RESEARCH

The *BMJ* is an Open Access journal.
We set no word limits on *BMJ* research articles, but they are abridged for print. The full text of each *BMJ* research article is freely available on bmj.com

Corticosteroids for pain relief in sore throat: systematic review and meta-analysis

Gail Hayward,[1] Matthew Thompson,[1] Carl Heneghan,[1] Rafael Perera,[1] Chris Del Mar,[2] Paul Glasziou[1]

EDITORIAL by Little

[1]Department of Primary Health Care, University of Oxford, Oxford OX3 7LF
[2]Bond University, Gold Coast, Queensland 4229, Australia
Correspondence to: M Thompson matthew.thompson@dphpc.ox.ac.uk

Cite this as: *BMJ* 2009;339:b2976
doi: 10.1136/bmj.b2976

ABSTRACT

Objective To evaluate whether systemic corticosteroids improve symptoms of sore throat in adults and children.
Design Systematic review and meta-analysis.
Data sources Cochrane Central, Medline, Embase, Database of Reviews of Effectiveness (DARE), NHS Health Economics Database, and bibliographies.
Outcome measures Percentage of patients with complete resolution at 24 and 48 hours, mean time to onset of pain relief, mean time to complete resolution of symptoms, days missed from work or school, recurrence and adverse events.
Results We included eight trials, consisting of 743 patients in total (369 children, 374 adults). 348 (47%) had exudative sore throat, and 330 (44%) were positive for group A β-haemolytic streptococcus. In addition to antibiotics and analgesia, corticosteroids significantly increased the likelihood of complete resolution of pain at 24 hours (four trials) by more than three times (relative risk 3.2, 95% confidence interval 2.0 to 5.1), and at 48 hours (three trials) to a lesser extent (1.7, 1.3 to 2.1). Corticosteroids (six trials) reduced mean time to onset of pain relief by more than 6 hours (95% confidence interval 3.4 to 9.3, P<0.001), although significant heterogeneity was present. The mean time to complete resolution was inconsistent across trials and a pooled analysis was not undertaken. Reporting of other outcomes was limited.
Conclusions Corticosteroids provide symptomatic relief of pain in sore throat, in addition to antibiotic therapy, mainly in participants with severe or exudative sore throat.

WHAT IS ALREADY KNOWN ON THIS TOPIC

Corticosteroids are beneficial for symptoms of upper respiratory tract infections
Sore throat is a common condition in primary care
Recent guidelines recommend that antibiotics should not be prescribed for sore throat

WHAT THIS STUDY ADDS

At 24 hours, patients with severe sore throat who are given corticosteroids in addition to antibiotics are three times more likely to report complete resolution of symptoms than those who do not receive corticosteroids
Corticosteroids also reduce the time to mean onset of pain relief in this patient group by about 6 hours
The effect of corticosteroids independent of antibiotics is unknown and should be the focus of future research

This article is an abridged version of a paper that was published on bmj.com. Cite this article as: *BMJ* 2009;339:b2976

INTRODUCTION

Treatment of sore throat with antibiotics provides only modest beneficial effect in reducing symptoms and fever.[1 2] However, prescribing rates remain disproportionately high.[3] High rates of antibiotic prescriptions contribute to antibiotic resistance[4] and also lead to the "medicalising" of sore throat, which can result in increased rates of patient attendance.[5 6] In developed countries, prescribing is no longer justified to prevent complications from group A β-haemolytic streptococcus infection.[2 7 8]

The pressure for clinicians to reduce antibiotic prescriptions for sore throat leaves a therapeutic vacuum. Corticosteroids inhibit transcription of proinflammatory mediators in human airway endothelial cells which cause pharyngeal inflammation and ultimately symptoms of pain.[9] Corticosteroids are beneficial in other upper respiratory tract infections such as acute sinusitis, croup, and infectious mononucleosis.[10-12] We hypothesised that corticosteroids would offer similar symptomatic relief from sore throat because of their anti-inflammatory effects, and undertook a systematic review to examine the effect of systemic corticosteroids on adults and children with sore throat.

METHOD

Search strategy and selection

We included randomised controlled trials comparing systemic corticosteroids with placebo, in children or adults, in outpatient settings. We included studies of patients with clinical signs of acute tonsillitis or pharyngitis and patients with a clinical syndrome of "sore throat" (painful throat, odynophagia).

