337:a2509
doi:10.1136/bmj.a2509

ABSTRACT

Objective To assess the effectiveness of a primary care based programme of exercise on prescription among relatively inactive women over a two year period.
Design Randomised controlled trial.
Setting 17 primary care practices in Wellington, New Zealand
Participants 1089 women aged 40-74 not undertaking 30 minutes of moderate intensity physical activity on at least five days of the week
Intervention Brief physical activity intervention led by nurse with six month follow-up visit and monthly telephone support over nine months.
Main outcome measure Physical activity assessed at baseline and 12 and 24 months. Secondary outcomes were quality of life (SF-36), weight, waist circumference, blood pressure, concentrations of fasting serum lipids, glycated haemoglobin (HbA_{1c}), glucose, insulin, and physical fitness.
Results Mean age was 58.9 (SD 7) years. Trial retention rates were 93% and 89% at 12 and 24 months, respectively. At baseline, 10% of intervention participants and 11% of control participants were achieving 150 minutes of at least moderate intensity physical activity a week. At 12 months rates increased to 43% and 30% and at 24 months to 39.3% and 32.8% (P<0.001), respectively. SF-36 physical functioning (P=0.03) and mental health (P<0.05) scores improved more in intervention compared with control participants, but role physical scores were significantly lower (P<0.01). There were no significant differences in clinical outcomes. More falls (P<0.001) and injuries (P=0.03) were recorded in the intervention group.
Conclusions This programme of exercise on prescription increased physical activity and quality of life over two years, although falls and injuries also increased. This finding supports the use of exercise on prescription programmes as part of population strategies to reduce physical inactivity.



Proportion of participants achieving at least 150 minutes of moderate intensity physical activity at baseline and 12 and 24 months

Trial registration Australian New Zealand Clinical Trials Registry (ANZCTR) ANZCTRN012605000490673.

INTRODUCTION

Physical inactivity is a major contributor to chronic disease, including ischaemic heart disease, stroke, breast cancer, colon cancer, and type 1 diabetes.[1-4] "Exercise on prescription" interventions that involve a health professional's written advice to a patient to be more physically active have been used with variable success.[5 6] We used a randomised controlled trial design to test the effectiveness and sustainability of a primary care based programme of exercise on prescription over two years.

METHODS

Participants

Eligibility criteria included women aged 40-74 who were physically inactive, as determined by a one question screening tool: "As a rule, do you do at least half an hour of moderate or vigorous exercise (such as walking or a sport) on five or more days of the week?"[7 8]

Women were excluded if they had a medical condition that might be adversely affected by increasing their physical activity, as determined by the physical activity readiness questionnaire (PAR-Q)[9] and subsequent assessment by their own general practitioner.

We recruited from two sources. The first source was an existing cohort of 50-74 year old women recruited by invitation letter from their general practitioner to a previous observational study of postmenopausal women between 1999 and 2002 from 10 primary care practices in Wellington.[10] The remainder of the participants were 50-70 year old women (40-60 years for Maori and Pacific women) recruited from 13 primary care practices in 2004-5, including two Maori health clinics. (See bmj.com for further details.)

Outcome measures

Our primary outcome measure was the proportion of those achieving the recommended 150 minutes of at least moderate intensity physical activity, as assessed by the long form of the physical activity questionnaire.[11-13] See bmj.com for a description of secondary outcomes, randomisation, and blinding.

Intervention

The intervention we assessed was built on an existing primary care programme, the green prescription, in which the general practitioner or practice nurse briefly counsels (7-13 minutes) patients using motivational interviewing techniques to increase physical activity among those who were physically inactive. The details of the exercise advice are written on a "green script," which is given to the patient and faxed to a community based exercise facilitator who provides telephone support over a three month period, assisting with choice of activity, goal setting, and ways to overcome personal barriers to physical activity.[14 15] In our study a primary care nurse delivered the green prescription and follow-up was extended to include telephone calls over a nine month period (average of five calls, each lasting 15 minutes) with an added 30 minute visit with the primary care nurse at six months. The recommended goal was moderate intensity physical activity such as brisk walking, with a goal of achieving 30 minutes five days a week. The nurse noted clinical details including weight, height, waist circumference, smoking status, and any relevant medical conditions on the faxed script.

During the visit at six months the nurse established whether the participant had increased her physical activity to the target level, provided encouragement and motivation, and measured blood pressure, weight, and waist circumference. Tools to assist with choosing appropriate types of activities and motivational aids, such as fridge magnets and activity record charts, were also offered. Further details about the intervention are available elsewhere.[16] Control participants received usual care from their primary care practice.

Sample size calculation

We needed a sample size of 880 to detect a minimum difference between the groups of 7% change in the proportion of women reaching the target level of physical activity, allowing for a 10% attrition rate (α=0.05, 80% power).[14 17] (See bmj.com.)

Statistical analyses

We carried out an intention to treat analysis. For missing data at follow-up assessments, we assumed no change from baseline. Final analyses were undertaken with regression models, adjusted for repeated measures and baseline values.

RESULTS

We recruited participants from 17 primary healthcare practices in Wellington and in 2004-5 enrolled 1089 less active women. Characteristics of the study participants were balanced at baseline (see bmj.com). Trial retention rates were 93% and 89% at 12 and 24 months, respectively.

This article is an abridged version of a paper that was published on bmj.com. Cite this article as: *BMJ* 2008;337:a2509

Both groups increased their physical activity over the two years. Mean physical activity levels, however, were higher (P=0.01) and a greater proportion reached the target of physical activity in the intervention group compared with the control group at 12 months (233 (43%) *v* 165 (30%), (P<0.001), with levels declining but still significantly different at two years (214 (39%) *v* 179 (33%), figure and table). SF-36 physical functioning (P=0.03) and mental health (P<0.05) subscores were also significantly better in the intervention group at 12 and 24 months, although role physical scores were lower in the intervention group (P<0.01) (see bmj.com). There were no significant differences between groups in any of the secondary clinical outcomes (table).

Although a large proportion of control participants increased their physical activity during the trial, only 2.4% (11/480 women contacted) recalled having received a green prescription from their doctor or nurse during the data collection period. (See bmj.com for full results.)

DISCUSSION

Exercise on prescription can increase physical activity and improve some variables of quality of life over two years among physically inactive women recruited through primary care compared with usual care. In our study, however, the increased levels of physical activity did not produce significant improvements in clinical or biochemical variables, and there were increases in falls and injuries and a reduction in the SF-36 role physical score.

Comparison with other studies

The two year follow-up makes this study one of the first primary prevention physical activity trials to show a significant effect of physical activity counselling over two years. The activity counselling trial compared two behavioural counselling approaches of different intensities (intervention arms) with brief advice from a physician and written materials (control arm). The intensive interventions produced more improvement in cardiorespiratory fitness in women but not men over two years, but there was no change in self reported physical activity in men or women.[18]

The recently reported "ProActive" randomised controlled trial in primary care involved adults with a family history of diabetes who were randomised to a one year behaviour change programme (delivered by telephone or face to face) or to a control group (given an advice leaflet on physical activity).[19] At one year, daytime physical activity had improved in both groups, but there were no significant differences between them.

