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

# Effect of fibre, antispasmodics, and peppermint oil in the treatment of irritable bowel syndrome: systematic review and meta-analysis

Alexander C Ford,[1] Nicholas J Talley,[2] Brennan M R Spiegel,[3] Amy E Foxx-Orenstein,[4] Lawrence Schiller,[5] Eamonn M M Quigley,[6] Paul Moayyedi[1]

EDITORIAL by Jones

[1]Gastroenterology Division, McMaster University, Health Sciences Centre, 1200 Main Street West, Hamilton, ON, L8N 3Z5, Canada
[2]Department of Medicine, Mayo Clinic Florida, Jacksonville, FL, USA
[3]VA Greater Los Angeles Healthcare System; UCLA/VA Center for Outcomes Research and Education, Los Angeles, CA, USA
[4]Division of Gastroenterology and Hepatology, Mayo Clinic, Rochester, MN, USA
[5]Digestive Health Associates of Texas, Baylor University Medical Center, Dallas, TX, USA
[6]Department of Medicine, Cork University Hospital, Ireland
Correspondence to: A C Ford
alexf12399@yahoo.com

Cite this as: *BMJ* 2008;337:a2313
doi:10.1136/bmj.a2313

This article is an abridged version of a paper that was published on bmj.com. Cite this article as: *BMJ* 2008;337:a2313

## ABSTRACT

**Objective** To determine the effect of fibre, antispasmodics, and peppermint oil in the treatment of irritable bowel syndrome.

**Design** Systematic review and meta-analysis of randomised controlled trials.

**Data sources** Medline, Embase, and the Cochrane controlled trials register up to April 2008.

**Review methods** Randomised controlled trials comparing fibre, antispasmodics, and peppermint oil with placebo or no treatment in adults with irritable bowel syndrome were eligible for inclusion. The minimum duration of therapy considered was one week, and studies had to report either a global assessment of cure or improvement in symptoms, or cure of or improvement in abdominal pain, after treatment. A random effects model was used to pool data on symptoms, and the effect of therapy compared with placebo or no treatment was reported as the relative risk (95% confidence interval) of symptoms persisting.

**Results** 12 studies compared fibre with placebo or no treatment in 591 patients (relative risk of persistent symptoms 0.87, 95% confidence interval 0.76 to 1.00). This effect was limited to ispaghula (0.78, 0.63 to 0.96). Twenty two trials compared antispasmodics with placebo in 1778 patients (0.68, 0.57 to 0.81). Various antispasmodics were studied, but otilonium (four trials, 435 patients, relative risk of persistent symptoms 0.55, 0.31 to 0.97) and hyoscine (three trials, 426 patients, 0.63, 0.51 to 0.78) showed consistent evidence of efficacy. Four trials compared peppermint oil with placebo in 392 patients (0.43, 0.32 to 0.59).

**Conclusion** Fibre, antispasmodics, and peppermint oil were all more effective than placebo in the treatment of irritable bowel syndrome.

## INTRODUCTION

Guidelines for the management of irritable bowel syndrome in the United Kingdom recommend that the diagnosis should be made on clinical grounds alone, without invasive investigations, unless alarm symptoms such as rectal bleeding are present.[1 2] General practitioners therefore need efficacious treatments that do not require monitoring and are cheap, safe, and readily available. This is particularly relevant as newer and more expensive drugs have failed to show efficacy or have been withdrawn because of concerns about adverse events. Traditionally, people with irritable bowel syndrome were instructed to increase their intake of dietary fibre.[3] When this failed, smooth muscle relaxants and antispasmodics were used.[2] More recently, peppermint oil, shown to have antispasmodic properties,[4] has been available over the counter and used to treat irritable bowel syndrome.

Whether these agents are effective in treating irritable bowel syndrome is controversial. Results of randomised controlled trials are conflicting and systematic reviews have come to different conclusions.[5-12] We carried out a systematic review and meta-analysis to determine the effect of fibre, antispasmodics, and peppermint oil in the treatment of irritable bowel syndrome.

## METHODS

We carried out an electronic search for randomised controlled trials of adults with irritable bowel syndrome diagnosed by a clinician or that met specific diagnostic criteria (for example, the Rome classification), combined with investigations to exclude organic disease when necessary. The studies had to compare fibre, antispasmodics, and peppermint oil with placebo or no treatment. Participants were required to be treated for a minimum of one week and to be followedup for at least one week, and studies had to report on either global symptoms of irritable bowel syndrome or abdominal pain after treatment. This was preferably as reported by the patient, but could be documented by a doctor. We also considered the first period of cross over randomised controlled trials. (See bmj.com for details of search strategy.)

Two reviewers assessed articles according to eligibility criteria. Disagreements were resolved by consensus. The primary outcomes assessed were the efficacy of fibre, antispasmodics, and peppermint oil compared with placebo or no treatment on global symptoms of irritable bowel syndrome or abdominal pain after treatment. Secondary outcomes included

efficacy according to type of fibre or antispasmodic, and adverse events.

Two reviewers extracted data on the primary outcomes. In addition we extracted data on setting, number of centres, country, dose and duration of treatment, number of adverse events, definition of irritable bowel syndrome, primary outcome measure to define improvement in symptoms or cure, method for generating the randomisation schedule, method for allocation concealment, blinding, proportion of females, disease subtype (predominant stool pattern), and duration of follow-up. Data were extracted as intention to treat analyses whenever possible. If this was not clear from the article then we analysed patients with evaluable data. Two reviewers assessed study quality according to the Jadad scale.[13]

### Data synthesis and statistical analysis

We pooled data using a random effects model, allowing for heterogeneity between studies.[14] The effects of interventions were expressed as relative risks (95% confidence intervals) of global symptoms of irritable bowel syndrome or abdominal pain persisting with fibre, antispasmodics, or peppermint oil compared with placebo or no treatment. From the reciprocal of the risk difference from the meta-analysis we calculated the number needed to treat (95% confidence interval). We used the $I^2$ statistic (cut-off point 25%),[15] to assess heterogeneity between studies and the $\chi^2$ test with a P value <0.10 to define significant heterogeneity. We planned sensitivity analyses a priori according to type of fibre or antispasmodic, predominant stool pattern, and study quality. If adverse events were significantly increased with active treatment we calculated the number needed to harm (95% confidence interval).

We used Review Manager and StatsDirect software to generate forest plots of pooled relative risks and risk differences (95% confidence intervals) for outcomes. We used Egger and Begg tests to assess funnel plots for publication bias.[16]

## RESULTS

Thirty five of 101 potentially relevant randomised controlled trials were included in the study (see bmj.com).[w1-w35]

### Fibre

Twelve trials (n=591) compared fibre with placebo or a low fibre diet (see bmj.com for details).[w1-w9 w33-w35] The proportion of women in the trials ranged between 20% and 90%. Only three studies reported on disease subtype according to predominant stool pattern.[w1 w4 w6] Two recruited only patients with predominant constipation,[w1 w4] and in the other trial 49% of patients had predominant constipation.[w6] Five studies used bran,[w1-w3 w7 w34] six used ispaghula,[w5 w6 w8 w9 w33 w35] and one used "concentrated" fibre of an unspecified type.[w4] Seven of the studies scored 4 or more on the Jadad scale.[w2 w4-w6 w8 w9 w33] None reported the method of allocation concealment.

Overall, 155 of 300 (52%) patients assigned to fibre had persistent or unimproved symptoms compared with 168 of 291 (57%) allocated to placebo or a low fibre diet (relative risk 0.87, 95% confidence interval 0.76 to 1.00, P=0.05), with no significant heterogeneity detected between studies ($I^2$=14.2%, P=0.31; see bmj.com). The number needed to treat with fibre to prevent one patient with persistent symptoms was 11 (95% confidence interval 5 to 100). The funnel plot showed no significant asymmetry (Egger test, P=0.84). When the seven studies scoring 4 or more on the Jadad scale were considered in the analysis the treatment effect for fibre was no longer significant (relative risk 0.90, 0.75 to 1.08).[w2 w4-w6 w8 w9 w33]

#### *Bran*

Five studies (n=221) compared bran with placebo or a low fibre diet.[w1-w3 w7 w34] Only one scored 4 or more on the Jadad scale.[w2] Sixty two of 114 (54%) patients assigned to bran had persistent symptoms compared with 58 of 107 (54%) allocated to placebo or a low fibre diet. Bran had no significant effect on symptoms (relative risk 1.02, 0.82 to 1.27; see bmj.com). No significant heterogeneity was detected between studies ($I^2$=0%, P=0.91) and there was no evidence of funnel plot asymmetry (Egger test, P=0.28).

#### *Ispaghula*

Six studies (n=321) compared ispaghula with placebo.[w5 w6 w8 w9 w33 w35] Eighty three of 161 (52%) patients allocated to ispaghula had persistent symptoms compared with 103 of 160 (64%) receiving placebo. Ispaghula was effective in treating symptoms (relative risk 0.78, 0.63 to 0.96; see bmj.com), with significant heterogeneity detected between studies ($I^2$=34.4%, P=0.18). The number needed to treat was 6 (3 to 50). No evidence of funnel plot asymmetry was found (Egger test, P=0.43). Five of the six studies scored 4 or more on the Jadad scale.[w5 w6 w8 w9 w33] When only these studies were considered in the analysis the treatment effect for ispaghula was no longer significant (relative risk 0.86, 0.74 to 1.01, P=0.06), with no significant heterogeneity detected between studies ($I^2$=2.6%, P=0.39).

#### *Adverse events*

Data on the total number of adverse events were provided by only four trials (n=251).[w4 w6 w34 w35] As the number of adverse events was small the data were not pooled. Three patients receiving fibre reported adverse events compared with two receiving placebo.

### Antispasmodics

Twenty two studies (n=1778) compared 12 different antispasmodics with placebo (see bmj.com for details).[w10-w28 w33-w35] The proportion of women in each trial ranged from 39% to 83%. Six studies reported on disease subtype.[w11 w12 w14 w15 w19 w24] One study recruited patients only with predominant constipation,[w15] and in the remaining five studies between 22%

and 64% of patients had predominant constipation. None of the trials reported the method of allocation concealment. Four trials used otilonium,[w16 w23 w25 w26] three cimetropium,[w11-w13] three hyoscine,[w10 w33 w35] three pinaverium,[w22 w24 w27] two trimebutine,[w15 w20] one trimebutine and rociverine,[w14] and one each of alverine,[w18] dicycloverine (dicyclomine),[w21] mebeverine,[w34] pirenzipine,[w17] prifinium,[w19] and propinox.[w28]

In total, 350 of 905 (39%) patients assigned to antispasmodics had persistent symptoms compared with 485 of 873 (56%) allocated to placebo (relative risk 0.68, 0.57 to 0.81), with significant heterogeneity detected between studies ($I^2$=62.6%, P<0.001; see bmj.com). The number needed to treat was 5 (4 to 9). The Egger test suggested funnel plot asymmetry (P=0.03), but this seemed to be driven by one small study; the Begg test did not confirm asymmetry (P=0.25). The treatment effect in favour of antispasmodics remained when only the 12 trials that scored 4 or more on the Jadad scale were considered in the analysis (relative risk 0.65, 0.48 to 0.89),[w11-w13 w15 w17-w19 w21 w25 w27 w33 w34] although the heterogeneity observed between studies persisted ($I^2$=70.2%, P=0.0001) and there was evidence of publication bias (Egger test, P=0.007).

The effect of different antispasmodics on symptoms was examined separately. Four trials (n=435) used otilonium.[w16 w23 w25 w26] Symptoms persisted in 111 of 216 (51%) patients assigned to otilonium compared with 155 of 219 (71%) receiving placebo (relative risk 0.55, 0.31 to 0.97, $I^2$=59.5%), and a number needed to treat of 4.5 (3.0 to 10.0). Three trials used cimetropium,[w11-w13] with 15 of 79 (19%) patients randomised to cimetropium having persistent symptoms compared with 42 of 79 (53%) receiving placebo (relative risk 0.38, 0.20 to 0.71, $I^2$=37.2%), with a number needed to treat of 3.0 (2.0 to 12.5). Three studies (n=426) compared hyoscine with placebo.[w10 w33 w35] Symptoms persisted in 63 of 215 (29%) patients receiving hyoscine compared with 97 of 211 (46%) allocated to placebo (relative risk 0.63, 0.51 to 0.78, $I^2$=0%), with a number needed to treat of 3.5 (2.0 to 25.0). Three trials (n=188) used pinaverium.[w22 w24 w27] Symptoms persisted in 26 of 94 (28%) patients assigned to pinaverium compared with 57 of 94 (61%) receiving placebo (relative risk 0.47, 0.33 to 0.67, $I^2$=0%) The number needed to treat was 3 (2 to 5). Three trials (n=140) studied trimebutine.[w14 w15 w20] Twenty eight of 70 (40%) patients assigned to trimebutine had persistent symptoms compared with 27 of 70 (39%) allocated to placebo (relative risk 1.08, 0.72 to 1.61, $I^2$=0%).

*Adverse events*

Thirteen studies (n=1379) reported a total number of adverse events.[w10-w16 w18 w20 w21 w25 w27 w34] Overall, 101 of 704 (14%) patients assigned to antispasmodics experienced adverse events compared with 62 of 675 (9%) allocated to placebo (relative risk of adverse events 1.62, 95% confidence interval 1.05 to 2.50), with significant heterogeneity detected between studies ($I^2$=37.9%, P=0.07), but no evidence of publication bias (Egger test, P=0.53). No trials reported serious adverse events. The number needed to harm was 17.5 (7.0 to 217.0).