We searched Medline, Embase, the Cochrane Library, the Database of reviews of effectiveness (DARE), and the NHS Health Economics Database. Search terms included "upper respiratory tract infection," "pharyngitis," "tonsillitis," "sore throat," and "corticosteroids" and viral and bacterial upper respiratory pathogens. We did citation searches of full-text papers.

Data extraction and quality assessment

We assessed methodological quality of studies by allocation concealment, randomisation, comparability of groups on baseline characteristics, blinding, treatment adherence, and percentage participation.

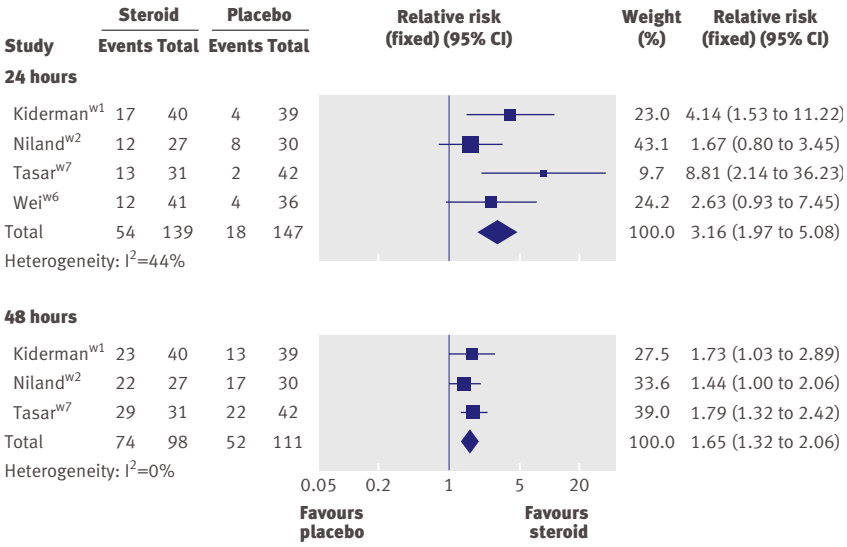


Fig 1 | Effect of corticosteroids on number of patients experiencing complete pain relief at 24 and 48 hours. See web appendix for references

Primary outcomes included the proportion of participants with improvement or complete resolution of symptoms, mean times to onset of pain relief, and complete resolution of pain. Secondary outcomes included adverse events, relapse rates, and days missed from school or work.

We did sensitivity analyses, excluding each study in turn, to determine the stability of the effect. A priori subgroup analyses included age, route of corticosteroid, presence of positive bacterial culture or direct antigen test, and severity of sore throat including presence of exudate. Meta-regression tested subgroup interaction on the outcomes.

Data synthesis and analysis

We expressed dichotomous outcomes as relative risks, and expressed continuous variables as weighted mean difference. We calculated the number needed to treat where data were sufficient. We measured heterogeneity for each outcome.[13] Where no heterogeneity was present, we performed a fixed effects meta-analysis. If substantial heterogeneity was detected, we used a random effects analysis.

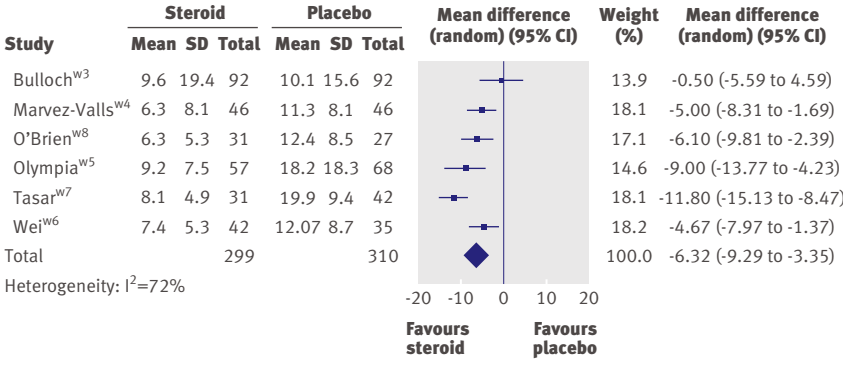


Fig 2 | Effect of corticosteroids on mean time to onset of pain relief in hours. See web appendix for references

RESULTS

Study characteristics

Of 3257 records identified, 26 were relevant to sore throat, tonsillitis, or pharyngitis. Of these, 18 studies that did not meet our inclusion criteria were excluded. The eight remaining studies included 743 patients (369 children, 374 adults): 348 (47%) had exudative sore throat, and 330 (44%) were positive for group A β-haemolytic streptococcus. Methodological quality was high with a low risk of bias. All eight trials prescribed antibiotics to both intervention and placebo groups and allowed simple analgesia.