The incremental increase in physical activity in our study was more than was found with the briefer green prescription intervention over one year (12.5% *v* 10% at 12 months) and similar improvements were observed in quality of life variables.[14]

Limitations

We were not able to blind participants to the intervention. Participants in both groups showed a high uptake of physical activity. This positive effect in the control group might have reflected a trend in the population, participation in a study about physical activity, or the fact that those who agreed to take part in a trial were more motivated to change anyway than those who declined. Furthermore, the time spent with the research nurse at each assessment might have acted as an intervention in itself. Contamination is unlikely as few participants in the control group reported having received a green prescription (the basic version of the intervention used) during the study.

Although participants were recruited through primary care, their participation was by special invitation and the delivery of the intervention was not part of routine care. Even so, the basic green prescription is already part of routine care in New Zealand. The focus on older women, the self selected nature of participants, and the overall participation rate of 19.5% (1089/5913 invited minus 317 returned to sender) might further

Primary and secondary clinical outcome measures and adverse events in intervention and control groups at baseline and 12 and 24 months. Values are means (SD) unless stated otherwise

Outcome measures	Intervention (n=544)	Control (n=545)	P value*
Primary			
No (%) completing at least 150 minutes physical activity/week:			
Baseline	56 (10)	62 (11)	
12 months	233 (43)	165 (30)	<0.001
24 months	214 (39)	179 (33)	
Secondary			
Systolic blood pressure (mm Hg):			
Baseline	122.8 (0.7)	123.4 (0.8)	
12 months	120.6 (0.7)	121.9 (0.7)	0.50
24 months	119.1 (0.7)	119.5 (0.7)	
Diastolic blood pressure (mm Hg):			
Baseline	73.8 (0.4)	74.7 (0.4)	
12 months	71.5 (0.4)	72.4 (0.4)	0.96
24 months	71.6 (0.4)	71.7 (0.4)	
Weight (kg):			
Baseline	73.2 (0.6)	72.7 (0.6)	
12 months	72.6 (0.6)	72.7 (0.6)	0.60
24 months	72.6 (0.6)	72.5 (0.6)	
Adverse events			
No (%) of falls:			
Baseline	138 (25)	155 (29)	
12 months	158 (32)	127 (25)	<0.001
24 months	179 (37)	143 (29)	
No (%) of injuries:			
Baseline	77 (14)	103 (19)	
12 months	91 (18)	86 (17)	0.03
24 months	92 (19)	66 (14)	

*Analyses took into account repeated measures and adjusted for baseline values. Data that were not normally distributed were log transformed.

WHAT IS ALREADY KNOWN ON THIS TOPIC

Achieving 30 minutes of moderate intensity physical activity on five or more days a week is associated with a substantial reduction in the risk of many chronic diseases and improves quality of life

Secondary prevention studies have shown the effectiveness of programmes for reducing inactivity and chronic disease

Few primary prevention programmes have produced sustainable increases in physical activity

WHAT THIS STUDY ADDS

This programme of exercise on prescription increased physical activity and improved some variables of quality of life over two years

This finding supports the role of exercise on prescription programmes in reducing population levels of physical inactivity

There was a small increase in falls and injuries associated with the programme

limit the generalisability of results to younger women and to men. Of those assessed for eligibility, however, we excluded 47% as they were already physically active, which is in line with national data on physical activity levels.[20]

The absence of a significant difference in secondary clinical end points is not surprising and might be partly due to the increase in overall physical activity in the control group. Therefore, while this sample size was large enough to detect quite small differences in physical activity between the groups, it was not large enough to detect as significant any differences in clinical outcomes. There is, however, a well established relation between increasing physical activity and health benefit, so showing a small increase across a population has health benefit and can be cost effective considering the low cost of the intervention.[21]

The use of a self reported measure as the main outcome is a potential weakness but, when validated against an objective measure, the physical activity questionnaire performed well.[12,13] Objective measures of physical activity, such as activity monitors, would have added to the validity of these findings. Adverse events of falls and injuries were also self reported so were open to recall bias.

Conclusion

Programmes of exercise on prescription can produce sustained increases in physical activity among less active middle aged and older women over two years.

We thank the women who participated in this trial and the primary care practices from which participants were recruited.

Contributors: See bmj.com.

Funding: This work was supported by the National Heart Foundation of New Zealand (grant number 1091 and grant-in-aid numbers 1091 and 1222), the Lottery Health Research Grants Board, the Hutt Valley District Health Board, Sport and Recreation New Zealand, and the Maori Health Directorate (Ministry of Health).

Ethical approval: This work was approved by the Central Region Ethics Committee (formerly the Wellington Ethics Committee) in September 2004 (WGT/04/08/061). Participants gave written informed consent to participate in this study.

Competing interests: None declared.

Provenance and peer review: Not commissioned; externally peer reviewed.

1 Chief Medical Officer. *At least five a week: evidence on the impact of physical activity and its relationship to health: a report from the chief medical officer*. London: Department of Health, 2004. www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_4080994

2 Schulze MB, Hu FB. Primary prevention of diabetes: what can be done and how much can be prevented? *Ann Rev Public Health* 2005;26:445-67.

3 Hu FB, Stampfer MJ, Colditz GA, Ascherio A, Rexrode KM, Willett WC, et al. Physical activity and risk of stroke in women. *JAMA* 2000;283:2961-7.

4 Colditz GA, Samplin-Salgado M, Ryan CT, Dart H, Fisher L, Tokuda A, et al. Harvard report on cancer prevention. Vol 5. Fulfilling the potential for cancer prevention: policy approaches. *Cancer Causes Control* 2002;13:199-212.

5 Sorensen JB, Skovgaard T, Puggaard L. Exercise on prescription in general practice: a systematic review. *Scand J Prim Health Care* 2006;24:69-74.

6 NICE. *A rapid review of the effectiveness of exercise referral schemes to promote physical activity in adults*. London: Public Health Collaborating Centre-Physical Activity, 2006. www.nice.org.uk/nicemedia/pdf/PA-Exercise_Referral_Review_Final_May_2006.pdf

7 Elley CR, Kerse NM, Arroll B. Why target sedentary adults in primary health care? Baseline results from the Waikato Heart, Health, and Activity Study. *Prev Med* 2003;37:342-8.

8 Rose SB, Elley CR, Lawton BA, Dowell AC. A single question reliably identifies physically inactive women in primary care. *N Z Med J* 2008;121:U2897.

9 Canadian Society for Exercise Physiology. *Physical activity readiness questionnaire (PAR-Q)*. 2002. www.csep.ca/main.cfm?cid=574&nid=5110

10 Rose S, Lawton B, Dowell A, Fenton A. Risk factors for type 2 diabetes in postmenopausal New Zealand women: a cross-sectional study. *N Z Med J* 2004;117:U1206.

11 Craig CL, Marshall AL, Sjostrom M, Bauman AE, Booth ML, Ainsworth BE, et al. International physical activity questionnaire: 12-country reliability and validity. *Med Sci Sports Exerc* 2003;35:1381-95.

12 McLean G, Tobias MI. *The New Zealand physical activity questionnaires: report on the validation and use of the NZPAQ-LF and NZPAQ-SF self-report physical activity survey instruments*. Wellington, New Zealand: Sport and Recreation New Zealand (SPARC), 2004. www.sparc.org.nz/research-policy/research/nzspas-97-01/nzpaq

13 Moy K, McFarlane K, Scragg R, Robinson S. *Validation of MOH-short and SPARC-long physical activity questionnaires: final report*. Auckland: University of Auckland and Sport and Recreation New Zealand, 2003.