### Peppermint oil

Four studies (n=392) compared peppermint oil with placebo (see bmj.com for details).[w29-w32] The proportion of women in each trial ranged from 40% to 76%. Only one study reported on disease subtype and recruited 25% of patients with predominant constipation and 75% with predominant diarrhoea.[w31] None of the trials reported the method of allocation concealment.

Fifty two of 197 (26%) patients randomised to peppermint oil had persistent symptoms compared with 127 of 195 (65%) receiving placebo (relative risk 0.43, 0.32 to 0.59; see bmj.com), with significant heterogeneity detected between studies ($I^2$=31.1%, P=0.23). The number needed to treat was 2.5 (2.0 to 3.0). When only the three studies that scored 4 or more on the Jadad scale were considered in the analysis the relative risk of persistent symptoms was of a similar magnitude (0.40, 0.29 to 0.55), with no significant heterogeneity detected between studies ($I^2$=22.0%, P=0.28).[w30-w32]

*Adverse events*

As only three studies reported data on adverse events,[w30-w32] the data were not pooled. Five adverse events occurred among 174 patients assigned to peppermint oil compared with no adverse events in 171 patients receiving placebo.

## DISCUSSION

This systematic review and meta-analysis has shown that fibre, antispasmodics, and peppermint oil are all more effective than placebo in the treatment of irritable bowel syndrome. The number needed to treat to prevent one patient having persistent symptoms was 11 for fibre, 5 for antispasmodics, and 2.5 for peppermint oil. Adverse events were significantly more frequent in those receiving antispasmodics in than those receiving placebo, but none was serious. As several different treatments were studied we carried out subgroup analyses (see bmj.com).

Wheat bran was no more effective at treating irritable bowel syndrome than placebo or a low fibre diet. The beneficial effect of fibre seemed to be limited to ispaghula (number needed to treat of 6). However, significant heterogeneity was detected between trials. When only high quality studies were considered in the analysis this heterogeneity was diminished, but the difference in effect in favour of ispaghula only reached marginal statistical significance.

Antispasmodics were of benefit, but heterogeneity between study results was significant, and there was evidence of publication bias. Data were limited for many of the drugs licensed for use in the United Kingdom. Most data were available for otilonium,

## WHAT IS ALREADY KNOWN ON THIS TOPIC

Irritable bowel syndrome is a chronic, relapsing and remitting disorder, which can be difficult to treat

Safe, effective treatments are required, as newer more expensive therapies have been withdrawn because of concerns about safety

Fibre, antispasmodics, and peppermint oil may fulfil this role, but evidence for their use is conflicting owing to methodological errors in previous systematic reviews

## WHAT THIS STUDY ADDS

Fibre, antispasmodics (particularly hyoscine and otilonium), and peppermint oil were all more effective than placebo for treating irritable bowel syndrome

The numbers needed to treat with these therapies were 11, 5, and 2.5, respectively

Doctors should consider ispaghula, antispasmodics (preferably hyoscine as first line treatment), and peppermint oil to treat irritable bowel syndrome

trimebutine, cimetropium, hyoscine, and pinaverium. Trimebutine seemed to have no benefit over placebo, whereas the other four drugs all significantly reduced the risk of persistent symptoms. Considerable heterogeneity was, however, detected between individual trials using otilonium and cimetropium and, although this was not the case when studies of pinaverium were pooled, the number of patients was small. Hyoscine seemed to have the best evidence for an individual compound, which was studied in over 400 patients. No statistically significant heterogeneity was detected, and the number need to treat was 3.5. It would seem reasonable for general practitioners who choose antispasmodics to use hyoscine as first line treatment and to consider other antispasmodics when necessary.

Peppermint oil was also superior to placebo, although significant heterogeneity was detected between studies, and only four trials were identified including fewer than 400 patients. Three of these trials scored more than 4 on the Jadad scale,[w30-w32] but the treatment effect was similar when only these studies were included in the meta-analysis, and heterogeneity between studies was no longer detected.

Limitations of this study arise from the quality of the included studies (mostly moderate to good, according to the Jadad scale). None of the included trials reported the method of allocation concealment, however, and therefore the numbers needed to treat may have been overestimated. Most trials were done before the Rome committee published their recommendations for the design of randomised controlled trials of therapies in functional gastrointestinal disorders.[17] Only five of the included studies used the Rome criteria,[w1 w16 w18 w31 w32] although only nine were published after the first Rome classification was proposed in 1990,[w1 w4 w16 w18 w26 w28 w30-w32] and only two used a validated outcome measure to define improvement in symptoms.[w18 w32] However, many of the included trials met some of the other suggested methodological criteria, and we preferentially extracted patient reported improvement in symptoms or abdominal pain whenever this was allowed by trial reporting. Blinding of patients may not have been entirely successful owing to differences in consistency and texture between fibre and placebo, adverse events with antispasmodics, and the smell and taste of peppermint oil. The pooling of data from trials to give an overall treatment effect and a number needed to treat could be criticised owing to differences in the methodology of individual studies. We carried out sensitivity analyses and in all cases identified potential reasons for heterogeneity between studies, while still showing a significant treatment effect for most of the treatments.

Current guidelines for the management of irritable bowel syndrome are equivocal or conflicting,[1 2 18 19] but most have been informed by systematic reviews, which are potentially flawed (see bmj.com). In the UK, guidelines from both the National Institute for Health and Clinical Excellence and the British Society of Gastroenterology provide similar advice.[1 2] Antispasmodics are recommended as first line treatment, and if fibre supplementation is required then soluble fibres such as ispaghula should be used. Neither guideline mentions peppermint oil.

This systematic review and meta-analysis shows that ispaghula, antispasmodics (particularly hyoscine), and peppermint oil are all effective treatments for irritable bowel syndrome. Many are safe and available over the counter but, with the advent of newer more expensive drugs, are often overlooked.

This study was done to inform the American College of Gastroenterology monograph on irritable bowel syndrome. We thank William Chey, Lawrence Brandt, Phillip Schoenfeld, and Edgar Achkar for their contributions to the discussion concerning the role of fibre, antispasmodics, and peppermint oil in the treatment of irritable bowel syndrome; and Premysl Bercik, Peter Bytzer, and Heidi Krall for assisting us with the translation of foreign language articles.

**Contributors:** See bmj.com.

**Funding:** This study was funded by the American College of Gastroenterology. The authors' work was independent of the funders.

**Competing interests:** NJT has received consultancy fees from Procter and Gamble, Lexicon Genetics, Astellas Pharma US, Pharma Frontiers, Callisto Pharmaceuticals, AstraZeneca, Addex Pharma, Ferring Pharma, Salix, M GI Pharma, McNeil Consumer, Microbia, Dynogen, Conexus, Novartis, and Metabolic Pharmaceuticals, and has received research support from Novartis, Takeda, GlaxoSmithKline, Dynogen, and Tioga. EMMQ has received consultant's and speaker's bureau fees from Nycomed, Boehringer Ingelheim, Procter and Gamble, Reckitt Benckiser, and Prometheus, and holds equity in Alimentary Health. PM holds a chair at McMaster University partly funded by an unrestricted donation by AstraZeneca, and has received consultant's and speaker's bureau fees from AstraZeneca, AxCan Pharma, Nycomed, and Johnson and Johnson.

**Ethical approval:** Not required.

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

1 National Institute for Health and Clinical Excellence. Irritable bowel syndrome in adults: diagnosis and management of irritable bowel syndrome in primary care. 2008. www.nice.org.uk/nicemedia/pdf/IBSFullGuideline.pdf.
2 Spiller R, Aziz Q, Creed FEA, Houghton L, Hungin P, Jones R, et al. Guidelines on the irritable bowel syndrome: mechanisms and practical management. *Gut* 2007;56:1770-98.
3 Harvey RF, Pomare EW, Heaton KW. Effects of increased dietary fibre on intestinal transit. *Lancet* 1973;301:1278-80.
4 Hills JM, Aaronson PI. The mechanism of action of peppermint oil on gastrointestinal smooth muscle. *Gastroenterology* 1991;101:55-65.
5 Jaliwala J, Imperiale TF, Kroenke K. Pharmacologic treatment of the irritable bowel syndrome: a systematic review of randomized, controlled trials. *Ann Intern Med* 2000;133:136-47.
6 Brandt LJ, Bjorkman D, Fennerty MB, Locke III GR, Olden K, Peterson W, et al. Systematic review on the management of irritable bowel syndrome in North America. *Am J Gastroenterol* 2002;97(suppl):S7-26.

7 Tack J, Fried M, Houghton LA, Spicak J, Fisher G. Systematic review: the efficacy of treatments for irritable bowel syndrome—a European perspective. *Aliment Pharmacol Ther* 2006;24:183-205.
8 Poynard T, Regimbeau C, Benhamou Y. Meta-analysis of smooth muscle relaxants in the treatment of irritable bowel syndrome. *Aliment Pharmacol Ther* 2001;15:355-61.
9 Lesbros-Pantoflickova D, Michetti P, Fried M, Beglinger C, Blum A. Meta-analysis: the treatment of irritable bowel syndrome. *Aliment Pharmacol Ther* 2004;20:1253-69.
10 Quartero AO, Meineche-Schmidt V, Muris J, Rubin G, de Wit N. Bulking agents, antispasmodic and antidepressant medication for the treatment of irritable bowel syndrome. *Cochrane Database Syst Rev* 2005;Apr 18;(2):CD003460.
11 Pittler MH, Ernst E. Peppermint oil for irritable bowel syndrome: a critical review and meta-analysis. *Am J Gastroenterol* 1998;93:1131-5.
12 Bijkerk CJ, Muris JWM, Knottnerus JA, Hoes AW, de Wit NJ. Systematic review: the role of different types of fibre in the treatment of irritable bowel syndrome. *Aliment Pharmacol Ther* 2004;19:245-51.
13 Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Control Clin Trials* 1996;17:1-12.
14 DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986;7:177-88.
15 Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003;327:557-60.
16 Egger M, Davey-Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997;315:629-34.
17 Veldhuyzen Van Zanten SJ, Talley NJ, Bytzer P, Klein KB, Whorwell PJ, Zinsmeister AR. Design of treatment trials for functional gastrointestinal disorders. *Gut* 1999;45(suppl 2):II69-77.
18 Drossman DA, Camilleri M, Mayer EA, Whitehead WE. AGA technical review on irritable bowel syndrome. *Gastroenterology* 2002;123:2108-31.
19 American College of Gastroenterology Functional Gastrointestinal Disorders Task Force. Evidence-based position statement on the management of irritable bowel syndrome in North America. *Am J Gastroenterol* 2002;97(suppl 2):S2-5.


**Accepted:** 24 September 2008

# Hypersensitivity reactions to human papillomavirus vaccine in Australian schoolgirls: retrospective cohort study

Liew Woei Kang,[1,2] Nigel Crawford,[3,4] Mimi L K Tang,[1,5] Jim Buttery,[3,4] Jenny Royle,[3] Michael Gold,[6] Christine Ziegler,[6] Patrick Quinn,[6] Sonja Elia,[3] Sharon Choo[1]

[1]Department of Allergy and Immunology, Royal Children's Hospital, Flemington Road, Parkville, Victoria 3052, Australia
[2]Paediatric Allergy, Immunology and Rheumatology, Department of Paediatric Medicine, KK Women's and Children's Hospital, Singapore
[3]Department of General Medicine, Royal Children's Hospital, Melbourne, Australia
[4]NHMRC Centre for Clinical Research Excellence in Child and Adolescent Immunisation, Surveillance of Adverse Events Following Vaccination in the Community, Murdoch Children's Research Institute, Department of Paediatrics, University of Melbourne, Melbourne, Australia
[5]Murdoch Children's Research Institute, Royal Children's Hospital, Melbourne, and Department of Paediatrics, University of Melbourne, Australia
[6]Department of Allergy and Immunology, Women's and Children's Hospital, Adelaide, Australia

Correspondence to: S Choo
sharon.choo@rch.org.au

Cite this as: *BMJ* 2008;337:a2642
doi:10.1136/bmj.a2642

This article is an abridged version of a paper that was published on bmj.com. Cite this article as: *BMJ* 2008;337:a2642


## ABSTRACT

**Objective** To describe the outcomes of clinical evaluation, skin testing, and vaccine challenge in adolescent schoolgirls with suspected hypersensitivity to the quadrivalent human papillomavirus vaccine introduced in Australian schools in 2007.

**Design** Retrospective cohort study.

**Setting** Two tertiary paediatric allergy centres in Victoria and South Australia, Australia.

**Participants** 35 schoolgirls aged 12 to 18.9 years with suspected hypersensitivity reactions to the quadrivalent human papillomavirus vaccine.

**Main outcome measures** Clinical review and skin prick and intradermal testing with the quadrivalent vaccine and subsequent challenge with the vaccine.

**Results** 35 schoolgirls with suspected hypersensitivity to the quadrivalent human papillomavirus vaccine were notified to the specialised immunisation services in 2007, after more than 380 000 doses had been administered in schools. Of these 35 schoolgirls, 25 agreed to further evaluation. Twenty three (92%) experienced reactions after the first dose. Thirteen (52%) experienced urticaria or angio-oedema, and of these, two experienced anaphylaxis. Thirteen had generalised rash, one with angio-oedema. The median time to reaction was 90 minutes. Nineteen (76%) underwent skin testing with the quadrivalent vaccine: all were skin prick test negative and one was intradermal test positive. Eighteen (72%) were subsequently challenged with the quadrivalent vaccine and three (12%) elected to receive the bivalent vaccine. Seventeen tolerated the challenge and one reported limited urticaria four hours after the vaccine had been administered. Only three of the 25 schoolgirls were found to have probable hypersensitivity to the quadrivalent vaccine.