Complete resolution of pain at 24 or 48 hours

In a pooled analysis of four trials[w1 w2 w6 w7] patients treated with corticosteroids were three times more likely to have complete resolution of pain at 24 hours (relative risk 3.2, 95% confidence interval 2.0 to 5.1, P<0.001, I^2=44%) (fig 1). The number needed to treat was 3.7 (95% confidence interval 2.8 to 5.9). Significant effects were recorded in adult patients only (relative risk 4.3, 2.3 to 8.1, P<0.001)[w1 w6 w7] and in those receiving oral corticosteroids only (2.6, 1.6 to 4.3, P<0.001).[w1 w2 w6]

In three trials[w1 w2 w7] corticosteroids also increased the likelihood of complete resolution of pain at 48 hours (1.7, 1.3 to 2.1, P<0.001), number needed to treat was 3.3 (2.4 to 5.6) (fig 1). Results were similar in trials with adult patients only (1.8, 1.3 to 2.3, P<0.001)[w1 w7] and in those with patients receiving oral corticosteroids only (1.6, 1.2 to 2.1, P=0.004).[w1 w2]

Mean time to onset of pain relief

Six trials reported the mean time to onset of pain relief, which occurred at an average of 6.3 hours earlier with corticosteroids than without (95% confidence interval 9.3 to 3.4, P<0.001) (fig 2).[w3-w8] The wide variation in individual response times caused high heterogeneity (I^2=72%). The majority of the heterogeneity arose from the trial of Tasar et al.[w7] Removal of this trial from the meta-analysis gave a mean time to onset of pain relief 5.1 hours earlier in patients given corticosteroids.

In patients with an exudative sore throat, corticosteroids also reduced the mean time to onset of pain relief (weighted mean difference 6.2 hours, 8.4 to 4.0). Similarly, we recorded a reduction in mean time to pain relief in sore throat that was bacterial-pathogen-positive (5.3, 8.0 to −2.6) and in trials selecting for severe sore throat (7.2, 10.1 to 4.3). All three categories of sore throat (exudative, bacterial pathogen positive, and severe) were significant (P<0.001) with no heterogeneity (I^2=0).

The direction of effect for mean time to onset of pain relief was similar in trials with adults only, in trials using intramuscular and oral routes of steroid administration, and in trials in which severe sore throat was not selected. We did not find significant changes in mean time to onset of pain relief in the subgroup analyses of children only, trials with less than 50% exudative sore throat, and participants with sore throats not positive for bacterial pathogens. Meta-regression analysis revealed no significant differences across all subgroups.

Time to complete resolution of symptoms

Five trials assessed the mean time to complete resolution of pain.[w3-w5 w7 w8] High heterogeneity prevented pooling: three studies showed a benefit of corticosteroids,[w5 w7 w8] and two showed non-significant effects in opposing directions. Time to complete resolution ranged from 15 to 45 hours in the corticosteroid group and 35 to 54 hours in the placebo group.

Adverse events, relapse rates and days missed from school/work

Only one trial[w5] of 125 participants reported adverse events: five patients were hospitalised for fluid rehydration, and three patients developed peritonsillar abscess. Three studies reported no significant differences in days missed from school or work.[w1 w2 w4] Four trials reported no difference in the incidence of recurrent symptoms,[w1-w4] whereas one trial found significantly increased recurrence in the placebo group.[w6]

DISCUSSION

Corticosteroids significantly increase the proportion of patients with sore throat who will experience complete resolution of pain at both 24 and 48 hours. Fewer than four patients need to be treated to prevent one patient continuing to experience a painful sore throat at 24 hours. Although corticosteroids decreased the mean time to onset of pain relief by 6 hours, pooled analysis showed significant heterogeneity. All effects were in addition to antibiotic use.

We found that the effects of corticosteroids on mean time to onset of pain relief were homogeneous in severe, exudative, or bacterial-pathogen-positive sore throat alone. Our data do not support an effect in mild sore throat because only one study included patients with milder symptoms at baseline and showed no significant effect. A meta-regression analysis showed no evidence of interactions across different subgroups (such as route of corticosteroid, age, severity) on the outcome of mean time to onset of pain relief.