14 Elley CR, Kerse N, Arroll B, Robinson E. Effectiveness of counselling patients on physical activity in general practice: cluster randomised controlled trial. *BMJ* 2003;326:793.

15 Sport and Recreation New Zealand. *The green prescription*. 1997www.sparc.org.nz/getting-active/green-prescription/overview.

16 Rose SB, Lawton BA, Elley CR, Dowell AC, Fenton AJ. The Women's Lifestyle Study, 2-year randomized controlled trial of physical activity counselling in primary health care: rationale and study design. *BMC Public Health* 2007;7:166.

17 Stevens W, Hillsdon M, Thorogood M, McArdle D. Cost-effectiveness of a primary care based physical activity intervention in 45-74 year old men and women: a randomised controlled trial. *Br J Sports Med* 1998;32:236-41.

18 Writing Group for the Activity Counselling Trial Research Group. Effects of physical activity counselling in primary care. *JAMA* 2001;286:677-87.

19 Kinmonth A-L, Wareham NJ, Hardeman W, Sutton S, Prevost AT, Fanshawe T, et al. Efficacy of a theory-based behavioural intervention to increase physical activity in an at-risk group in primary care (ProActive UK): a randomised trial. *Lancet* 2008;371:41-8.

20 Van Aalst I, Kazakov D, Mclean G. *SPARC facts. Results of the New Zealand sport and physical activity surveys (1997-2001)*. Wellington, NZ: Sport and Recreation New Zealand, 2003.

21 Dalziel K, Segal L, Elley CR. Cost utility analysis of physical activity counselling in general practice. *Aust N Z J Public Health* 2006;30:57-63.

Accepted: 29 September 2008

Renal screening in children after exposure to low dose melamine in Hong Kong: cross sectional study

Hugh S Lam,[1] Pak C Ng,[1] Winnie C W Chu,[2] William Wong,[1] Dorothy F Y Chan,[1] Stella S Ho,[2] Ka T Wong,[2] Anil T Ahuja,[2] Chi K Li[1]

[1]Department of Paediatrics, The Chinese University of Hong Kong
[2]Department of Diagnostic Radiology and Organ Imaging, The Chinese University of Hong Kong
Correspondence to: H S Lam, 6/F Clinical Sciences Building, Department of Paediatrics, Prince of Wales Hospital, Sha Tin, New Territories, Hong Kong
hshslam@cuhk.edu.hk

Cite this as: *BMJ* 2008;337:a2991
doi:10.1136/bmj.a2991

ABSTRACT

Objective To investigate the renal outcomes of children after exposure to low dose melamine in Hong Kong.

Design Cross sectional study.

Setting Special assessment centres, Hong Kong.

Participants 3170 children (1422 girls and 1748 boys) aged 12 years or less referred from territory-wide primary care clinics after daily consumption for one month or more of milk products tainted with melamine.

Main outcome measures Presence of renal stones and haematuria.

Results One child had a confirmed renal stone, seven were suspected of having melamine related renal deposits, and 208 (6.6%) were positive for blood in urine by reagent strip. A proportion of these children were followed up at the special assessment centre, but only 7.4% of those positive for blood on reagent strip were confirmed by microscopy, suggesting an overall estimated prevalence of less than 1% for microscopic haematuria.

Conclusions No severe adverse renal outcomes, such as acute renal failure or urinary tract obstruction, were detected in children after exposure to low dose melamine. Our results were similar to territory-wide findings in Hong Kong. Even including the seven children with suspected renal deposits, the prevalence of suspected melamine related abnormalities on ultrasonography was only 0.2%. None of these children required specific treatment. The prevalence of microscopic haematuria was probably overestimated by the reagent strip. These data suggest that large scale and urgent screening programmes may not be informative or cost effective for populations who have been exposed to low dose melamine.

INTRODUCTION

Since the first media reports on 11 September 2008 linking an outbreak of renal diseases among children on the mainland of China to consumption of milk products contaminated with melamine, an estimated 50 000 or more children have been affected.[1 2] More than 100 became seriously ill and at least four deaths have been attributed to melamine.[3]

Melamine is a triazine compound and is an essential component of materials such as flame retardants, glues, and plastics.[4 5] In the mainland, addition of this compound to milk intended for human consumption allowed diluted milk to appear to contain satisfactory amounts of protein and to pass food quality tests. Little is known about the adverse effects of melamine consumption in humans. The main adverse effects shown in studies on animals were kidney related, including renal calculi, renal tubular necrosis, melamine crystalluria, and haematuria.[4 6] On the mainland where young children were exposed to high doses of melamine, adverse renal effects, including renal stones and acute renal failure, have been reported.[4 7] Although milk products tainted with melamine were found to be commercially available outside the mainland—for example, the Hong Kong Special Administrative Region, these exported products had much lower concentrations of melamine. Melamine contaminated milk products available on the mainland could contain as much as 2563 mg/kg of melamine, whereas the most severely contaminated milk product manufactured on the mainland but commercially available in Hong Kong contained 68 mg/kg.[8]

In view of the wide range of food products found to be contaminated with melamine and the proximity of Hong Kong to the mainland, parents throughout the territory were anxious about the possible adverse health effects on their children of consuming tainted milk products. In the absence of good evidence to determine the risk of melamine related complications in children in Hong Kong, the government initiated a territory-wide screening programme for renal complications in children aged 12 years or less who had consumed the contaminated milk products. In view of the lack of published data on the consequences of exposure to low dose melamine—that is, doses below the tolerable daily intake recommended by the US Food and Drug Administration of 0.63 mg/kg/day,[4] we report the short term findings from our cohort.

METHODS

General outpatient clinics at primary care level throughout Hong Kong were designated as first level screening centres. Hong Kong residents could walk in for assessment without appointment. Children who were aged 12 years or less and had consumed melamine tainted milk products daily for one month or more were referred to special assessment centres for renal assessments. We defined melamine tainted milk products as those known to be contaminated with melamine according to test results posted by the Hong Kong government.[8] We recorded the brands of contaminated products, the daily intake, and the duration of continuous consumption. In view of the large degree of uncertainty and rapidly expanding list of confirmed tainted products during the early weeks of screening, the Hong Kong government refrained from setting minimum consumption volumes or estimated

daily intake thresholds for melamine below which the child would not be referred for renal assessment.

All children at our special assessment centre underwent renal ultrasonography and urinalysis. Ultrasound examinations were done by experienced sonographers and radiologists. Positive findings were cross checked by the chief consultant radiologist before being reported. Although the sonographers and radiologists were not purposely blinded to history of exposure, this information was not available to those who carried out the renal ultrasonography. Furthermore, with the large numbers of children for each scanning session, the sonographers or radiologists lacked the time to find out about consumption or to estimate the level of exposure to melamine. In practice therefore ultrasonography was essentially done blinded to exposure history.

Urine was analysed using Combur-Test strips (Roche Diagnostics, Mannheim Germany). In view of the large number of children screened daily, those with urine positive for blood on reagent strip testing were referred back to the local primary care clinics for further investigation and follow-up if no abnormalities were found on ultrasonography or for renal function. Consequently, with smaller numbers of children seen, urine microscopy was done on all those with urine positive for blood on reagent strips to further assess the haematuria. In addition, blood samples were taken for assessment of urea and creatinine levels in all but one of the children who had abnormalities on ultrasonography or urinalysis.