**Conclusion** True hypersensitivity to the quadrivalent human papillomavirus vaccine in Australian schoolgirls was uncommon and most tolerated subsequent doses.

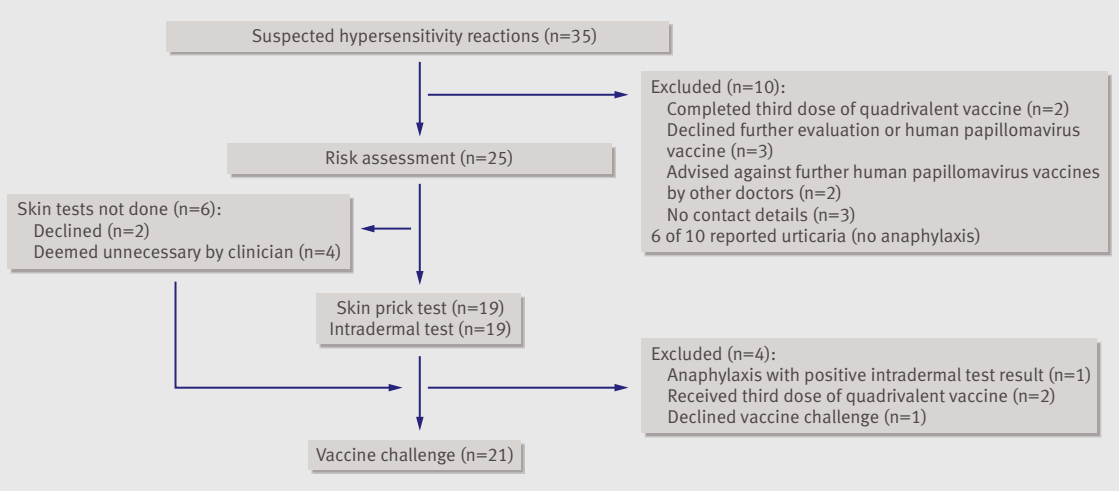

## INTRODUCTION

A quadrivalent human papillomavirus vaccine (Gardasil; Merck, NJ, USA) was included in the Australian national immunisation programme in April 2007 for females aged 12-26 years. Adolescent schoolgirls received the vaccine in a secondary school vaccination programme and reports of vaccine related adverse events soon followed.[1] Constituents of the quadrivalent vaccine, such as aluminium salts,[2,3] polysorbate 80,[4] and yeast,[5] have been associated with hypersensitivity reactions. The vaccine also shares constituents with other vaccines, such as hepatitis B (H-B-Vax II; Merck, NJ, USA) and diphtheria, tetanus, and pertussis (Boostrix; GlaxoSmithKline, Rixensart, Belgium), which are given to Australian adolescents at age 13 and 15 years, respectively. A bivalent human papillomavirus vaccine (Cervarix; GlaxoSmithKline, Rixensart, Belgium) lacks these constituents and may be an alternative for patients with hypersensitivity to the quadrivalent vaccine (see bmj.com).

We describe the outcomes of clinical evaluation, skin testing, and vaccine challenge in Australian adolescent schoolgirls with suspected hypersensitivity to the quadrivalent human papillomavirus vaccine.

## METHODS

In the Australian states of Victoria and South Australia, specialised immunisation services are notified of




**Flow chart of clinical evaluation through trial**

reported vaccine related adverse events. Adolescent schoolgirls with suspected hypersensitivity reactions to the quadrivalent human papillomavirus vaccine, including urticaria, generalised rash, angio-oedema, or anaphylaxis, were referred to tertiary paediatric allergy centres for further evaluation and are included in this retrospective cohort study. We include only girls who received the vaccine in school and not those who may have received the vaccine elsewhere. A detailed history of the reaction was obtained, including previous doses of the quadrivalent vaccine, concomitant vaccines, and time and severity of reaction. We also recorded any history of atopic disease, recurrent urticaria, or drug or vaccine related adverse reactions.

Skin prick and intradermal tests were carried out with 1:10 dilutions of both the quadrivalent and the bivalent human papillomavirus vaccines and 100 mg/ml polysorbate 80 (Tween 80; Merck, Darmstadt, Germany).[6] We used histamine and normal saline as positive and negative controls. Additional skin prick tests to other potential allergens were done as guided by clinical history. We measured skin wheals 15 and 20 minutes after skin prick and intradermal testing, respectively, and considered diameters of 3 mm or more above the saline control as a positive result.

Vaccine challenges were administered intramuscularly under medical supervision. All the girls were offered challenge with the quadrivalent vaccine unless there was previous anaphylaxis or a positive skin test result to the vaccine. A 0.1 ml dose was followed 30 minutes later by a 0.4 ml dose. The bivalent vaccine (0.5 ml) was given if requested by the recipient. We followed up the schoolgirls by telephone one week after vaccination and recorded any adverse events. Further vaccinations were planned for those who tolerated the challenge, to complete the three dose schedule.

## RESULTS

Thirty five schoolgirls with suspected hypersensitivity to the quadrivalent vaccine were reported in 2007, after more than 380 000 vaccine doses had been administered in schools in Victoria and South Australia. Twenty five of these schoolgirls (71%) agreed to undergo further evaluation and were reviewed between August 2007 and February 2008, at a median of 5.7 months (range 1.6-9.9 months) after the reaction (figure). The age of the schoolgirls, proportion with reactions to the first dose, and proportion with urticaria reactions were similar in those excluded and those evaluated. No cases of angio-oedema or anaphylaxis occurred in the excluded group (six in the evaluated group) and time to reaction was significantly longer (median 24 hours) and positively skewed than in the evaluated group.

The median time to reaction after vaccination in the evaluated group was 90 minutes. Thirteen of the 25 evaluated schoolgirls experienced urticaria or angio-oedema, and of these, two experienced anaphylaxis (table). Thirteen experienced generalised rash, one with angio-oedema.

Nineteen (76%) of the 25 evaluated schoolgirls received only the quadrivalent vaccine, whereas six had concomitant vaccines (table). Twenty three of the 25 (92%) reported reactions after the first dose of quadrivalent vaccine. Four of the 25 reported reactions after the second dose, and of these, three reported reactions after the first and the second doses. One patient reported a reaction after the third dose.

Fifteen (60%) of the 25 evaluated schoolgirls had a history of current atopic disease: allergic rhinitis in 12 (48%), asthma in eight (32%), atopic dermatitis in five (20%), allergic conjunctivitis in five (20%), and food allergy in three (12%). Two girls had recurrent urticaria and none had a history of hypersensitivity to yeast, drugs, or vaccines. Food, environmental allergens, and drug allergens that may have been associated with the vaccine related adverse event were excluded by a detailed history taking and, if clinically indicated, skin prick tests.

Nineteen of the 25 evaluated schoolgirls (76%) underwent skin prick testing to the quadrivalent

**Details of 25 girls reporting adverse reactions to the quadrivalent human papillomavirus vaccine**

| Vaccine category, dose, and concomitant vaccines | Suspected hypersensitivity reaction | Onset of reaction (min) | Skin prick test result | Intradermal test result | Vaccine challenge | Challenge reaction | Notes |
|---|---|---|---|---|---|---|---|
| Probable hypersensitivity (median 17.5 minutes): | | | | | | | |
| Third dose | Urticaria, angio-oedema, laryngeal oedema, tachypnoea, palpitations | 390 | Negative | Negative | NA | NA | Anaphylaxis after third dose |
| First (and second) dose | Urticaria (urticaria, angio-oedema, hoarse voice, laryngeal oedema) | 20 (15) | Negative | Positive | NA | NA | Anaphylaxis after second dose |
| First dose | Urticaria | 15 | Negative | Negative | Quadrivalent HPV vaccine | Reported limited urticaria four hours later | |
| Possible hypersensitivity (median 16 hours): | | | | | | | |
| Second dose | Urticaria | 960 | Negative | Negative | Elected not to proceed with challenge before evaluation | Elected not to proceed with challenge before evaluation | Hyperventilating after intradermal test. Reported non-specific limited rash several hours after intradermal test |
| Unlikely hypersensitivity (median 19 hours): | | | | | | | |
| First dose plus H-B-Vax II | Generalised rash, angio-oedema | 2 | Negative | Negative | Bivalent HPV vaccine | None | Hypersensitivity unlikely as did not receive quadrivalent vaccine |
| First dose plus H-B-Vax II | Generalised rash | 120 | Negative | Negative | Bivalent HPV vaccine | None | Hypersensitivity unlikely as did not receive quadrivalent vaccine |
| First dose | Generalised rash | 2160 | Negative | Negative | Quadrivalent HPV vaccine | Reported nausea, vomiting, and lethargy two days later | Hypersensitivity unlikely as reaction was different to previous reaction |
| First dose plus H-B-Vax II | Urticaria, angio-oedema | 2880 | Negative | Negative | Bivalent HPV vaccine | None | Hypersensitivity unlikely as did not receive quadrivalent vaccine |
| Not hypersensitivity (median 90 minutes): | | | | | | | |
| First dose plus Varilrix plus tetanus | Generalised rash | 1440 | Negative | Negative | Quadrivalent HPV vaccine | None | |
| First dose plus H-B-Vax II | Generalised rash (eczema) | 1440 | NA | NA | Quadrivalent HPV vaccine | None | Skin testing deemed unnecessary |
| First dose | Generalised rash | 1080 | NA | NA | Quadrivalent HPV vaccine | None | Skin testing deemed unnecessary |
| First dose | Generalised rash | 1080 | NA | NA | Quadrivalent HPV vaccine | None | Declined skin testing |
| First dose | Generalised rash | 180 | Negative | Negative | Quadrivalent HPV vaccine | None | |
| First dose plus Boostrix | Generalised rash | 720 | Negative | Negative | Quadrivalent HPV vaccine | None | |
| First dose | Urticaria | 2880 | Negative | Negative | Quadrivalent HPV vaccine | None | |
| First dose | Angio-oedema | 5 | Negative | Negative | Quadrivalent HPV vaccine | None | |
| First dose | Generalised rash | 90 | Negative | Negative | Quadrivalent HPV vaccine | None | |
| First dose | Generalised rash | 1440 | NA | NA | Quadrivalent HPV vaccine | None | Declined skin testing |
| First dose | Angio-oedema | 1440 | Negative | Negative | Quadrivalent HPV vaccine | None | |
| First dose | Urticaria | 15 | Negative | Negative | Quadrivalent HPV vaccine | None | |
| First (and second) dose | Urticaria (urticaria) | 30 (20) | Negative | Negative | Quadrivalent HPV vaccine | None | Twin of schoolgirl |
| Third dose | Urticaria | 10 | Negative | Negative | NA | NA | Twin of schoolgirl |
| First dose | Generalised rash, tachypnoea | 20 | Negative | Negative | Quadrivalent HPV vaccine | None | Thought to hyperventilate after first dose |
| First dose | Generalised rash | 30 | Negative | Negative | Quadrivalent HPV vaccine | None | |
| First (and second) dose | Urticaria (urticaria) | 10 (10) | Negative | Negative | Quadrivalent HPV vaccine | None | |

HPV=human papillomavirus; NA=not applicable; H-B-Vax II=vaccine against hepatitis B (Merck); Varilix=vaccine against varicella (GlaxoSmithKline); Boostrix=vaccine against diphtheria, tetanus, and pertussis (GlaxoSmithKline).

vaccine, polysorbate 80, and bivalent vaccine. All the results were negative. The 19 schoolgirls underwent intradermal testing, of which one (table) had a positive result to the quadrivalent vaccine and negative results to polysorbate 80 and the bivalent vaccine. One schoolgirl experienced hyperventilation during intradermal testing (table) and reported a limited non-specific rash several hours later. She declined further vaccination against human papillomavirus.

Challenge with the quadrivalent vaccine was carried out in 18 (72%) of the 25 evaluated schoolgirls. Three of the seven schoolgirls who were not challenged with the

quadrivalent vaccine elected to receive the bivalent vaccine as they had concerns about the quadrivalent vaccine despite a negative skin test result. Vaccine challenges were not done in the two schoolgirls who had completed the three doses of the schedule or the one girl who declined further vaccination, and challenge was contraindicated in one girl who had anaphylaxis and a positive skin test result to the quadrivalent vaccine.

Seventeen of the 18 schoolgirls challenged with the quadrivalent vaccine and all three challenged with the bivalent vaccine remained well one week after vaccination. One schoolgirl reported limited urticaria over the limbs and trunk four hours after challenge with the quadrivalent vaccine (table). Supervised challenge with the bivalent vaccine for her third dose of human papillomavirus vaccine was well tolerated.

The 25 evaluated schoolgirls were classified into one of four categories (table): probable hypersensitivity—those with anaphylaxis, a positive skin test result for the quadrivalent vaccine, or reproducible reactions to challenge with the quadrivalent vaccine; possible hypersensitivity—those with reactions to skin testing who were not challenged with the quadrivalent vaccine; unlikely hypersensitivity—those with negative skin test results to the quadrivalent vaccine who were not challenged with the quadrivalent vaccine, or were challenged with the quadrivalent vaccine but did not experience a reproducible reaction; and not hypersensitivity—those with negative skin test results for the quadrivalent vaccine and no adverse reaction to subsequent challenge with the quadrivalent vaccine.