The effects of corticosteroids on resolution of pain were most apparent in the initial 24 hours, which implies that a single dose of corticosteroids may be sufficient.

Limitations

Our analysis had some limitations. Firstly, all of the included trials provided antibiotics to patients in both corticosteroid and placebo groups. Therefore, we do not know the effects of corticosteroids on sore throat symptoms independent of antibiotics.

Secondly, significant heterogeneity occurred in some of our analyses; this was attributable mainly to one trial.[w7] However, our results remained robust to the removal of this trial.

Thirdly, the outcome measure of mean time to onset of pain relief was limited by recall bias, because the estimation of the time when pain relief begins relies on patients' subjective recall and recording.

Finally, the limited number of trials meant that we were unable to assess publication bias. Included studies were also underpowered to detect rare adverse effects of corticosteroid therapy, as well as relapse rates and days missed from work or school.

Implications for practice

We could not fully assess the best type, route, or dosing regimen of corticosteroids because of small sample sizes. Two studies which directly compared intramuscular and oral routes found no differences, and our subgroup comparison also showed no differences.[14 w6]

Most of the trials were performed in North America, and additional trial data are warranted in European populations before the results can be deemed generalisable. We are also unsure of the benefits of corticosteroids in children because of the limitations on the reporting in these trials.

Recommendations for research

Further research should target corticosteroid use in antibiotic-naïve patients. Further trials in children are warranted. Trials should report the effects of corticosteroids on rates of antibiotic prescription as well as longer-term measures such as reattendance with recurrent sore throats.

We thank Nia Roberts, Ed Diggines, and Emma Meats for their assistance with literature searching and organisation.

Contributors: See bmj.com.

Funding: Funding for this work was provided in part by a grant from the British Society for Antimicrobial Chemotherapy Systematic Review Grant (GA722SRG). The Department of Primary Health care is part of the NIHR School of Primary Care Research. The study sponsors had no role in the study design; in the collection, analysis, or interpretation of data; in the writing of the report; nor in the decision to submit the article for publication.

Competing interests: None.

Ethical approval: Not required.

1 Del Mar CB, Glasziou PP, Spinks AB. Antibiotics for sore throat. *Cochrane Database Syst Rev* 2006;(4):CD000023.
2 NICE guideline. Respiratory Tract Infections - antibiotic prescribing. Prescribing of antibiotics for self limiting respiratory tract infections in adulst and children in primary care. 2008. National Institute for Health and Clinical Excellence. Ref Type: Report.
3 Linder JA, Stafford RS. Antibiotic treatment of adults with sore throat by community primary care physicians: a national survey, 1989-1999. *JAMA* 2001;286(10):1181-6.
4 Standing Medical Advisory Committee S-GoAR. The path of least resistance. London: Department of Health, 1998. Ref Type: Generic.
5 Little P, Gould C, Williamson I, Warner G, Gantley M, Kinmonth AL. Reattendance and complications in a randomised trial of prescribing strategies for sore throat: the medicalising effect of prescribing antibiotics. *BMJ* 1997;315:350-2.
6 Thomas M, Del MC, Glasziou P. How effective are treatments other than antibiotics for acute sore throat? *Br J Gen Pract* 2000;50:817-20.
7 Howie JG, Foggo BA. Antibiotics, sore throats and rheumatic fever. *J R Coll Gen Pract* 1985;35:223-4.
8 Petersen I, Johnson AM, Islam A, Duckworth G, Livermore DM, Hayward AC. Protective effect of antibiotics against serious complications of common respiratory tract infections: retrospective cohort study with the UK General Practice Research Database. *BMJ* 2007;335:982.
9 Mygind N, Nielsen LP, Hoffmann HJ, Shukla A, Blumberga G, Dahl R, et al. Mode of action of intranasal corticosteroids. *J Allergy Clin Immunol* 2001;108(1 Suppl):S16-25.
10 Zalmanovici A, Yaphe J. Steroids for acute sinusitis. *Cochrane Database Syst Rev* 2007;(2):CD005149.
11 Russell K, Wiebe N, Saenz A, Ausejo SM, Johnson D, Hartling L et al. Glucocorticoids for croup. *Cochrane Database Syst Rev* 2004;(1):CD001955.
12 Candy B, Hotopf M. Steroids for symptom control in infectious mononucleosis. *Cochrane Database Syst Rev* 2006;3:CD004402.
13 Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003;327:557-60.
14 Marvez-Valls EG, Stuckey A, Ernst AA. A randomized clinical trial of oral versus intramuscular delivery of steroids in acute exudative pharyngitis. *Acad Emerg Med* 2002;9:9-14