RESULTS

Between 25 September and 30 October 2008 the special assessment centre at the Prince of Wales Hospital assessed 3170 children (1422 girls and 1748 boys), mean age 6.4 (range 0.1 to 12.9) years, who had consumed milk products tainted with melamine. Milk intake had been on a regular basis for at least one month, from twice a week to daily. The amount of milk consumed was variable, with self reported daily volumes ranging from 250 ml to more than 1.5 litres. All reported regularly consuming at least one of the milk products that had been found to contain melamine.[8] The estimated melamine intake of the eight children with renal stones or deposits was between 0.01 mg/kg/day and 0.21 mg/kg/day, which did not exceed the tolerable daily intake of 0.63 mg/kg/day.[4]

A renal calculus (7 mm) without associated obstructive hydronephrosis was found in one child (0.03%). Seven children (0.2%) had small hyperechoic renal foci (<4 mm) at or close to the renal papillae. They all had a history of good health and only one child with small hyperechoic renal foci had urinary symptoms (dysuria and urinary frequency). The urinary calcium to creatinine ratio was normal in all children with renal stones or deposits. Incidental findings were identified in 17 children (0.5%), which were considered unrelated to melamine intake (table).

On reagent strip urinalysis, 115 girls (8.0%) and 93 boys (5.3%) were positive for blood, 34 girls (2.4%) and 25 boys (1.4%) were positive for protein, and 106 girls (7.6%) and two boys (0.1%) were positive for leucocytes. Of the 54 children with positive results for blood between 15 and 29 October, red blood cells in the urine could be confirmed by microscopy in only four. This would reduce the overall estimated prevalence of haematuria to less than 1%. One 3 year old boy and one 5 year old girl had low calculated creatinine clearance (54 ml/min/1.73 m^2 and 59 ml/min/1.73 m^2, respectively)[9]; however, no other abnormalities were found on further assessment. Only one of the eight children with suspected renal stones or deposits showed any abnormalities on urinalysis (blood ++++ by reagent strip), and all had normal plasma creatinine concentrations.

Results of renal ultrasonography and urinalysis in 3170 Chinese children who consumed melamine tainted milk products between 25 September and 30 October 2008

Abnormality	No with abnormality	Prevalence (%)
Renal calculus	1	0.03
Renal deposit	7	0.22
Other renal abnormalities*	17	0.54
Haematuria	208	6.56
Proteinuria	59	1.86
Leucocytes	108	3.40
Other abnormalities on urinalysis†	5	0.16

*Pelviectasia (n=4), tiny milk of calcium cysts (n=3), renal agenesis (n=2), polycystic renal disease (n=2), horseshoe kidney (n=1), transient dilation of ureter (n=1), ureterocele (n=1), small scarred kidney (n=1), ovarian dermoid (n=1), and ovarian cystadenoma (n=1).
†Other abnormalities on urinalysis: ketones (n=1), bilirubin (n=1), nitrites (n=1), and glucose (n=4).

DISCUSSION

In the absence of any relevant data on acute outcomes in children exposed to low dose melamine during the months before September 2008 in China, the government of the Hong Kong Special Administrative Region reacted promptly by screening large numbers of children for renal complications. Unlike children living on the mainland of China, who were exposed to much higher doses of melamine than children in Hong Kong, no serious adverse renal outcomes were detected in our screening programme. We speculate that the disparity in outcomes between children who were assessed in Hong Kong and those living on the mainland was the large difference in concentrations of melamine in the milk products consumed.

Comparison with other studies

The pattern of complications detected in the territory-wide screening in the Hong Kong Special Administrative Region was similar to that of our cohort. Of 17 667 children screened at all special assessment centres up to 5 November 2008, only 10 (0.06%) had renal stones (≥4 mm) detected. Echogenic foci closely related to the renal papillae are unusual findings on ultrasonography in asymptomatic children. In a report on 196 cases of microlithiasis in children, all but 13 presented with abdominal or genitourinary symptoms.[10] A large proportion also had urinary abnormalities, including haematuria (61%) and hypercalcuria (38%). In our children with renal deposits, only one had haematuria, but all had a normal urinary calcium:creatinine ratio.

WHAT IS ALREADY KNOWN ON THIS TOPIC

Exposure to high dose melamine was linked to an outbreak of severe renal diseases among children in China

Little is known about the renal outcomes of children after exposure to low dose melamine

WHAT THIS STUDY ADDS

Large numbers of children in Hong Kong were screened for acute renal complications within a short period

No severe adverse renal outcomes were detected

Urgent and large scale renal screening in children with a history of exposure to low dose melamine may not be necessary

These hyperechoic lesions were noted to have less acoustic shadowing than usual renal stones. The features of these suspected melamine related renal deposits on ultrasonography resemble gallbladder sludge rather than calcified stones.[11] Even including these seven children with possible renal deposits, only a small percentage (0.2%) of children exposed to relatively low dose melamine were affected. None of our children developed acute renal failure or urinary tract obstruction, or required treatment. However, the high prevalence of abnormalities in urine is unusual. It is possible that a large proportion was false positive, as suggested by the small number of children with red blood cells in urine confirmed by microscopy. As reagent strips can be more sensitive than microscopy,[12] however, some of these children may genuinely have had mild haematuria. Results from a screening programme in Japan showed that in 6197 school aged children, dipstick alone detected occult blood in 4.1% and protein in 2.1%.[13] The proportion of children with haematuria detected by reagent strip testing was slightly higher in our cohort (6.6%). Using a multilevel screening algorithm in a larger cohort (23 121 Japanese children), haematuria was detected in 0.7% and proteinuria in less than 0.01%.[14] The proportion of children with haematuria confirmed by microscopy in Hong Kong is similar to that in Japan.

Strengths and limitations of study

The catchment area of our centre includes the northern areas of Hong Kong closest to the border with the mainland. Our cohort therefore includes a large mobile population that travels often and regularly between the Hong Kong Special Administrative Region and the mainland. A large population of the children are born in Hong Kong but live across the border. Children seen at our centre would therefore be expected to be at highest risk of exposure to melamine tainted milk products within Hong Kong, but probably of lower risk compared with infants and children living on the mainland. To date our centre has screened the largest number of children in Hong Kong.

A limitation of our analyses is the absence of a reliable biomarker of melamine exposure. In view of the short plasma half life of melamine (three hours), blood concentrations are unreliable.[4] The concentration of melamine in urine is also an unhelpful marker as levels decrease rapidly after exposure stops. As all our children had stopped consuming the contaminated milk products for some time before assessment, the concentrations of melamine in both blood and urine are unhelpful. In view of the large number of children seen at our centre, a detailed history of consumption of melamine tainted milk products was not systematically recorded and was only rechecked if abnormalities found on ultrasonography were confirmed. Furthermore, as different batches of the same brand of milk product could be contaminated to varying degrees[8] it was not possible to accurately calculate exposure to melamine. There was also no easy way to avoid overestimation of children's consumption of contaminated products by anxious parents. Despite these problems, dietary history and parental perceived risk of exposure are the only surrogate measures of melamine exposure available and we are therefore subject to misclassification bias.