Schoolgirls in the probable hypersensitivity group were more likely to present with urticaria than those in the unlikely hypersensitivity group (likelihood ratio 9.0) and not hypersensitivity group (10.2), and had a median time to reaction of 17.5 minutes compared with 19 hours in the unlikely hypersensitivity group and 90 minutes in the not hypersensitivity group (table). Other clinical features, including number of doses of the quadrivalent vaccine, concomitant vaccines, recurrence of reactions to the quadrivalent vaccine, and current atopic disease or recurrent urticaria, did not predict hypersensitivity to the quadrivalent vaccine.

## DISCUSSION

We evaluated suspected hypersensitivity in adolescent females immunised with a human papillomavirus vaccine in Australian schools. Only three of the 25 evaluated schoolgirls had probable hypersensitivity to the quadrivalent human papillomavirus vaccine after 380 000 doses had been administered in schools. Seventeen of the 18 girls subsequently challenged with the quadrivalent vaccine tolerated revaccination. Our data suggest that true hypersensitivity to the quadrivalent vaccine is uncommon and that suspected hypersensitivity reactions such as urticaria are often idiosyncratic and not usually a contraindication to further vaccinations. Studies of other vaccines have found that most reactions after immunisation are not due to hypersensitivity and revaccination is usually well tolerated.[7-9]

### WHAT IS ALREADY KNOWN ON THIS TOPIC

Hypersensitivity reactions to vaccines are uncommon

### WHAT THIS STUDY ADDS

True hypersensitivity to the quadrivalent human papillomavirus vaccine is uncommon and most females tolerate subsequent doses

Although we excluded 10 of 35 schoolgirls with suspected hypersensitivity to the quadrivalent vaccine from our evaluation, reactions in the excluded group were mostly mild and delayed in presentation, suggesting that we did not miss any important cases of suspected hypersensitivity to the quadrivalent vaccine. All reported cases of anaphylaxis were evaluated.

Time to anaphylaxis was 15 minutes in one girl and 6.5 hours in another. As anaphylaxis after childhood vaccinations usually occurs within one hour,[10 11] 6.5 hours is beyond any standard observation period after immunisation. Consistent with the delayed presentation, one of the girls had no evidence of IgE mediated hypersensitivity to the quadrivalent vaccine and we postulate her reaction was mediated by IgG or complement, or both. As she was not rechallenged with the quadrivalent vaccine, however, hypersensitivity was not confirmed.

One of the girls had a positive intradermal test result to the quadrivalent vaccine that was consistent with IgE mediated hypersensitivity. We were unable to determine whether her reaction was due to the recombinant viral-like particles or other constituents of the vaccine such as aluminium hydroxyphosphate sulphate. As she had no history of reactions to yeast, and skin testing for polysorbate 80 gave a negative result, IgE mediated hypersensitivity to these components was unlikely. For females with probable hypersensitivity to the quadrivalent vaccine, immunoblot analysis and measurement of specific IgG and IgE to the individual vaccine components would provide further information.

Our study describes two cases of anaphylaxis after 380 000 doses of the quadrivalent vaccine had been administered. Although we have a passive surveillance system for reporting vaccine related adverse events in Australia, the quadrivalent human papillomavirus vaccine is a new vaccine and there is a high level of awareness of the importance of reporting adverse events in the school immunisation programme. One study estimated that if 80% of eligible US adolescent females were to receive a saline injection according to the vaccination schedule for human papillomavirus, 3 per 100 000 adolescents would require emergency care for asthma or allergy within 24 hours of vaccination.[12] As allergic symptoms are common, studies of adverse events to the quadrivalent vaccine should take these "baseline" rates into consideration. An Australian human papillomavirus vaccination programme

register (www.hpvregister.org.au/), established in August 2008, will facilitate more accurate determination of rates of hypersensitivity reactions not possible from current data sources.

In conclusion, suspected hypersensitivity reactions to the human papillomavirus quadrivalent vaccine require further evaluation to exclude IgE mediated reactions. Most females with suspected hypersensitivity to this vaccine tolerate revaccination. Our clinical recommendation is that females with suspected hypersensitivity to the quadrivalent vaccine should be evaluated before receiving more doses, and any challenges with the same vaccine should be carried out in a supervised setting. Further studies are required to investigate the mechanisms of hypersensitivity to this vaccine.

We thank the allergy and immunology department and immunisation nurse consultants from the Royal Children's Hospital and the South Australian Immunisation Coordination Unit for their assistance. NC acknowledges support from a National health and Medical Research Council PhD postgraduate public health research scholarship.
**Contributors:** See bmj.com.
**Funding:** None.
**Competing interests:** MT is chairperson of an incorporated association Asia Pacific Immunoglobulins in Immunology Expert Group that is supported by an unrestricted grant from CSL. JB has served on an advisory board for GSK and serves on a data safety monitoring board for CSL. MCRI receives reimbursement from both GSK and CSL for JB's attendance at advisory board and scientific meetings.
**Ethical approval:** Not required.
**Provenance and peer review:** Not commissioned; externally peer reviewed.

1 Tanne JH. Questions over human papillomavirus vaccine in US and Australia. *BMJ* 2007;334:1182-3.
2 Cox NH, Moss C, Forsyth A. Cutaneous reactions to aluminium in vaccines: an avoidable problem. *Lancet* 1988;2:43.
3 Baylor NW, Egan W, Richman P. Aluminium salts in vaccines—US perspective. *Vaccine* 2002;(Suppl 3):S18-23.
4 Shelley WB, Talanin M, Shelley ED. Polysorbate 80 hypersensitivity. *Lancet* 1995;345:1312-3.
5 Brightman CA, Scadding GK, Dumbreck LA, Latchman Y, Brostoff J. Yeast-derived hepatitis B vaccine and yeast hypersensitivity. *Lancet* 1989;22:903.
6 Wood RA, Setse R, Halsey N. Irritant skin test reactions to common vaccines. *J Allergy Clin Immunol* 2007;120:478-81.
7 Ponvert C, Scheinmann P. Vaccine allergy and pseudo-allergy. *Eur J Dermatol* 2003;13:10-5.
8 Andrews RM, Kempe AE, Herceg A. Vaccinating children with a history of serious reactions after vaccination or of egg allergy. *Med J Austr* 1998;168:491.
9 Gold M, Goodwin H, Botham S, Burgess M, Nash M, Kempe A. Revaccination of 421 children with a past history of an adverse reaction in a special immunization service. *Arch Dis Child* 2000;83:128-31.
10 Nakayama T, Aizawa C, Kuno-Sakai H. A clinical analysis of gelatin allergy and determination of its causal relationship to the previous administration of gelatin-containing acellular pertussis vaccine combined with diphtheria and tetanus toxoid. *J Allergy Clin Immunol* 1993;103:321-5.
11 Patja A, Davidkin I, Kurki T, Kallio MJ, Valle M, Peltola H. Serious adverse events after measles-mumps-rubella vaccination during a fourteen year prospective follow-up. *Pediatr Infect Dis J* 2000;19:1127-34.
12 Siegrist CA, Lewis EM, Eskola J, Evans SJ, Black SB. Human papilloma virus immunisation in adolescent and young adults: a cohort study to illustrate what events might be mistaken for adverse reactions. *Pediatr Infect Dis J* 2007;26:979-84.

**Accepted:** 1 October 2008

# Effects of algorithm for diagnosis of active labour: cluster randomised trial

Helen Cheyne,[1] Vanora Hundley,[1] Dawn Dowding,[2,3] J Martin Bland,[2] Paul McNamee,[4] Ian Greer,[3] Maggie Styles,[5] Carol A Barnett,[6] Graham Scotland,[4] Catherine Niven[1]

[1]Nursing Midwifery and Allied Health Professions Research Unit, University of Stirling, Stirling FK9 4LA
[2]Department of Health Sciences, University of York, York YO10 5DD
[3]Hull York Medical School, University of York
[4]Health Economics Research Unit, Institute of Applied Health Sciences, University of Aberdeen, Aberdeen AB25 2ZD
[5]Department of Nursing and Midwifery, University of Stirling
[6]NHS Tayside, Kings Cross Hospital, Dundee DD3 8EA
**Correspondence to:** H Cheyne
h.l.cheyne@stir.ac.uk

Cite this as: *BMJ* 2008;337:a2396
doi:10.1136/bmj.a2396

## ABSTRACT

**Objective** To compare the effectiveness of an algorithm for diagnosis of active labour in primiparous women with standard care in terms of maternal and neonatal outcomes.

**Design** Cluster randomised trial.

**Setting** Maternity units in Scotland with at least 800 annual births.

**Participants** 4503 women giving birth for the first time, in 14 maternity units. Seven experimental clusters collected data from a baseline sample of 1029 women and a post-implementation sample of 896 women. The seven control clusters had a baseline sample of 1291 women and a post-implementation sample of 1287 women.

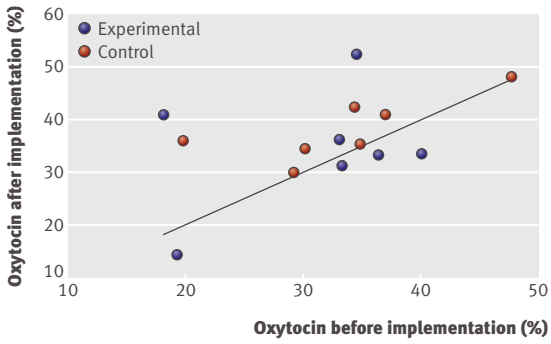
**Intervention** Use of an algorithm by midwives to assist their diagnosis of active labour, compared with standard care.

**Main outcomes** Primary outcome: use of oxytocin for augmentation of labour. Secondary outcomes: medical interventions in labour, admission management, and birth outcome.

**Results** No significant difference was found between groups in percentage use of oxytocin for augmentation of labour (experimental minus control, difference=0.3, 95% confidence interval −9.2 to 9.8; P=0.9) or in the use of medical interventions in labour. Women in the algorithm group were more likely to be discharged from the labour suite after their first labour assessment (difference= −19.2, −29.9 to −8.6; P=0.002) and to have more pre-labour admissions (0.29, 0.04 to 0.55; P=0.03).

**Conclusions** Use of an algorithm to assist midwives with the diagnosis of active labour in primiparous women did not result in a reduction in oxytocin use or in medical intervention in spontaneous labour. Significantly more women in the experimental group were discharged home after their first labour ward assessment.

**Trial registration** Current Controlled Trials ISRCTN00522952.




Oxytocin use before and after trial implementation

## INTRODUCTION

Between 30% and 45% of women admitted to labour wards in the United Kingdom and other developed countries are subsequently found not to be in labour,[1-3] and several studies have shown that women admitted in the latent phase or not yet in labour are more likely to receive medical intervention than those admitted in active labour.[3-6] Only one randomised controlled trial has tested the efficacy of adhering to strict criteria for diagnosis of labour.[7] This reported a reduction in the use of augmentation with oxytocin and analgesics. The study was small, however, and a Cochrane review concluded that a multicentre randomised controlled trial was needed to determine the risks and benefits of using explicit criteria to diagnose active labour.[8]

We hypothesised that improving the diagnosis of labour in primiparous women through the use of an algorithm would result in a reduction in the use of oxytocin for augmentation of labour and other labour interventions compared with standard care. We chose a cluster randomised trial for this purpose because the algorithm was aimed at the clinical practice of midwives. We could not use individual randomisation of women or midwives because of the risk of contamination between groups.

## METHODS

### Recruitment and randomisation

The trial took place between April 2005 and June 2007. Three levels of participation existed: the unit of randomisation was the maternity unit, midwives were participants at the level of the intervention, and we measured trial outcomes for women receiving maternity care.

Maternity units in Scotland with at least 800 annual births were eligible. We used minimisation to allocate maternity units to experimental or control groups. After random allocation of the first maternity unit, we purposively allocated clusters in order to maximise balance between groups. We chose presence or absence of an on-site midwife managed birth unit as the balancing variable. In the experimental group, we invited midwives who admitted women in labour to participate in the study. We provided workshops and individual contacts for each midwife. We made minimum contact with midwives in control units.

We considered women to be eligible for recruitment if they attended a participating maternity unit for assessment of labour and were primiparous, at term, and assessed as low risk (see bmj.com). We used the same eligibility criteria in both experimental and control groups. In order to reduce confounding variables, we excluded multiparous women from the trial.

Women in both arms of the trial were given information at a clinic visit between 34 and 36 weeks' gestation. Women in the experimental group were asked for consent on admission to the labour suite. Women in the control group were asked for consent in the postnatal wards. We asked midwives to recruit women who would have been eligible for the trial when they first presented for labour admission, regardless of subsequent labour outcome.

### Intervention

The algorithm comprised three levels: level one confirmed the woman's eligibility for involvement in the study; level two prompted a general physical assessment (for example, temperature, pulse, and blood pressure); and level three presented, in a stepwise fashion, key informational cues needed for diagnosis of labour. Active labour was diagnosed when painful, regular, moderate or strong uterine contractions were present, as well as at least one of the following cues: cervix effacing and at least 3 cm dilated, spontaneous rupture of membranes, or "show." We pre-tested the algorithm with three samples of midwives and found it to have good face validity and content validity and a high level of inter-rater reliability.[9] Subsequently, we did a feasibility study in two maternity units.[10] During the trial, women in both groups contacted the hospital by telephone when they thought that they were in labour and then attended for assessment in a similar way.