Accepted: 23 March 2009

Alarm symptoms and identification of non-cancer diagnoses in primary care: cohort study

Roger Jones,[1] Judith Charlton,[2] Radoslav Latinovic,[2] Martin C Gulliford[2]

[1]Department of General Practice and Primary Care, Division of Health and Social Care Research, King's College London School of Medicine at Guy's, King's College and St Thomas' Hospitals, London SE11 6SP

[2]Department of Public Health Sciences, Division of Health and Social Care Research, King's College London School of Medicine at Guy's, King's College and St Thomas' Hospitals, London SE1 3QD

Correspondence to: R Jones
roger.jones@kcl.ac.uk

Cite this as: *BMJ* 2009;339:b3094
doi: 10.1136/bmj.b3094

ABSTRACT

Objective To evaluate the predictive value of alarm symptoms for specified non-cancer diagnoses and cancer diagnoses in primary care.

Design Cohort study using the general practice research database.

Setting 128 general practices in the UK contributing data, 1994-2000.

Participants 762 325 patients aged 15 or older.

Main outcome measures Up to 15 pre-specified, non-cancer diagnoses associated with four alarm symptoms (haematuria, haemoptysis, dysphagia, rectal bleeding) at 90 days and three years after the first recorded alarm symptom. For each outcome analyses were implemented separately in a time to event framework. Data were censored if patients died, left the practice, or reached the end of the study period.

Results We analysed data on first episodes of haematuria (11 108), haemoptysis (4812), dysphagia (5999), or rectal bleeding (15 289). Non-cancer diagnoses were common in patients who presented with alarm symptoms. The proportion diagnosed with either cancer or non-cancer diagnoses generally increased with age. In patients presenting with haematuria, the proportions diagnosed with either cancer or non-cancer diagnoses within 90 days were 17.5% (95% confidence interval 16.4% to 18.6%) in women and 18.3% (17.4% to 19.3%) in men. For the other symptoms the proportions were 25.7% (23.8% to 27.8%) and 24% (22.5% to 25.6%) for haemoptysis, 17.2% (16% to 18.5%) and 22.6% (21% to 24.3%) for dysphagia, and 14.5% (13.7% to 15.3%) and 16.7% (15.8% to 17.5%) for rectal bleeding.

Conclusion Clinically relevant diagnoses are made in a high proportion of patients presenting with alarm symptoms. For every four to seven patients evaluated for haematuria, haemoptysis, dysphagia, or rectal bleeding, relevant diagnoses will be identified in one patient within 90 days.

WHAT IS ALREADY KNOWN ON THIS TOPIC

Certain symptoms, such as haematuria, haemoptysis, dysphagia, and rectal bleeding, are generally regarded as "red flags" because of their association with serious disease

The predictive value of these red flag or alarm symptoms for a diagnosis of cancer have now been established, but little is known about their predictive value for non-cancer diagnoses, which might also have considerable implications for patients' health

WHAT THIS STUDY ADDS

In patients with haemoptysis, haematuria, dysphagia, and rectal bleeding around one in five have an associated diagnosis at 90 days and approaching half of all patients at three years

The "number needed to evaluate" to identify an associated diagnosis in this group of patients is between four and seven

Patients presenting with these symptoms merit timely investigation for non-cancer diagnoses and potential cancer diagnoses, rather than a policy of watchful waiting

INTRODUCTION

General practitioners have the often difficult task of separating the minority of patients whose symptoms could indicate serious disease, and who require urgent diagnostic attention, from the majority with less serious, self limiting illness.[1] There are few studies in primary care that provide accurate information about the predictive value of common symptoms.

Previously we used the general practice research database to study the incidence of cancers in patients presenting in primary care with four "alarm symptoms"—haematuria, dysphagia, haemoptysis, and rectal bleeding.[2] We now report on the incidence of a range of pre-specified non-cancer diagnoses and provide predictive values for these diagnoses when associated with alarm symptoms.