Conclusions and policy implications

The data from our cohort suggest that large scale and urgent screening programmes may not be informative or cost effective in regions outside of the mainland of China. Evidence to show that urgent massive screening after exposure to low dose melamine will lead to any health benefits is lacking. It is possible that hyperechoic lesions at the renal papillae may be associated with exposure to low dose melamine—that is, levels below 0.63 mg/kg/day, but the clinical significance of these lesions is uncertain at this stage and long term follow-up is mandatory. In view of the lack of evidence to guide the government initially, and the large number of severely affected children on the mainland, a large scale, territory-wide urgent screening programme was probably justifiable. In light of results of our screening programme, it may now be acceptable to arrange renal assessments for select groups on routine clinical service to avoid stressing the already overworked public health system of Hong Kong.

The use of reagent strip testing of urine as the primary means of identifying haematuria may lead to a large number of false positive results,[12] resulting in unnecessary anxiety for children and their parents. We believe, however, that reagent strip testing remains a valuable tool for identifying people for further confirmatory testing by microscopy of urine in screening programmes handling large numbers of patients in a short period. Arranging microscopy as soon as possible after a positive reagent strip result would help minimise anxiety for children and their parents.

We postulate that the difference in prevalence and severity of renal complications between our children and their counterparts on the mainland can be explained by the difference in levels of exposure to melamine. The severe acute complications observed on the mainland seem unlikely to occur elsewhere. Further medium and long term follow-up studies of these children are warranted to assess more comprehensively the public health impact of consuming milk products contaminated with melamine.

HSL and PCN contributed equally to this article. We thank our research team, So Hung Kwan, Ngai Hoi Yan, Simmy Yeung, and Jessie Mak, who helped organise and enter a large number of data within a short period.
Contributors: See bmj.com.
Funding: The government of the Hong Kong Special Administrative Region provided out of hours pay to staff working extra time during screening. The researchers were independent of the funders.
Competing interests: None declared.
Ethical approval: In light of the severe adverse renal outcomes associated with consumption of melamine tainted milk products in children on the mainland of China, an expert committee appointed by the government of the Hong Kong Special Administrative Region endorsed a screening programme comprising urinalysis, renal ultrasonography, and blood tests for renal function. There were no adverse health risks from the renal ultrasound scans and urinalysis, and only 1 ml of whole blood was taken for plasma urea and creatinine determination. Participants were free to join the programme and therefore implied consent was assumed for the assessments and no formal written consent was required. The data presented in this report were obtained from the participants who attended our centre and are anonymised. Ethical approval for analysis and publication of the screening programme results was granted by the joint The Chinese University of Hong Kong—New Territories East Cluster Clinical Research Ethics Committee after retrospective review.
Provenance and peer review: Not commissioned; externally peer reviewed.

1 Parry J. China's tainted infant formula sickens nearly 13 000 babies. *BMJ* 2008;337:a1802.
2 Parry J. Contaminated infant formula sickens 6200 babies in China. *BMJ* 2008;337:a1738.
3 Chan EY, Griffiths SM, Chan CW. Public-health risks of melamine in milk products. *Lancet* 2008;372:1444-5.
4 World Health Organization. Melamine and cyanuric acid: toxicity, preliminary risk assessment and guidance on levels in food. Geneva: WHO, 2008. www.who.int/foodsafety/fs_management/Melamine.pdf.
5 Dobson RL, Motlagh S, Quijano M, Cambron RT, Baker TR, Pullen AM, et al. Identification and characterization of toxicity of contaminants in pet food leading to an outbreak of renal toxicity in cats and dogs. *Toxicol Sci* 2008;106:251-62.
6 Cianciolo RE, Bischoff K, Ebel JG, Van Winkle TJ, Goldstein RE, Serfilippi LM. Clinicopathologic, histologic, and toxicologic findings in 70 cats inadvertently exposed to pet food contaminated with melamine and cyanuric acid. *J Am Vet Med Assoc* 2008;233:729-37.
7 Parry J. China's tainted milk scandal spreads around world. *BMJ* 2008;337:a1890.
8 Centre for Food Safety. Frequently asked questions. 2008. www.cfs.gov.hk/english/whatsnew/whatsnew_fstr/whatsnew_fstr_Test_dairy_product_FAQ.html.
9 Schwartz GJ, Haycock GB, Edelmann CM Jr, Spitzer A. A simple estimate of glomerular filtration rate in children derived from body length and plasma creatinine. *Pediatrics* 1976;58:259-63.
10 La Manna A, Polito C, Cioce F, De Maria G, Capacchione A, Rocco CE, et al. Calyceal microlithiasis in children: report on 196 cases. *Pediatr Nephrol* 1998;12:214-7.
11 Good LI, Edell SL, Soloway RD, Trotman BW, Mulhern C, Arger PA. Ultrasonic properties of gallstones. Effect of stone size and composition. *Gastroenterology* 1979;77:258-63.
12 Robbiani JH, Simonetti GD, Crosazzo L, Ferrarini A, Pronzini F, Bianchetti MG. False positive dipstick for urinary blood in childhood. *J Nephrol* 2006;19:605-6.
13 Pugia MJ, Lott JA, Kajima J, Saambe T, Sasaki M, Kuromoto K, et al. Screening school children for albuminuria, proteinuria and occult blood with dipsticks. *Clin Chem Lab Med* 1999;37:149-57.
14 Pugia MJ, Murakami M, Lott JA, Ohta Y, Kitagawa T, Yamauchi K, et al. Screening for proteinuria in Japanese schoolchildren: a new approach. *Clin Chem Lab Med* 2000;38:975-82.

Accepted: 8 December 2008

Comprehensive warm-up programme to prevent injuries in young female footballers: cluster randomised controlled trial

Torbjørn Soligard,[1] Grethe Myklebust,[1] Kathrin Steffen,[1] Ingar Holme,[1] Holly Silvers,[2] Mario Bizzini,[3] Astrid Junge,[3] Jiri Dvorak,[3] Roald Bahr,[1] Thor Einar Andersen[1]

EDITORIAL by Brooks and Erith

[1]Oslo Sports Trauma Research Centre, Norwegian School of Sport Sciences, PO Box 4014 Ullevaal Stadion, 0806 Oslo, Norway
[2]Santa Monica Orthopaedic and Sports Medicine Research Foundation, 1919 Santa Monica Blvd, Suite 350, Santa Monica, CA 90404 USA
[3]FIFA Medical Assessment and Research Centre, Schulthess Clinic, Lengghalde 2, CH-8008 Zurich, Switzerland
Correspondence to: T Soligard
torbjorn.soligard@nih.no

Cite this as: *BMJ* 2008;337:a2469
doi:10.1136/bmj.a2469

This article is an abridged version of a paper that was published on bmj.com. Cite this article as: *BMJ* 2008;337:a2469

ABSTRACT

Objective To examine the effect of a comprehensive warm-up programme designed to reduce the risk of injuries in female youth football.
Design Cluster randomised controlled trial with clubs as the unit of randomisation.
Setting 125 football clubs from the south, east, and middle of Norway (65 clusters in the intervention group; 60 in the control group) followed for one league season (eight months).
Participants 1892 female players aged 13-17 (1055 players in the intervention group; 837 players in the control group).
Intervention A comprehensive warm-up programme to improve strength, awareness, and neuromuscular control during static and dynamic movements.
Main outcome measure Injuries to the lower extremity (foot, ankle, lower leg, knee, thigh, groin, and hip).
Results During one season, 264 players had relevant injuries: 121 players in the intervention group and 143 in the control group (rate ratio 0.71, 95% confidence interval 0.49 to 1.03). In the intervention group there was a significantly lower risk of injuries overall (0.68, 0.48 to 0.98), overuse injuries (0.47, 0.26 to 0.85), and severe injuries (0.55, 0.36 to 0.83).
Conclusion Though the primary outcome of reduction in lower extremity injury did not reach significance, the risk of severe injuries, overuse injuries, and injuries overall was reduced. This indicates that a structured warm-up programme can prevent injuries in young female football players.
Trial registration ISRCTN10306290.