### Trial groups

In the experimental group, we asked midwives to use the algorithm during the admission assessment of women to assist in the diagnosis of active labour, recording their judgment on the algorithm. Women identified as not yet in active labour were encouraged to return home or were admitted to an antenatal area, depending on local maternity unit policy.

Eligible women who attended for admission assessment in the control group units received normal care. This comprised admission assessment by midwives using clinical judgment alone.

Owing to the Scotland-wide, multisite nature of the trial, dictating where the woman should go after the admission assessment was not appropriate. Some units served remote areas where women may have had difficulty returning home. Therefore, after the admission assessment, women in both groups received standard care for their maternity unit.

This article is an abridged version of a paper that was published on bmj.com. Cite this article as: *BMJ* 2008;337:a2396

### Outcomes

The primary outcome was use of oxytocin (any dose) for augmentation of labour. This is the principal treatment (and key marker) of slow progress in labour. Furthermore, primiparous women who receive oxytocin have a reduced likelihood of an unassisted vaginal delivery.[11] Secondary outcomes were interventions in labour (artificial rupture of membranes, vaginal examination, continuous electronic fetal monitoring, and use of analgesia), admission management (number of admissions before labour, time spent in labour ward, and duration of active labour), and labour outcomes (mode of delivery, intrapartum complications, neonatal outcome, and unplanned out of hospital births).

### Data collection

We collected baseline data retrospectively from the case records of 200 women who gave birth in each unit before implementation of the study and who fitted the trial eligibility criteria on first labour ward assessment. These data were anonymised. We collected data for the post-implementation sample from case records after delivery. From the power calculation we needed at least 12 hospitals with 200 women observed before and 200 after the trial implementation point (see bmj.com).

### Analysis

Data analysis was appropriate for a cluster randomised trial and accounted for clustering of observations within maternity units. The primary analysis used multiple regression of maternity unit level data adjusted for baseline. This meant that for each outcome we calculated a summary statistic (the mean or proportion) for each cluster, at baseline and after study implementation. In each case, the baseline value was the covariate. This provided a confidence interval and test of significance for the difference in proportions of women receiving oxytocin. We did other analyses at the level of the individual woman or using data aggregated to cluster level as appropriate.

## RESULTS

Of the 15 maternity units in Scotland that were available to participate in the trial, 14 agreed and were allocated to experimental or control groups. Two units in each group had an on-site midwife managed birth unit. Overall, 80% of midwives consented to participate (unit range 57-100%). Baseline data were collected for 1029 women in the experimental group and 1291 women in the control group.

The steering group did a routine review of study procedures after the first few months of data collection and recommended a protocol change to minimise any potential risk of selection bias in the control group. Midwives might have been reluctant to approach women who had experienced complications of labour or negative outcomes. We asked midwives in the control group to continue to recruit women as planned up to a sample of 100 women, needed for a health economics evaluation questionnaire. We asked them to then go back to the recruitment start date and review the case records of women who had given birth from that time and who had been eligible but not recruited. Anonymous study outcome data were collected from consecutive cases up to the total target sample of 200 cases. This resulted in near complete data collection in control units. The second sample, recruited after implementation of the study, comprised 892 women in the experimental group and 1279 women in the control group (see bmj.com).

The table shows cluster level data for the primary outcome. We found no significant difference in the percentage of oxytocin use attributable to the application of the algorithm (difference=0.3, 95% confidence interval −9.2 to 9.8; P=0.9) (figure).

We found no significant difference between groups for any of the labour interventions considered (see bmj.com). Significantly more women in the control group had only one admission; 398 women (44.6%) in the intervention group had only one admission compared with 795 women (62.6%) in the control group (difference −19.2 (−29.9 to −8.6) (P=0.002)). This means that women in the control group were more likely to remain in the labour suite after their first admission assessment until delivery. In contrast, women in the experimental group were significantly more likely to have several admissions and discharges before their eventual admission leading to delivery.

We found no significant difference between groups for duration of active labour, time from the first labour assessment to delivery, or time from final admission to labour suite until delivery (see bmj.com). We found no significant difference in mode of delivery between study groups. Overall, 45% (n=2028) of women had at least one intrapartum complication. We found no significant difference in maternal complications between groups. Overall, 67 babies were admitted to

**Oxytocin use "before" and "after" study implementation**

| Unit No | Total women per cluster (before and after) | Oxytocin use—% (No) | |
|---|---|---|---|
| | | At baseline | After study implementation |
| **Experimental (n=1921)** | | | |
| 2 | 398 | 18 (36/198) | 41 (82/200) |
| 4 | 112 | 33 (16/48) | 31 (20/64) |
| 7 | 139 | 19 (16/83) | 14 (8/56) |
| 9 | 401 | 40 (81/202) | 34 (67/199) |
| 10 | 260 | 37 (73/200) | 33 (20/60) |
| 12 | 362 | 35 (56/162) | 53 (105/200) |
| 14 | 249 | 33 (45/136) | 36 (41/113) |
| **Control (n=2570)** | | | |
| 1 | 401 | 35 (70/201) | 36 (71/200) |
| 3 | 398 | 48 (95/199) | 48 (96/199) |
| 5 | 399 | 29 (58/199) | 30 (60/200) |
| 6 | 400 | 37 (74/200) | 41 (82/200) |
| 8 | 399 | 30 (60/199) | 35 (69/200) |
| 11 | 397 | 20 (39/197) | 36 (72/200) |
| 13 | 176 | 34 (33/96) | 43 (34/80) |

## WHAT IS ALREADY KNOWN ON THIS TOPIC

Up to 30% of women admitted to labour wards in the UK may not be in active labour

Women admitted to labour wards in the latent phase of labour are more likely to receive medical intervention than are those admitted in active labour

Previous studies have indicated that the introduction of explicit criteria for diagnosis of labour may reduce oxytocin use

## WHAT THIS STUDY ADDS

The introduction of an algorithm to assist with the diagnosis of labour did not reduce oxytocin use

Use of an algorithm for the diagnosis of labour increased the number of times women were admitted and subsequently discharged from hospital before finally being admitted for delivery

Increased rates of intervention in women admitted to labour suites early cannot be fully explained by failure of clinicians to distinguish between the latent and active phases of labour

the neonatal unit for more than 48 hours, but this did not differ significantly between groups.

## DISCUSSION

This trial, involving 14 maternity units and 4503 women, tested the effectiveness of an algorithm to assist midwives with the diagnosis of active labour in primiparous women. We found that use of the algorithm did not reduce the number of women who received oxytocin or other medical interventions compared with standard care. Women in the experimental group were more likely to be discharged home and subsequently have significantly more admissions before labour.

One other trial has specifically evaluated the efficacy of using strict diagnostic criteria.[7] This trial found that significantly fewer women received oxytocin to augment labour compared with no labour assessment, and less pain relief was used. Although both trials included similar diagnostic criteria, the interventions were not identical and the other study evaluated both diagnosis and management of early labour.[7] Also, the other study took place in one hospital with 209 women and was underpowered to test the effects of the intervention on several important maternal and neonatal outcomes.

Although our target sample of 400 in each maternity unit was not achieved at all sites, this deficit was partially offset by the recruitment of an additional two maternity units. The study had sufficient power to test the primary outcome (see bmj.com).

### Limitations of the trial

We could not accurately determine the number of eligible women in each maternity unit, and estimates were based on routinely collected data. We could not identify the number of women who would have been ineligible for medical or obstetric reasons, nor could we differentiate between women who were not eligible and those who were not approached for consent to data collection. In most of the units the proportion of eligible women not included was high and therefore selection bias could have occurred. The strength of the cluster design is that it avoids contamination between groups; however, the design is prone to selection bias, because consent to trial entry is given at cluster level but individuals can then decide whether to accept or refuse the trial intervention. Selection bias is also a common problem in trials of intrapartum care, in which difficulty in estimating numbers of potentially eligible participants and high losses to recruitment are often reported.[12] Although we could not recruit women in the control group in the same way as in the experimental group, we used the same trial entry criteria. We maximised the use of anonymised data after the protocol change, to minimise the potential for Hawthorne effects and reduce selection bias.

Studies of decision support suggest that it is the consistency of decision support tools, not the provision of new knowledge, that makes them effective.[13,14] However, the reluctance of healthcare professionals to use decision support has been widely reported.[15,16] In our trial, the rate of consent of midwives to use the algorithm varied between units from 57% to 100%. In most (although not all) units, this consent rate reflected the success or otherwise of subsequent data collection. We found evidence that using the algorithm did alter the midwives' judgments, as women in the experimental group were significantly more likely to be discharged after their first labour assessment than were women in the control group. However, these women quickly returned to the hospital, creating a "revolving door" effect. This implies that the observation from other studies of higher rates of intervention in women admitted to labour suites early cannot be fully explained by a failure of clinicians to distinguish between the latent and active phases of labour.

We thank the midwives and mothers who participated in or facilitated this study.

**Contributors:** See bmj.com.

Local principal investigators, responsible for the conduct of the trial at the individual sites: Angela Watt, Annette Lobo, Susan Stewart, Angela Cunningham, Joyce Linton, Carol Powrie, Gillian Morton, Cathy Harkins, Lynn Wojciechowska, Joan Milne, Eleanor Stenhouse, Lorraine Wilson, and Liz Terrace.

Local trial coordinators, who supported the trial locally and collected trial outcome data: Morag Grant, Anne Paterson, Sylvia Morrison, Gillian McMurray, Alison Hourston, Liz McMurchie, Lesley Darroch, Fiona Mundell, Caron Cruikshank, Liz Main, Carol Beatts, Anne Marie Brolly, Karen McIntosh, Bernie McStea, and Janie Cunning.

Sponsor: The University of Stirling was the study sponsor. The sponsor had no role in study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the article for publication.

**Funding:** This work was supported by the Scottish Executive Chief Scientist Office Health Service Research Committee (CZH/4/245). The research was independent of the funders.

**Competing interests:** None declared.

**Ethical approval:** The Multicentre Research Ethics Committee for Scotland B approved the study (05/MRE10/31). The local research ethics committees in each area granted site specific approval. All women gave informed consent.

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

1 Ball JA, Washbrook M. *Birthrate plus: a framework for workforce planning and decision making for midwifery services*. London: Books for Midwives Press, 1996.

2 Janssen P, Iker C, Carty E. Early labour assessment and support at home: a randomised controlled trial. *J Obstet Gynaecol Can* 2003;25:734-41.

3 Klein MC, Kelly A, Kaczorowski J, Grzybowski S. The effect of family physician timing of maternal admission on procedures in labour and maternal and infant morbidity. *J Obstet Gynaecol Can* 2003;26:641-5.
4 Hemminki E, Simukka R. The timing of hospital admission and progress of labour. *Eur J Obstet Gynecol Reprod Biol* 1986;22:85-94.
5 Holmes P, Oppenheimer LW, Wen SW. The relationship between cervical dilatation at initial presentation in labour and subsequent intervention. *BJOG* 2001;108:1120-4.
6 Jackson DJ, Lang JM, Ecker J, Swartz WH, Heeren T. Impact of collaborative management and early admission in labour on method of delivery. *J Obstet Gynecol Neonatal Nurs* 2003;32:147-57.
7 McNiven PS, Williams JI, Hodnett E, Kaufman K, Hannah ME. An early labor assessment program: a randomized, controlled trial. *Birth* 1998;25:5-10.
8 Lauzon L, Hodnett E. Labour assessment programs to delay admission to labour wards. *Cochrane Database Syst Rev* 2001;(3):CD000936.
9 Cheyne H, Dowding D, Hundley V, Aucott L, Styles M, Mollison J, et al. The developing and testing of an algorithm for diagnosis of active labour in primiparous women. *Midwifery* 2008;24:199-213.
10 Cheyne H, Hundley V, Dowding D, Mollison J, Greer IA. *A feasibility study for a cluster randomised trial to investigate the use of a decision aid for the diagnosis of active labour in term pregnancy*. Edinburgh: Scottish Executive Health Department Chief Scientist Office, 2004 (available from www.sehd.scot.nhs.uk/cso/Publications/ExecSumms/needs04.htm).
11 Green J, Baston H, Easton S, McCormick F. *Greater expectations? Summary report*. Leeds: Mother and Infant Research Unit, 2003.
12 Hundley V, Cheyne H. The trials and tribulations of intrapartum studies. *Midwifery* 2004;20:27-36.
13 Rosenthal GE, Mettler G, Pare S, Riegger M, Ward M, Landefeld CS. Diagnostic judgements of nurse practitioners providing primary gynaecologic care: a quantitative analysis. *J Gen Intern Med* 1992;7:304-11.
14 Dawes RM, Faust D, Meehl PE. Clinical versus actuarial judgement. In: Gilovich T, Griffin D, Kahneman D, eds. *Heuristics and biases: the psychology of intuitive judgement*. Cambridge: Cambridge University Press, 2002:716-29.
15 Thompson C, Cullum N, McCaughan D, Sheldon T, Raynor P. Nurses, information use, and clinical decision making—the real world potential for evidence-based decisions in nursing. *Evid Based Nurs* 2004;7:68-72.
16 McCaughan D, Thompson C, Cullum N, Sheldon T, Raynor P. Nurse practitioner and practice nurses' use of research information in clinical decision making: findings from an exploratory study. *Fam Pract* 2005;22:490-7.