INTRODUCTION

Playing football entails a substantial risk of injury, and female footballers have rates of injury similar to those in men.[1-4] Women, however, might have a higher risk of serious injury; the rate of anterior cruciate ligament injuries is three to five times higher for girls than for boys.[5] The high injury rate constitutes a considerable problem, both short term and in the dramatic increase in the risk of early osteoarthritis.[6] There are only a few small or non-randomised studies on prevention of injury in female football players.[7-9]

Fig 1 | Example of running exercise illustrating key objectives of all running, jumping, cutting, and landing exercises: core stability and correct lower extremity alignment. Left: correct technique; right: incorrect technique with pelvic tilt and knee valgus alignment to right

Our recent randomised controlled trial examined the effect of a structured training programme ("The 11")[10] over one season among 2000 female players aged 13-17.[11] We found no difference in the risk of injury between the intervention group and control group, though the study was limited by low compliance among the intervention teams.

This led us to develop an exercise programme to improve both the preventive effect of the programme and the compliance of coaches and players. The revised programme ("The 11+") included key exercises and additional exercises to provide variation and progression. It also included a new set of structured running exercises that made it better suited as a comprehensive warm-up programme for training and matches.

We conducted a cluster randomised controlled trial to examine the effect of the revised programme on rates of lower extremity (foot, ankle, lower leg, knee, thigh, groin, and hip) injury in young female footballers.

METHODS

We randomised 125 clubs study to the intervention or control group. All teams from one club were randomised to the same treatment arm. Five clubs included two teams each.

Intervention

An expert group convened by the international football federation (FIFA) developed the warm-up programme. It consisted of three parts: running exercises at slow speed combined with active stretching; six different sets of exercises, including strength, balance, and jumping exercises, each with three levels of increasing difficulty; and speed running combined with football specific movements with sudden changes in direction

From February to mid-April 2007, we invited the coaches and team captains from all clubs in the intervention group to a three hour instructional course in which we introduced the warm-up programme. Instructors from the Oslo Sports Trauma Research Centre arranged courses at different locations in each of the eight regional districts.

The coaches received an instructional DVD showing all of the exercises in the programme (see fig 1 for example), a loose leaf exercise book, and small exercise cards attached to a neck strap. In addition, the coaches and every player received a poster explaining every exercise. We asked the coaches to use the complete exercise programme as the warm up for every training session throughout the season and to use the running exercises in the programme as part of their warm up for every match. Once players were familiar with the exercises the programme took about 20 minutes to complete.

Throughout the season, researchers contacted the coaches regularly by email and telephone. Clubs in both groups were offered an incentive in the form of high quality footballs, provided they completed data registration throughout the study period. Despite these measures, 13 clubs in the intervention group did not start the warm-up programme nor did they deliver any data on injury or exposure. Nineteen clubs in the control group did not provide any data.

Outcome measures

We defined the primary outcome as an injury to the lower extremity (foot, ankle, lower leg, knee, thigh, groin, and hip) and secondary outcomes as any injury, or an injury to the ankle, knee, or other body parts. We included all injuries reported after an intervention club had completed the first prevention training session (matched with the same date in a control club) to compare the risk of injury between the groups.

Exposure and injury registration

The coaches reported injuries and details of an individual player's participation for each training session and match, as well as to what extent the warm-up programme was carried out each session

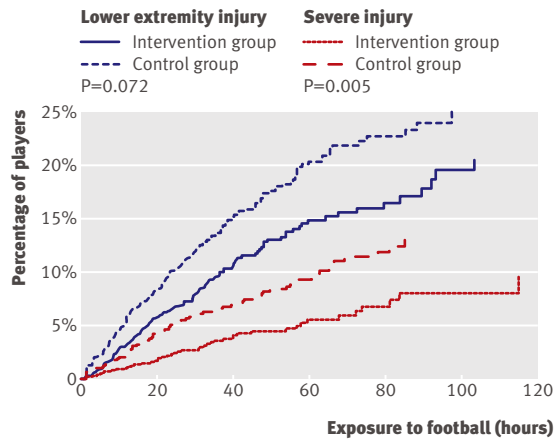


Fig 2 | Survival curves based on Cox regression for players with lower extremity injury and severe injury

(intervention clubs) on weekly registration forms throughout the study period. These were submitted by email, mail, or fax to the research centre. Data on players who dropped out during the study period were included for the entire period of their participation.

At the research centre one physical therapist and one medical student, who were blinded to group allocation, recorded injuries. Every injured player was contacted to assess aspects of the injury based on a standardised injury questionnaire.[12]

Sample size and statistics

We calculated our sample size on the basis of data on incidence of injury in young female footballers in Norway during the 2005 season.[11] We estimated that 16% would incur an injury to the lower extremities and about 10-12% would injure a knee or ankle during one season. Our model was based on 18 players per club and a dropout rate of 15%, which means that we needed to include about 120 clubs with 2150 players.

We conducted all statistical analyses according to a prespecified plan using Stata, version 10.0 (StataCorp, College Station, TX) (see bmj.com for details of analysis).

RESULTS

The final sample consisted of 52 clubs (1055 players) in the intervention group and 41 clubs (837 players) in the control group. The players in the two groups were similar in age (15.4 (SD 0.7) years in both groups) and the dropout rate was similar between the groups (23 (2.1%) *v* 24 (2.9%)).

Numbers and severity of injuries in young female footballers according to use of warm-up exercise programme (intervention)

	Intervention (n=1055)	Control (n=837)	Rate ratio (95% CI)*	P value (z test)
All injuries:				
Total	161	215	0.68 (0.56 to 0.84)	0.0003
Match	109	138	0.71 (0.55 to 0.91)	0.007
Training	51	74	0.63 (0.44 to 0.90)	0.012
Minimal injuries (1-3 days)	27	32	0.77 (0.46 to 1.28)	0.313
Mild injuries (4-7 days)	24	34	0.64 (0.38 to 1.08)	0.097
Moderate injuries (8-28 days)	63	70	0.82 (0.58 to 1.15)	0.250
Severe injuries (>28 days)	47	79	0.54 (0.38 to 0.78)	0.0009
Overuse injuries:				
Total	25	52	0.44 (0.27 to 0.71)	0.0007
Minimal injuries	5	10	0.46 (0.16 to 1.33)	0.142
Mild injuries	3	7	0.39 (0.10 to 1.51)	0.174
Moderate injuries	9	11	0.75 (0.31 to 1.80)	0.509
Severe injuries	8	24	0.30 (0.14 to 0.68)	0.004
Acute injuries:				
Total	136	163	0.76 (0.61 to 0.95)	0.017
Minimal injuries	22	22	0.91 (0.50 to 1.64)	0.757
Mild injuries	21	27	0.71 (0.40 to 1.25)	0.234
Moderate injuries	54	59	0.83 (0.58 to 1.21)	0.332
Severe injuries	39	55	0.65 (0.43 to 0.97)	0.037
Contact	53	76	0.64 (0.45 to 0.90)	0.011
Non-contact	55	58	0.86 (0.60 to 1.25)	0.435
Acute knee injuries	27	37	0.66 (0.41 to 1.09)	0.105
Acute ankle injuries	51	52	0.89 (0.61 to 1.31)	0.562

*Rate ratio obtained from Poisson model.

Exposure and injury characteristics

Those in the intervention group played 49 899 hours of football (16 057 hours of matches and 33 842 hours of practice). The figure for the control group was 45 428 hours (14 342 and 31 086). During the eight month season, 301 (16%) of the 1892 players included in the study sustained a total of 376 injuries; 161 in the intervention group, 215 in the control group. There were 299 (80%) acute injuries and 77 (20%) overuse injuries. The overall incidence of injuries was 3.9 (SD 0.2) per 1000 player hours (8.1 (SD 0.5) in matches and 1.9 (SD 0.2) in training).

Effect of revised injury prevention programme

The rate ratio for players with a lower extremity injury between the intervention and the control group was 0.71 (0.49 to 1.03, P=0.072). There was a significantly lower risk of injuries overall (0.68, 0.48 to 0.98), overuse injuries (0.47, 0.26 to 0.85), and severe injuries (0.55, 0.36 to 0.83) in the intervention group. The reduction in the risk of match injuries, training injuries, knee injuries, and acute injuries (from 26% to 38%) did not reach significance. The number needed to treat to prevent one injury varied from 15 to 63 players. Figure 2 shows survival curves for lower extremity injuries and severe injuries in the two groups.

Compared with the control group, significantly fewer players in the intervention group had two or more injuries (rate ratio 0.51, 95% confidence interval 0.29 to 0.87), while a reduction in the risk of re-injuries did not reach significance (0.46, 0.20 to 1.01). The table shows the severity distribution for different types of injury. The rate of severe injuries, severe overuse injuries, and severe acute injuries was significantly lower in the intervention group.

Compliance with programme

The 52 clubs in the intervention group performed the injury prevention programme for 44 (SD 22, range 11-104) sessions (77%) throughout the season. None of the clubs in the control group reported performing structured warm-up exercises comparable with the intervention. The risk of injury was 35% lower in intervention players in the third with the highest compliance (2.6 (20. to 3.2) injuries/1000 player hours, mean (range 33-95) 49.2 sessions) compared with players in the intermediate third (4.0 (3.0 to 5.0) injuries/1000 player hours, mean 23.4 (15-32) sessions) (rate ratio 0.65, 0.44 to 0.94, P=0.02). The 32% reduction in risk of injury compared with the third with the lowest compliance (3.7 (2.2 to 5.3) injuries/1000 player hours, mean 7.7 (0-14) sessions) did not reach significance (rate ratio 0.68, 0.41 to 1.12, P=0.13).

DISCUSSION

This randomised controlled trial of a structured warm-up programme in young female footballers showed

WHAT IS ALREADY KNOWN ON THIS TOPIC

The injury rate among female footballers, regardless of age and level of performance, approaches that of men

The risk of severe knee injuries, such as anterior cruciate ligament injuries, is three to five times higher for female than male football players

Studies from other sports indicate that it might be possible to reduce the rate of lower extremity injuries, but no programmes have been validated for female footballers

WHAT THIS STUDY ADDS

A comprehensive warm-up programme designed to improve strength, awareness, and neuromuscular control can prevent injuries in young female footballers

The risk of injury can be reduced by about one third and the risk of severe injuries by as much as a half

that the risk of injury can be reduced by about one third and severe injuries by as much as one half. Although the rate ratios for the different outcome variables indicated a consistent effect on risk of injury across most types of injury, the primary outcome—lower extremity injury—did not reach significance when we adjusted for the cluster sampling. There was, however, a significant reduction in several secondary outcome variables, including the rate of severe injuries, overuse injuries, and injuries overall.

Methodological considerations

Of the 125 clubs randomised, we could not include 13 intervention clubs and 19 control clubs in the analyses because they did not deliver any data on injury or exposure. In most cases the coaches were volunteers, such as parents, and the most common reason for not reporting any data was the additional work of registering and reporting data weekly.

With respect to the internal validity, we found no differences between the two groups in their training or match exposure during the study. Having recorded individual exposure, we could adjust for playing time. Individual exposure also takes censorship into account, such as abbreviated lengths of follow-up for reasons other than injury (such as illness, moving, quitting the sport).[13] Another advantage of this approach is that it provides accurate data about each player's exposure to the intervention, in this case the injury prevention programme. Given the individual activity logs kept by the coaches, it is unlikely that injuries would go unreported, and we see no reason to expect a reporting bias between the groups.

Compliance

Compliance in the current trial was higher than with the previous programme (77% *v* 52%),[11] and we saw effects on the risk of injury. Site visits indicated that not all of the players seemed to concentrate fully on the performance of the exercises and the compliance logs documented that not all clubs completed the requested minimum of two training sessions a week. We included all clubs and players in the intention to treat analysis, which means that the preventive effect of the programme might be higher than reported. This is supported by subgroup analyses within the intervention group, indicating a trend towards a lower risk of injury among the most compliant players.

Structured programme of warm-up exercises to prevent injuries

The revised exercises include both variety and progression of difficulty, absent from the previously tested training programme[11] but present in other successful prevention programmes.[14-17] Our prevention programme is multifaceted and addresses many factors that could be related to the risk of injury . Further studies are needed to determine what the key components are so that future programmes might require less time and effort.

Except for a few reports from coaches on muscular soreness in the beginning of the intervention period and one report about a minor hamstring strain, we observed no negative effects of the programme.

Implications

We do not know if the results can be generalised to both sexes, other age groups, or other youth sports. Similar preventive programmes, however, were effective in senior elite football,[15 18] young male footballers,[19] and in both sexes in other sports.[17 20] Mechanisms for serious knee injuries seem to be comparable across many sports (mostly non-contact, resulting from pivoting and landing movements). It therefore seems reasonable to assume that the programme we used could be modified for use in other similar sports.

We thank the project assistants (Birgitte Lauersen, Agnethe Nilstad, Ellen Blom, Olav Kristianslund, Tone Wigemyr, Monika Bayer, Heidi M Pedersen, Vegar Vallestad, and John Bjørneboe), the coaches, and the players who participated in this study. A poster illustrating various exercise components and progressions of programme is available at www.ostrc.no/en/Project/144—The-11-plus/. Also, videos displaying every exercise in the programme (with Norwegian text and narrator) are available at www.klokavskade.no/no/Skadefri/Fotball/SPILLEKLAR/.

Contributors: See bmj.com.

Funding: This study was supported by grants from the FIFA Medical Assessment and Research Centre. The Oslo Sports Trauma Research Center has been established at the Norwegian School of Sport Sciences through grants from the Royal Norwegian Ministry of Culture and Church Affairs, the South-Eastern Norway Regional Health Authority, the Norwegian Olympic Committee and Confederation of Sport, and Norsk Tipping AS. No author or related institution has received any financial benefit from research in this study.