**Accepted:** 9 September 2008

# Incidence of cardiovascular disease and cancer in advanced age: prospective cohort study

Jane A Driver,[1] Luc Djoussé,[1] Giancarlo Logroscino,[2] J Michael Gaziano,[1,3,4] Tobias Kurth[1,3,5]

**EDITORIAL by Strandberg**

[1]Division of Aging, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02215, USA
[2]Department of Neurology and Psychiatry, School of Medicine, University of Bari, Italy
[3]Preventive Medicine, Department of Medicine, Brigham and Women's Hospital
[4]Massachusetts Veterans Epidemiology Research Information Center, VA Boston Healthcare System, Boston, MA
[5]Department of Epidemiology, Harvard School of Public Health, Boston, MA
Correspondence to: J A Driver
jdriver@partners.org

Cite this as: *BMJ* 2008;337:a2467
doi:10.1136/bmj.a2467

This article is an abridged version of a paper that was published on bmj.com. Cite this article as: *BMJ* 2008;337:a2467

## ABSTRACT

**Objective** To investigate the influence of increasing age on the incidence and remaining lifetime risk of cardiovascular disease and cancer in a cohort of older men.

**Design** Prospective cohort study.

**Setting** United States.

**Participants** 22 048 male doctors aged 40-84 who were free of major disease in 1982.

**Main outcome measures** Incidence and remaining lifetime risk of major cardiovascular disease (myocardial infarction, stroke, and death from cardiovascular disease) and cancer.

**Results** 3252 major cardiovascular events and 5400 incident cancers were confirmed over 23 years of follow-up. The incidence of major cardiovascular disease continued to increase to age 100. Beginning at age 80, however, major cardiovascular disease was more likely to be diagnosed at death. The incidence of cancer peaked in those aged 80-89 and then declined. Cancers detected by screening accounted for most of the decline, whereas most cancers for which there was no screening continued to increase to age 100. Unadjusted cumulative incidence overestimated the risk of cardiovascular disease by 16% and cancer by 8.5%. The remaining lifetime risk of cancer at age 40 was 45.1% (95% confidence interval 43.8% to 46.3%) and at age 90 was 9.6% (7.2% to 11.9%). The remaining lifetime risk of major cardiovascular disease at age 40 was 34.8% (33.1% to 36.5%) and at age 90 was 16.7% (12.9% to 20.6%).

**Conclusions** In this prospective cohort of men, the incidence of new cardiovascular disease continued to increase after age 80 but was most often diagnosed at death. The decrease in incidence of cancer late in life seemed largely due to a decline in cancers usually detected by screening. These findings suggest that people aged 80 and older have a substantial amount of undiagnosed disease. The remaining lifetime risk of both diseases approached a plateau in the 10th decade. This may be due to decreased detection of disease and reporting of symptoms and increased resistance to disease in those who survive to old age. Accurate estimates of disease risk in an aging population require adjustment for competing risks of mortality.

## INTRODUCTION

Cardiovascular disease and cancer increase exponentially between ages 40 and 80, yet data on incidence in the ninth and 10th decades are sparse, particularly in men. We estimated the age specific incidence and remaining lifetime risk of these diseases up to age 100 in a prospective cohort of men with 23 years of follow-up.

## METHODS

The Physicians' Health Study is a completed randomised trial of aspirin and β carotene for the primary prevention of cardiovascular disease and cancer among 22 071 US male doctors.[12] At study entry in 1982, participants were aged between 40 and 84 and had no history of cardiovascular disease, cancer, or other serious illnesses. In total, 92.2% of the participants identified themselves as white. Baseline information on lifestyle variables and other risk factors for cardiovascular disease and cancer was self reported by questionnaire. Follow-up questionnaires in which patients reported new medical diagnoses and

procedures were sent twice in the first year and yearly thereafter. We used follow-up information to 30 March 2007.

Non-fatal cases of cancer and cardiovascular disease were self reported by participants and fatal cases by family members or next of kin. Reports of revascularisation procedures and new onset angina were also recorded. We obtained the medical records for all reported cancers and major cardiovascular events. Malignancies were confirmed by review of pathology reports. We used the World Health Organization criteria to confirm non-fatal myocardial infarction.[3] Non-fatal stroke was defined as a typical neurological deficit, sudden in onset and lasting more than 24 hours, and attributed to a cerebrovascular event. Death due to ischaemic heart disease or stroke was confirmed by convincing evidence based on information from medical records, death certificates, family members, or next of kin. For this analysis we used only confirmed events. Major cardiovascular disease was defined as non-fatal myocardial infarction, non-fatal stroke, or death from cardiovascular disease. For the end point all cardiovascular disease, we also considered self reported angina and revascularisation procedures. When more than one end point was reported, we used the first event to occur to define the onset of disease.

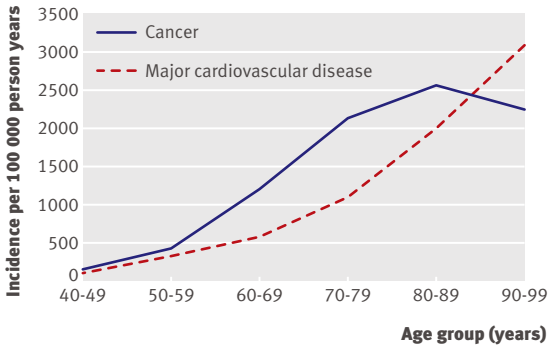
### Statistical analysis

After exclusions, 22 048 men remained for analysis. We used age as time scale starting at baseline age and we censored participants at the age they developed an end point, died, or finished follow-up. As only few participants lived to age 100, we censored data at this age. We calculated one year crude incidence (per 100 000 person years) for each age and collapsed these into 10 year age groups. We stratified age specific incidence by smoking status (ever *v* never). Using a modified Kaplan-Meier method we estimated the cumulative incidence of major cardiovascular disease or cancer conditional on survival to age 40.[4] To understand the influence of mortality on estimates of disease incidence in very old people, we calculated remaining lifetime risk[5] for cancer, major cardiovascular disease, and overall cardiovascular disease for those who reached the ages of 50, 60, 70, 80, and 90 free of the disease of interest.

We produced incidence estimates using the practical incidence estimators macro.[6] Statistical calculations were done using SAS version 9.1 software.

## RESULTS

A total of 3051 participants were aged 65 or older at baseline (see bmj.com). After 23 years of follow-up (478 692 person years), 76.5% of the cohort was still alive. Overall, 32 142 person years had accrued in men aged 80-89 and 3312 person years in those aged 90-99. During follow-up 3252 cases of major cardiovascular disease and 5400 incident cancers were confirmed. The incidence of major cardiovascular disease continued to rise through the 10th decade, with a rate of 3110 per 100 000 person years (figure). In contrast, the age specific incidence of overall cancer increased steadily from 160 per 100 000 person years in those aged 40-49 to 2555 per 100 000 person years in those aged 80-89. It then declined to 2264 per 100 000 person years in those aged 90-99.




Crude incidence of overall cancer and major cardiovascular disease by age

Whereas cardiovascular disease diagnosed at death increased dramatically with age, the incidence of non-fatal myocardial infarction declined, and the incidence of non-fatal stroke increased only slightly after age 89 (see bmj.com). Revascularisation procedures and angina declined noticeably with age as the first manifestation of cardiovascular disease, whereas confirmed major cardiovascular disease continued to increase with age (see bmj.com).

The most common cancers were prostate (47.2%), colorectal (10.3%), lymphoma (6.6%), lung (6.6%), and melanoma (5.7%; see bmj.com). Most of the cancers that declined before age 100 were those detected by screening, whereas the incidence of cancers for which there was no routine screening continued to increase up to age 99 (see bmj.com). The cancer rate among ever smokers peaked at age 80-89 (2883 per 100 000 person years) and then declined, whereas the rate among never smokers peaked in the ninth decade (2205 per 100 000 person years). In contrast, the incidence of major cardiovascular disease increased through the 10th decade in smokers and non-smokers (data not shown).

Adjustment for competing risks of death attenuated the estimate of cumulative incidence for cancer and major cardiovascular disease (see bmj.com). Whereas cumulative incidence continued to increase among the oldest participants, mortality adjusted curves flattened out in the 10th decade. The table shows the remaining lifetime risk of major diseases for men who reached various index ages free of the disease of interest. The risk of both cancer and cardiovascular disease decreased as remaining lifetime diminished. The risk of any cancer at age 40 was 45.1% (95% confidence interval 43.8% to 46.3%) and at 90 was 9.6% (7.2% to 11.9%) and of major cardiovascular disease at age 40 was 34.8% (33.1% to 36.5%) and at 90 was 16.7% (12.9% to 20.6%).

## DISCUSSION

In this large prospective cohort of US male doctors without serious diseases at study entry, the incidence of major cardiovascular disease increased exponentially through the 10th decade, whereas the incidence of cancer peaked at age 80-90 and then declined. In those aged more than 80, new cases of cardiovascular disease were most often diagnosed at death. The decrease in incidence of cancer late in life seemed largely due to a decline in cancers detected by screening. The cumulative incidence of cardiovascular disease and cancer were clearly attenuated by adjustment for risk of mortality. The remaining lifetime risk of both diseases decreased with age.

The rate of major cardiovascular disease continued to increase up to age 99. However, the rate of non-fatal myocardial infarction declined after age 89, whereas that of non-fatal stroke reached a plateau by age 90. The decline in non-fatal cardiovascular events in the face of increased deaths due to cardiovascular disease suggests that people of advanced age may have substantial undiagnosed cardiovascular disease. Self reported revascularisation procedures and angina declined noticeably after age 79. This may be due in part to decreased symptoms or decreased reporting, but may also reflect a decrease in aggressive medical care and investigations. When we included angina or coronary procedures as cardiovascular events, age specific incidence approached a plateau in the oldest participants, illustrating the importance of disease definition in incidence studies.

Our findings of a decline in incidence of cancer after age 89 are similar to results of an analysis using data from the Surveillance, Epidemiology, and End Results programme (SEER), a US cancer registry.[7] The incidence of invasive cancer in white men peaked at a similar age to ours (85-89) and at about the same incidence (2500 per 100 000), and then declined.

However, the decrease in incidence of cancer late in life seemed to be largely driven by cancers detected by screening. The most dramatic decline was in prostate cancer, which accounted for 47% of all cancers diagnosed. Prostate cancer peaked the earliest of all cancers, in those aged 70-79. Melanoma and colorectal cancer declined after age 89. With the exception of lymphoma and tumours of the urinary tract, the incidence of all other cancers continued to increase to age 100. Gastrointestinal malignancies and tumours of unknown origin became prominent with age, similar to other studies.[8 9] Our findings suggest that a substantial part of the decline in the overall incidence of cancer late in life is accounted for by decreased ascertainment of disease and may not represent a true decrease in risk. It is possible that our participants were less likely to report new diagnoses of cancer in advanced age. However, that cancers not detected by screening peaked later in our cohort than in other studies might reflect the higher level of medical surveillance and reporting of symptoms in this cohort of doctors than would be the case in a general population.

### Lifetime risks

While the cumulative incidence curve shows cardiovascular disease increasing sharply to age 100, adjustment for competing risks of death resulted in a substantial decrease in risk in men aged 80 or more (see bmj.com). A similar decrease was seen for cancer. Cumulative incidence overestimated the actual risk of cardiovascular disease in our population by 16% and cancer by 8.5%.

The remaining lifetime risk of major cardiovascular disease decreased from 1 in 2 for men aged 40 to 1 in 6 for men aged 90. When we used a broader definition of cardiovascular disease, the remaining lifetime risk at age 40 was higher (41.3% *v* 34.8%), whereas the risk at age 90 was lower (16.7% *v* 13.7%). This is because some participants reported angina or revascularisation procedures before a major event, thus shifting their date of diagnosis to earlier. The age specific and cumulative incidence curves would suggest that cardiovascular disease depends on age.[10] However, lifetime risk curves show that the incidence begins to plateau later in life, as any increased risk is outpaced by competing risks of death. Cancer is often thought of as inextricably linked with aging, but it seems to fit the pattern of age related diseases better, occurring in a particular age range then declining. The remaining lifetime risk of cancer decreased from 1 in 2 in 40 year olds to 1 in 10 in 90 year olds.

A plateau or decline in lifetime risk with age has previously been reported for overall cancer,[7] stroke,[11] coronary artery disease,[12] and Alzheimer's disease.[13] This may be a function of decreased life expectancy in very old people; it may also be due to the selective survival of those who are more resistant to disease. Finally, the observed decline in lifetime risk may not represent a true decrease in risk but rather decreased reporting of or diagnosis of disease.

It is important to emphasise that estimates of lifetime risk strongly depend on life expectancy and they cannot be directly compared across populations unless mortality rates are similar. The goal of our analysis was not to provide estimates that would be readily applicable to men in general, but to investigate the risk of disease in advanced age.

**Remaining lifetime risk of first cancer or major cardiovascular disease event by age reached free of events in participants of Physicians' Health Study**

| | Remaining lifetime risk (%) (95% CI) | | |
|---|---|---|---|
| Age (years) | All cancer | Major cardiovascular disease | All cardiovascular disease |
| 40 | 45.1 (43.8 to 46.3) | 34.8 (33.1 to 36.5) | 41.3 (39.9 to 42.7) |
| 50 | 44.5 (43.3 to 45.7) | 34.5 (32.8 to 36.2) | 40.5 (39.1 to 41.9) |
| 60 | 42.9 (41.6 to 44.2) | 33.0 (31.2 to 34.8) | 38.0 (36.5 to 39.5) |
| 70 | 36.6 (35.2 to 38.0) | 30.5 (28.5 to 32.4) | 33.2 (31.5 to 34.9) |
| 80 | 24.3 (22.5 to 26.1) | 25.7 (23.3 to 29.1) | 24.9 (22.7 to 27.1) |
| 90 | 9.6 (7.2 to 11.9) | 16.7 (12.9 to 20.6) | 13.7 (10.3 to 17.2) |

Remaining lifetime risk=mortality adjusted cumulative risk conditional on disease-free survival to age specified.