Competing interests: None declared.

Ethical approval: The study was approved by the regional committee for medical research ethics, South-Eastern Norway Regional Health Authority, Norway. Players and parents gave individual written informed consent.

Provenance and peer review: Not commissioned; externally peer reviewed.

1 Giza E, Mithofer K, Farrell L, Zarins B, Gill T. Injuries in women's professional soccer. *Br J Sports Med* 2005;39:212-6.

2 Jacobson I, Tegner Y. Injuries among Swedish female elite football players: a prospective population study. *Scand J Med Sci Sports* 2007;17:84-91.

3 Junge A, Dvorak J. Injuries in female football players in top-level international tournaments. *Br J Sports Med* 2007;41(suppl 1):i3-7.

4 Tegnander A, Olsen OE, Moholdt TT, Engebretsen L, Bahr R. Injuries in Norwegian female elite soccer: a prospective one-season cohort study. *Knee Surg Sports Traumatol Arthrosc* 2008;16:194-8.
5 Bjordal JM, Arnly F, Hannestad B, Strand T. Epidemiology of anterior cruciate ligament injuries in soccer. *Am J Sports Med* 1997;25:341-5.
6 Lohmander LS, Östenberg A, Englund M, Roos H. High prevalence of knee osteoarthritis, pain, and functional limitations in female soccer players twelve years after anterior cruciate ligament injury. *Arthritis Rheum* 2004;50:3145-52.
7 Heidt RS Jr, Sweeterman LM, Carlonas RL, Traub JA, Tekulve FX. Avoidance of soccer injuries with preseason conditioning. *Am J Sports Med* 2000;28:659-62.
8 Söderman K, Werner S, Pietila T, Engström B, Alfredson H. Balance board training: prevention of traumatic injuries of the lower extremities in female soccer players? A prospective randomized intervention study. *Knee Surg Sports Traumatol Arthrosc* 2000;8:356-63.
9 Mandelbaum BR, Silvers HJ, Watanabe DS, Knarr JF, Thomas SD, Griffin LY, et al. Effectiveness of a neuromuscular and proprioceptive training program in preventing anterior cruciate ligament injuries in female athletes: 2-year follow-up. *Am J Sports Med* 2005;33:1003-10.
10 Dvorak J, Junge A. *Football medicine manual*. Zurich: F-MARC, 2005:81-93.
11 Steffen K, Myklebust G, Olsen OE, Holme I, Bahr R. Preventing injuries in female youth football—a cluster-randomized controlled trial. *Scand J Med Sci Sports* 2008 Jan 14 [epub ahead of print].
12 Olsen OE, Myklebust G, Engebretsen L, Bahr R. Injury pattern in youth team handball: a comparison of two prospective registration methods. *Scand J Med Sci Sports* 2006;16:426-32.
13 Bahr R, Holme I. Risk factors for sports injuries—a methodological approach. *Br J Sports Med* 2003;37:384-92.
14 Olsen OE, Myklebust G, Engebretsen L, Holme I, Bahr R. Exercises to prevent lower limb injuries in youth sports: cluster randomised controlled trial. *BMJ* 2005;330:449.
15 Caraffa A, Cerulli G, Projetti M, Aisa G, Rizzo A. Prevention of anterior cruciate ligament injuries in soccer. A prospective controlled study of proprioceptive training. *Knee Surg Sports Traumatol Arthrosc* 1996;4:19-21.
16 Myklebust G, Engebretsen L, Brækken IH, Skjølberg A, Olsen OE, Bahr R. Prevention of anterior cruciate ligament injuries in female team handball players: a prospective intervention study over three seasons. *Clin J Sport Med* 2003;13:71-8.
17 Emery CA, Cassidy JD, Klassen TP, Rosychuk RJ, Rowe BH. Effectiveness of a home-based balance-training program in reducing sports-related injuries among healthy adolescents: a cluster randomized controlled trial. *CMAJ* 2005;172:749-54.
18 Arnason A, Andersen TE, Holme I, Engebretsen L, Bahr R. Prevention of hamstring strains in elite soccer: an intervention study. *Scand J Med Sci Sports* 2008;18:40-8.
19 Junge A, Rosch D, Peterson L, Graf-Baumann T, Dvorak J. Prevention of soccer injuries: a prospective intervention study in youth amateur players. *Am J Sports Med* 2002;30:652-9.
20 Hewett TE, Lindenfeld TN, Riccobene JV, Noyes FR. The effect of neuromuscular training on the incidence of knee injury in female athletes. A prospective study. *Am J Sports Med* 1999;27:699-706.

Accepted: 26 September 2008

Ideas, concerns, and expectations

Yesterday I was a foundation year 1 doctor in cardiology, today I am a foundation year 2 doctor in general practice. I have my own office, computer, chlamydia screening kits, and even a secretary. I also, however, have my own patients. The problem here is that they arrive expecting to see a confident and experienced general practitioner. Unfortunately, I am neither. Luckily my tutor, who is both, is on hand to help during my consulting sessions, and also offers a weekly tutorial.

This week the tutorial was on consultation skills, and in particular how a doctor should elicit a patient's ideas, concerns, and expectations during a consultation. After the tutorial, I am keen to put this into practice. My next patient seems an ideal candidate: he is an elderly man who has come to see me about his constipation. There are no worrying features in the history, and it would have been easy to send him off with a prescription for my favourite laxative, but I stop myself. I ask, "What do you think might be causing the constipation?"

"I dunno," comes the reply. "You're the doctor," he helpfully reminds me.

"But are you concerned about anything in particular?" I persist.

"Not really."

There must be something on his mind, I think. Perhaps I need to be more direct. "Are you worried about cancer?" I ask.

"No, not at all," he replies sharply.

With the patient's ideas and concerns skilfully elicited, I move on: "What do you think would be the best course of action?"

"You're the doctor," is the predictable response.

The patient leaves with his prescription and an appointment to see me in a few days if he isn't feeling better. I congratulate myself on putting my new skills into action, thinking that the consultation had gone well, and moved on to see the next patient.

The next day I get a message asking me to telephone the constipated patient I had seen yesterday. I naturally assume the worst—he's probably perforated his colon, and his family want my GMC number. However, when the telephone is picked up I get a worried answer from the man himself. He is concerned about the consultation we had yesterday and wants to see me to discuss it. I put him on the end of the afternoon's list and wonder what the problem could be.

When the patient returns, I see he's looking worried. "What's the problem?" I ask.

"Do you think I have cancer, doctor?"

"No. Why do you ask that?"

"Well, you mentioned it yesterday out of the blue, and it hadn't even crossed my mind. I've been worrying about it ever since."

Suddenly the penny drops. In my clumsy efforts to elicit his ideas, concerns, and expectations, I had actually given him something to be concerned about. Cancer had never crossed his mind, and why should it have? Such direct questioning did nothing but give the poor man a sleepless night of worry.

During my GP placement my consultations skills have improved considerably. Discussing ideas, concerns, and expectations is still high on the agenda, but now it is more delicately approached and more patient specific. If done correctly it allows the patient to give you all the answers you need, and often the diagnosis, without interruption.

J Lilleker foundation year 2 doctor, Bodey Medical Centre, Manchester
jimmylill@blueyonder.co.uk
Cite this as: *BMJ* 2008;337:a3135