### Strengths and limitations

Our study has several strengths, including the large number of participants and outcome events, prospective design, and well defined population with a long follow-up. End points were ascertained and confirmed after review of medical records. We adjusted cumulative incidence for competing risks of mortality.

Several limitations must also be considered. Firstly, our findings may not be generalisable to a broader population as our cohort consisted almost exclusively of highly educated white men. Our participants might have a lower risk of cancer and cardiovascular disease than a general population for several reasons. They were healthy at baseline, had a lower incidence of smoking and obesity than expected, and had participated in a primary prevention trial of aspirin and β carotene. Randomisation to aspirin was associated with a decreased risk of myocardial infarction but not of stroke or death from cardiovascular disease.[1] Most participants became regular users of aspirin after completion of the trial.[14] Aspirin use was not associated with risk of cancer. Treatment thus had no effect on the study outcomes, with the exception of lower rates of myocardial infarction. Our participants may also have had increased rates of screening. Nevertheless, the lifetime risk for overall cancer (45.1%) was nearly identical to that of the SEER estimate for white men based on a sample of the US population (44.9%).[7] The lifetime risk of stroke in men aged 55 in our cohort (15.4%) was also similar to that of men in the population based Framingham Heart Study (16.9%).[11 12]

Finally, despite the homogeneity of our population for race and education, biological associations are expected to be similar in our study compared with other male populations.

### Conclusion

In summary, the lifetime risks of cardiovascular disease and cancer approached a plateau after age 90. This may be due to decreased detection or reporting of disease as well as increased resistance to disease. Substantial amounts of undiagnosed cardiovascular disease and cancer may contribute to frailty in people of advanced age. Accurate estimates of long term risk require adjustment for competing risks of mortality, particularly in an aging population.

We thank the staff of the Physician's Health Study and the participants.

### WHAT IS ALREADY KNOWN ON THIS TOPIC

Age is the strongest risk factor for cardiovascular disease and cancer but the incidence of these diseases in people aged 80 or more is less clear

Studies of incidence in the ninth and 10th decade of life are sparse

### WHAT THIS STUDY ADDS

The incidence of cardiovascular disease in a cohort of US male doctors increased to age 100 whereas that of overall cancer decreased after age 89

The decline in cancer incidence was largely driven by a decrease in screening related cancers, whereas cardiovascular disease after age 80 was most commonly diagnosed at death

**Contributors:** See bmj.com.
**Funding:** The authors' work was independent of the funders. JAD received support for this research from the Hartford Foundation's Center of Excellence in Geriatric Medicine at Harvard Medical School and the Eleanor and the Miles Shore 50th anniversary fellowship programme for scholars in medicine at Harvard Medical School. The Physicians' Health Study is supported by grants CA-34944, CA-40360, and CA-097193 from the National Cancer Institute, and grants HL-26490 and HL-34595 from the National Heart, Lung, and Blood Institute, Bethesda, Maryland.
**Competing interests:** Although the authors believe they have no competing interests that could influence their decisions, work, or writing of the manuscript, they report a full disclosure for the past five years. LD received investigator initiated funding and support as principal investigator from the National Institutes of Health, the Alcoholic Beverage Medical Research Foundation, the Biomedical Research Institute at Brigham and Women's Hospital, and the Huntington Disease Society of America. GL received investigator initiated research funding from the National Institutes of Health and received honorariums from Pfizer and Lilly Pharmaceutical for speaking engagements in 2003. JMG received investigator initiated research funding and support as principal investigator from the National Institutes of Health, BASF, D SM Pharmaceuticals, Wyeth Pharmaceuticals, McNeil Consumer Products, and Pliva; received honorariums from Bayer and Pfizer for speaking engagements; and is a consultant for Bayer, McNeil Consumer Products, Wyeth Pharmaceuticals, Merck, Nutraquest, and GlaxoSmithKline. TK received investigator initiated research funding as principal investigator or co-investigator from the National Institutes of Health, Bayer AG, McNeil Consumer & Specialty Pharmaceuticals, Merck, and Wyeth Consumer Healthcare; he is a consultant to i3 Drug Safety, and received honorariums from Organon for contributing to an expert panel and from Genzym for educational lectures.
**Ethical approval:** This study analysed existing data from the Physicians' Health Study, and was approved by the institutional review board of the Brigham and Women's Hospital.
**Provenance and peer review:** Not commissioned; externally peer reviewed.

1 Final report on the aspirin component of the ongoing physicians' health study. Steering Committee of the Physicians' Health Study Research Group. *N Engl J Med* 1989;321:129-35.
2 Hennekens CH, Buring JE, Manson JE, Stampfer M, Rosner B, Cook NR, et al. Lack of effect of long-term supplementation with beta carotene on the incidence of malignant neoplasms and cardiovascular disease. *N Engl J Med* 1996;334:1145-9.
3 World Health Organization. *Ischemic heart disease registers: report of the Fifth Working Group* (including a second revision of the operating protocol), Copenhagen, 26-29 April 1971. Copenhagen, Denmark: Regional Office for Europe, WHO, 1971.
4 Kaplan E, Meier P. Nonparametric estimation from incomplete observations. *J Am Stat Assoc* 1958;53:457-81.
5 Gaynor J, Feuer E, Tan C, Wu D, Little C, Straus D, et al. On the use of cause-specific failure and conditional failure probabilities: examples from clinical oncology data. *J Am Stat Assoc* 1993;88:400-9.
6 Beiser A, D'Agostino RB Sr, Seshadri S, Sullivan LM, Wolf PA. Computing estimates of incidence, including lifetime risk: Alzheimer's disease in the Framingham study. The practical incidence estimators (PIE) macro. *Stat Med* 2000;19:1495-522.
7 Merrill RM, Weed DL. Measuring the public health burden of cancer in the United States through lifetime and age-conditional risk estimates. *Ann Epidemiol* 2001;11:547-53.
8 De Rijke JM, Schouten LJ, Hillen HF, Kiemeney LA, Coebergh JW, van den Brandt PA. Cancer in the very elderly Dutch population. *Cancer* 2000;89:1121-33.
9 Saltzstein SL, Behling CA, Baergen RN. Features of cancer in nonagenarians and centenarians. *J Am Geriatr Soc* 1998;46:994-8.
10 Brody JA, Schneider EL. Diseases and disorders of aging: an hypothesis. *J Chronic Dis* 1986;39:871-6.
11 Seshadri S, Beiser A, Kelly-Hayes M, Kase CS, Au R, Kannel WB, et al. The lifetime risk of stroke: estimates from the Framingham study. *Stroke* 2006;37:345-50.
12 Lloyd-Jones DM, Larson MG, Beiser A, Levy D. Lifetime risk of developing coronary heart disease. *Lancet* 1999;353:89-92.
13 Seshadri S, Wolf PA. Lifetime risk of stroke and dementia: current concepts, and estimates from the Framingham study. *Lancet Neurol* 2007;6:1106-14.
14 Cook NR, Hebert PR, Manson JE, Buring JE, Hennekens CH. Self-selected posttrial aspirin use and subsequent cardiovascular disease and mortality in the physicians' health study. *Arch Intern Med* 2000;160:921-8.

**Accepted:** 19 September 2008

# Discrepancies in sample size calculations and data analyses reported in randomised trials: comparison of publications with protocols

An-Wen Chan,[1] Asbjørn Hróbjartsson,[2] Karsten J Jørgensen,[2] Peter C Gøtzsche,[2] Douglas G Altman[3]

[1]Mayo Clinic, Rochester, USA
[2]Nordic Cochrane Centre, Copenhagen, Denmark
[3]Centre for Statistics in Medicine, University of Oxford, Oxford

Correspondence to: A-W Chan
chan.anwen@mayo.edu

Cite this as: *BMJ* 2008;337:a2299
doi:10.1136/bmj.a2299

## ABSTRACT

**Objective** To evaluate how often sample size calculations and methods of statistical analysis are pre-specified or changed in randomised trials.

**Design** Retrospective cohort study.

**Data source** Protocols and journal publications of published randomised parallel group trials initially approved in 1994-5 by the scientific-ethics committees for Copenhagen and Frederiksberg, Denmark (n=70).

**Main outcome measure** Proportion of protocols and publications that did not provide key information about sample size calculations and statistical methods; proportion of trials with discrepancies between information presented in the protocol and the publication.

**Results** Only 11/62 trials described existing sample size calculations fully and consistently in both the protocol and the publication. The method of handling protocol deviations was described in 37 protocols and 43 publications. The method of handling missing data was described in 16 protocols and 49 publications. 39/49 protocols and 42/43 publications reported the statistical test used to analyse primary outcome measures. Unacknowledged discrepancies between protocols and publications were found for sample size calculations (18/34 trials), methods of handling protocol deviations (19/43) and missing data (39/49), primary outcome analyses (25/42), subgroup analyses (25/25), and adjusted analyses (23/28). Interim analyses were described in 13 protocols but mentioned in only five corresponding publications.

**Conclusion** When reported in publications, sample size calculations and statistical methods were often explicitly discrepant with the protocol or not pre-specified. Such amendments were rarely acknowledged in the trial publication. The reliability of trial reports cannot be assessed without having access to the full protocols.

## INTRODUCTION

Sample size calculations and data analyses have an important impact on the planning, interpretation, and conclusions of randomised trials. Statistical analyses often involve several subjective decisions about which data to include and which tests to use, producing potentially different results and conclusions.[1-7] Methods of analysis that are chosen or altered after preliminary examination of the data can introduce bias if a subset of favourable results is then reported in the publication.

The study protocol plays a key role in reducing such bias by documenting a pre-specified blueprint for conducting and analysing a trial. To evaluate the completeness and consistency of reporting, we reviewed a comprehensive cohort of randomised trials and compared the sample size calculations and data analysis methods described in the protocols with those reported in the publications.

## METHODS

We included all published parallel group randomised trials approved by the scientific-ethics committees for Copenhagen and Frederiksberg, Denmark, from 1 January 1994 to 31 December 1995. We confirmed journal articles for each trial by surveying investigators and searching PubMed, Embase, and the Cochrane Controlled Trials Register as part of a separate study of outcome reporting.[8]

For each trial, we reviewed the protocol, statistical appendices, amendments, and publications that reported the primary outcome measures (see bmj.com). The pre-specified primary outcomes of our study were the proportion of trial protocols and publications that did not provide key information (described below) about sample size calculations and statistical methods and the proportion of trials with discrepancies between the information presented in the protocol and the publication. We considered a 10% or greater difference between calculated sample sizes in the protocol and publication to be a discrepancy, as well as any qualitative or quantitative difference in other information we examined (see box 1 and bmj.com).

## RESULTS

We identified 70 parallel group randomised trials that received ethics approval in 1994-5 and were subsequently published.[8]

### Sample size calculation

Overall, only 11 trials fully and consistently reported all of the requisite components of the sample size calculation in both the protocol and the publication.

*Completeness of reporting*—An a priori sample size calculation was reported for 62 trials; 28 were described only in the protocol and 34 in both the protocol and the publication. Thirty seven protocols and 21 publications reported all of the components of the sample size calculation (figure). Individual components were reported in 74-100% of protocols and 48-75% of publications. Nine protocols provided only the calculated sample size without any further details about the calculation. Among trials that reported an estimated minimum clinically important effect size (delta), 20/53 protocols and 10/33 publications stated the basis on which the figure was derived.

This article is an abridged version of a paper that was published on bmj.com. Cite this article as: *BMJ* 2008;337:a2299

Discrepancies in sample size calculations reported in trial publications compared with protocols

| Component of sample size calculation | No of trials with discrepancy | | |
|---|---|---|---|
| | Total | Not pre-specified* | Different from protocol description |
| Outcome measure (n=31)† | 7 | 3 | 4 |
| Estimated delta (n=33)† | 12 | 6 | 6: 3 larger in protocol; 3 larger in article |
| Estimated event rates (n=16)‡ | 3 | 3 | 0 |
| Estimated standard deviation (n=14)§ | 5 | 2 | 3: 2 larger in protocol; 1 larger in article |
| Alpha (n=33)† | 2 | 2 | 0 |
| Power (n=34)† | 9 | 2 | 7: 5 larger in protocol; 2 larger in article |
| Calculated sample size (n=30)† | 8 | 0 | 8¶: 7 larger in protocol; 1 larger in article |
| Any component (n=34)** | 18 | 8 | 16 |

*Reported in publication but not mentioned in protocol.
†Among trials reporting component in publication.
‡Among trials reporting event rates for binary outcome measures in publication.
§Among trials reporting standard deviations for continuous outcome measures in publication.
¶Greater than 10% difference in calculated sample size.
**Among trials reporting any component in publication.

*Comparison of calculated and actual sample sizes*—Sixty two trials provided a calculated sample size in the protocol. Of these, 30 subsequently recruited a sample size within 10% of the calculated figure from the protocol; 22 trials randomised at least 10% fewer participants than planned as a result of early stopping (n=3), poor recruitment (2), and unspecified reasons (17) (see bmj.com).

*Discrepancies between publications and protocols*—Both the publications and the protocols for 34 trials described a sample size calculation. Overall, we noted discrepancies in at least one component of the published sample size calculation when compared with the protocol for 18 trials (figure). Publications for eight trials reported components that had not been pre-specified in the protocol, and 16 had explicit discrepancies between information contained in the publication and protocol (table, box 2). None of the publications mentioned any amendments to the original sample size calculation.

**Box 1 Definitions of collected data**

**Design framework**

*Superiority trial*—Explicitly described as a study designed to show a difference in effects between interventions or not explicitly described as an equivalence or non-inferiority trial

*Non-inferiority trial*—Explicitly described as a study designed to show that one intervention is not worse than another, or a non-inferiority margin is specified, or a one sided confidence interval is presented

*Equivalence trial*—Explicitly described as a study designed to show that one intervention is neither inferior nor superior to another or an equivalence margin is specified

**Handling of protocol deviations**

*Intention to treat analysis*—All participants with available data are analysed in the original group to which they were randomly assigned (as randomised), regardless of adherence to the protocol. No data are excluded for reasons other than loss to follow-up

*Per protocol analysis*—Participants with available data are analysed as randomised provided they meet some defined level of adherence to the protocol

*As treated analysis*—Participants are analysed in the group corresponding to the actual intervention received (ignoring original randomisation)

**Primary outcome**

Main outcome(s) of interest, in the following hierarchical order:

1 Explicitly defined as primary or main
2 Outcome used in the power calculation
3 Main outcome stated in the trial objectives

### Protocol deviations

The specific method of handling protocol deviations in the primary statistical analysis (as defined in box 1) was named or described in 37 protocols and 43 publications (figure). Overall, the primary method described for handling protocol deviations in the publication differed from that described in the protocol for 19/43 trials. None of these discrepancies was acknowledged in the journal publication.

Thirty protocols and 33 publications used the term "intention to treat" analysis and applied a variety of definitions (see bmj.com). Few of these protocols (n=7) and publications (3) made it explicit whether study participants were analysed in the group to which they were originally randomised. Most protocols (22) and publications (18) incorrectly excluded participants from the intention to treat analysis for reasons other than loss to follow-up.

### Missing data

The method of handling missing data was described in only 16 protocols and 49 publications (figure). Methods reported in publications differed from the protocol for 39/49 trials. Published methods were often not pre-specified in the protocol (38/49). For one trial, the protocol stipulated that missing data would be counted as failures, whereas in the publication they were excluded from the analysis.

### Primary outcome analysis and overall number of tests

Fifty four trials designated at least one outcome measure as primary in the protocol (n=49) or publication (43). The statistical method for analysing the primary outcome measure was described in 39 protocols and 42 publications. Overall, 25 publications that described the statistical test for primary outcome measures differed from the protocol (figure, box 2).

The median number of between group statistical tests defined in 44 protocols was 30 (10th-90th centile range 8-218); the other 26 protocols contained

insufficient statistical detail. Publications for all 70 trials reported a median of 22 (8-71) tests. Half of the protocols (n=36) and publications (34) did not define whether hypothesis testing was one or two sided. Interestingly, we found one neurology trial that used two sided P values in one publication (all P values >0.1) and a one sided P value in another (P=0.028).

### Subgroup analysis

Overall, 25 trials described subgroup analyses in the protocol (n=13) or publication (20). All had discrepancies between the two documents (figure, box 2). Twelve of the trials with protocol specified analyses reported only some (n=7) or none (5) in the publication. Nineteen of the trials with published subgroup analyses reported at least one that was not pre-specified in the protocol. Protocols for 12 of these trials specified no subgroup analyses, whereas seven specified some but not all of the published analyses. Only seven publications explicitly stated whether the analyses were defined a priori; four of these trials claimed that the subgroup analyses were pre-specified even though they did not appear in the protocol.

**Box 2 Anonymised examples of unacknowledged discrepancies in sample size calculations and statistical analyses reported in publications compared with protocols**

**Sample size calculation**

***Changed delta (1)***
- Outcome: disease progression or death rate
- Protocol: delta 10%; event rates unspecified
- Publication: delta 6%; event rates 16% and 10%

***Changed delta (2)***
- Outcome: mean number of active joints
- Protocol: delta 2.5 joints
- Publication: delta 5 joints

***Changed standard deviation***
- Outcome: mean symptom score
- Protocol: 1.4
- Publication: 0.49

***Changed power***
- Outcome: survival without disease progression
- Protocol: 90%
- Publication: 80%

***Changed sample size estimate***
- Outcome: thromboembolic complication rate
- Protocol: 2200
- Publication: 1500

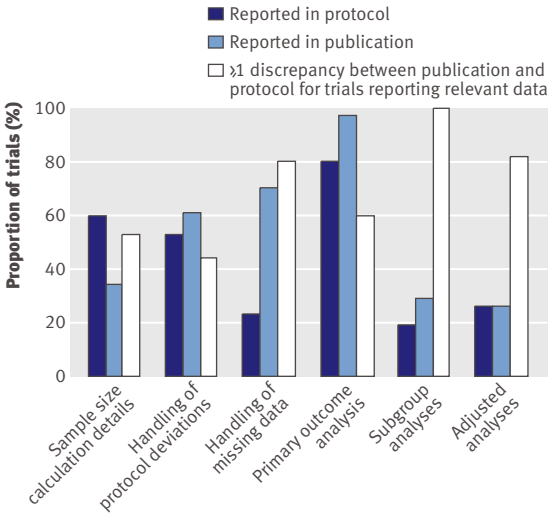
**Statistical analyses**

***Changed primary outcome analysis***
- Outcome: global disease assessment
- Protocol: $\chi^2$ test
- Publication: analysis of covariance

***New subgroups added to publication***
- Outcome: time to progression or death
- Protocol: baseline disease severity
- Publication: duration of previous treatment*, type of previous treatment*, blood count*, disease severity

***Omitted covariates for adjusted analysis in publication***
- Outcome: neurological score at six months
- Protocol: baseline neurological score, pupil reaction, age, CT scan classification, shock, haemorrhage
- Publication: no adjusted analysis reported

*Described explicitly as pre-specified despite not appearing in the protocol




**Reporting of sample size calculations and data analyses in publications compared with protocols**

### Adjusted analysis

Overall, 28 trials described adjusted analyses in the protocol (n=18) or publication (18). Of these, 23 had discrepancies between the two documents (figure, box 2). Twelve of the trials with protocol specified covariates reported no adjustment (n=10) or omitted at least one pre-specified covariate (2) from the published analysis. Twelve of the trials with published adjusted analyses used covariates that were not pre-specified in the protocol. Ten of these trials did not mention any adjusted analysis in the protocol, whereas two trials added new covariates to those specified in the protocol. Publications for only one trial explicitly stated whether the covariates were defined a priori.

### Interim analyses and data monitoring boards

Interim analyses were described in 13 protocols, but reported in only five corresponding publications. An additional two trials reported interim analyses in the publications, despite the protocol explicitly stating that there would be none. A data monitoring board was described in 12 protocols but in only five of the corresponding publications.

## DISCUSSION

We identified a high frequency of unacknowledged discrepancies and poor reporting of sample size calculations and data analysis methods in an unselected cohort of randomised trials. We reviewed key methodological information that can introduce bias if misrepresented or altered retrospectively. Our broad sample of protocols is

## WHAT IS ALREADY KNOWN ON THIS TOPIC

The results and conclusions of randomised trials are influenced by the choice of statistical analysis methods and individual components of sample size calculations

If these methodological choices are defined or altered after examination of the data, the potential for biased reporting of favourable results is substantial

## WHAT THIS STUDY ADDS

Trial protocols and publications are often missing important methodological information about sample size calculations and statistical analysis methods

When described, methodological information in journal publications is often discrepant with information in trial protocols

a key strength, as unrestricted access to such documents is often very difficult to obtain.[9]

One limitation is that our cohort may not reflect recent protocols and publications, as this type of review can be done only several years after protocol submission to allow time for publication. Whether the widespread adoption of CONSORT and other reporting guidelines for publications has improved the quality of protocols or reduced the prevalence of unacknowledged amendments in recent years is unclear.[10] However, our results are consistent with more recent examples of discrepancies.[3 11 12] Furthermore, we previously found that the prevalence of publication restrictions stated in industry initiated trial protocols did not change between 1995 and 2004.[13]

We also acknowledge that detailed statistical analysis plans may not always be included in the application for scientific or ethical review as a result of varying standards, even though this information has a role in evaluating the validity of a study. However, this does not explain the frequent discrepancies we found between explicit descriptions in protocols and publications (box 2).

The 70 trials in this study were part of a larger review that found unacknowledged changes to primary outcome measures in more than half of 102 trials.[8] Therefore, we are not surprised to find frequent discrepancies in other aspects of study conduct.

Accurate reporting of sample size calculations and data analysis methods is important not only for the sake of transparency but also because the choice of methods and the reasons for choosing them can directly influence the interpretation and conclusions of study results (see bmj.com).[1-7 14] Public access to full protocols is thus needed to reliably appraise trial publications. Several journals have recognised this principle and require submission of protocols with manuscripts.[15-17]

### Conclusions

Our findings support the need to improve the content of trial protocols and encourage transparent reporting of amendments in publications through research training. In collaboration with journal editors, trialists, methodologists, and ethicists, we have launched the SPIRIT (standard protocol items for randomised trials) initiative to establish evidence based recommendations for the key content of trial protocols.[18]

To improve the reliability of published results, investigators should document the sample size calculations and full analysis plans before the trial is started and should then analyse the results with fidelity to the study protocol or describe major amendments in the publication.[19] As the guardians of clinical research before study inception, scientific and ethical review committees can help to ensure that statistical analysis plans are well documented in protocols. Only with fully transparent reporting of trial methods and public access to protocols can the results be properly appraised, interpreted, and applied to care of patients.

**Contributors:** See bmj.com.
**Funding sources:** None.
**Competing interests:** A-WC chairs and AH, DGA, and PCG are members of the steering committee for the SPIRIT (standard protocol items for randomised trials) initiative, a project that aims to define key protocol content for randomised trials.
**Ethical approval:** The scientific-ethics committees for Copenhagen and Frederiksberg granted unrestricted access to trial protocols. The research did not undergo the usual ethical review process, as patients' data were not reviewed.
**Provenance and peer review:** Not commissioned; externally peer reviewed.

1 Psaty BM, Kronmal RA. Reporting mortality findings in trials of rofecoxib for Alzheimer disease or cognitive impairment: a case study based on documents from rofecoxib litigation. *JAMA* 2008;299:1813-7.
2 Porta N, Bonet C, Cobo E. Discordance between reported intention-to-treat and per protocol analyses. *J Clin Epidemiol* 2007;60:663-9.
3 Curfman GD, Morrissey S, Drazen JM. Expression of concern reaffirmed. *N Engl J Med* 2006;354:1193.
4 Melander H, Ahlqvist-Rastad J, Meijer G, Beermann B. Evidence b(i)ased medicine—selective reporting from studies sponsored by pharmaceutical industry: review of studies in new drug applications. *BMJ* 2003;326:1171-3.
5 Schiffner R, Schiffner-Rohe J, Gerstenhauer M, Hofstadter F, Landthaler M, Stolz W. Differences in efficacy between intention-to-treat and per-protocol analyses for patients with psoriasis vulgaris and atopic dermatitis: clinical and pharmacoeconomic implications. *Br J Dermatol* 2001;144:1154-60.
6 Murray GD. Research governance must focus on research training. *BMJ* 2001;322:1461-2.
7 Hrachovec JB, Mora M. Reporting of 6-month vs 12-month data in a clinical trial of celecoxib. *JAMA* 2001;286:2398-400.
8 Chan AW, Hróbjartsson A, Haahr MT, Gøtzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *JAMA* 2004;291:2457-65.
9 Chan AW, Upshur R, Singh JA, Ghersi D, Chapuis F, Altman DG. Research protocols: waiving confidentiality for the greater good. *BMJ* 2006;332:1086-9.
10 Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet* 2001;357:1191-4.
11 Hahn S, Williamson PR, Hutton JL. Investigation of within-study selective reporting in clinical research: follow-up of applications submitted to a local research ethics committee. *J Eval Clin Pract* 2002;8:353-9.
12 Hernandez AV, Steyerberg EW, Taylor GS, Marmarou A, Habbema JD, Maas AI. Subgroup analysis and covariate adjustment in randomized clinical trials of traumatic brain injury: a systematic review. *Neurosurgery* 2005;57:1244-53.
13 Gøtzsche PC, Hróbjartsson A, Johansen HK, Haahr MT, Altman DG, Chan AW. Constraints on publication rights in industry-initiated clinical trials. *JAMA* 2006;295:1645-6.
14 Lagakos SW. The challenge of subgroup analyses—reporting without distorting. *N Engl J Med* 2006;354:1667-9.
15 Jones G, Abbasi K. Trial protocols at the BMJ. *BMJ* 2004;329:1360.
16 McNamee D, James A, Kleinert S. Protocol review at the Lancet. *Lancet* 2008;372:189-90.
17 PLoS Med. Guidelines for authors. http://journals.plos.org/plosmedicine/guidelines.php#supporting.
18 Chan AW, Tetzlaff J, Altman DG, Gøtzsche PC, Hróbjartsson A, Krleža-Jeric K, et al. The SPIRIT initiative: defining standard protocol items for randomised trials. *German J Evid Quality Health Care (suppl)* 2008;102:S27.
19 Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med* 2001;134:663-94.

**Accepted:** 10 September 2008