METHODS

The methods have been described in our previous report.[2] We selected all 128 general practices that provided up to standard data from 1 January 1994 to 31 December 2000 and whose data were exclusively Read coded. We included only patients who had a first ever alarm symptom, with a complete date, recorded between 1 January 1995 and 31 December 2000, and whose data were up to standard at the date of the symptom. We constructed a list of pre-specified, potentially important diagnoses (conditions that generally require treatment or are likely to be progressive, or both) for each of the alarm symptoms: haematuria; haemoptysis; dysphagia; and rectal bleeding. See bmj.com for the lists of diagnoses.

Analyses were implemented in a time to event framework. Separate analyses were conducted for each outcome. The start date was the date of the first consultation for the alarm symptom. The end date was the date of the first recorded outcome event. Data were censored if patients left the practice or died. We estimated the proportion in whom the outcome was recorded before 90 days and three years; the former represents an upper limit of time in which a practitioner might aim to make a diagnosis after presentation, while three years might represent an upper limit of time during which serious clinical diagnoses would become evident. An individual patient could have a diagnosis of more than one of the outcome events. Tests for trend by age group were implemented with the log rank test.

This article is an abridged version of a paper that was published on bmj.com. Cite this article as: *BMJ* 2009;339:b3094

RESULTS

The study population consisted of 762 325 eligible patients aged 15 and older registered with 128 practices in 1994. We examined diagnoses made after the first occurrence of alarm symptoms in patients with no previous diagnosis of our specified conditions. We identified 11 108 first occurrences of haematuria, 4812 of haemoptysis, 5999 of dysphagia, and 15 289 of rectal bleeding between 1 January 1995 and 31 December 2000.[2] The mean age in years at first symptom was 58.5 (SD 18.9) for haematuria, 61.6 (18.0) for dysphagia, 54.5 (19.4) for haemoptysis, and 52.5 (18.8) for rectal bleeding. The figure shows the proportion of patients free of any of the selected outcomes by time since their first alarm symptom. Patients ceased to be at risk if they were diagnosed with one or more of the outcomes, died, left the practice, or reached the end of the study.

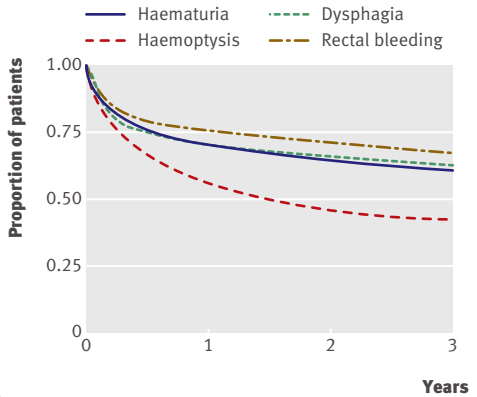
In patients presenting with macroscopic haematuria, 17.5% of women and 18.3% of men had one of the pre-specified diagnoses at 90 days (see bmj.com). At three years these figures rose to 42.0% and 36.6%, respectively, with cystitis and urinary tract infection being the commonest diagnosis in men and women at three years, followed by urinary tract cancers (8.0% in men) and benign prostatic hypertrophy (7.3%) in men and menstrual disorders (8.5%) in women. Urinary tract cancers were less common in women at three years (3.7%), with a further 0.4% being diagnosed with uterine cancer. Orchitis was reported in 2.6% of men. Renal calculi were reported in 3.8% of men and 1.5% of women. Although the event rates were similar across the three age ranges studied in women, there was a clear age gradient in men, with significantly higher event rates in men over the age of 64 (χ^2 test for trend: P=0.022 for women, P<0.001 for men).

Acute lower respiratory infection was the most common diagnosis in men with haemoptysis (10.2% at 90 days and 30.3% at three years) (see bmj.com). In women with haemoptysis the most common diagnosis was acute upper respiratory infection (10.6% and 47.4%, respectively). At 90 days the prevalence of a diagnosis of chronic obstructive pulmonary disease was 2.7% in women and 2.5% in men, with corresponding rates for asthma of 4.5% and 2.5%. Tuberculosis was rare, with rates of only 0.3% in women and 0.5% in men at three years. The event rates were related to age at 90 days in both men and women, and in men at three years, although event rates were fairly evenly distributed across age groups in women at this time (χ^2 test for trend: P=0.103 for women, P<0.001 for men).

In the patients with dysphagia, 22.6% of men and 17.2% of women had received a definite diagnosis at 90 days (see bmj.com). The commonest diagnosis in both men and women was oesophagitis (7.1% and 5.4%, respectively), followed by hiatus hernia (4.6% and 4.8%). Oesophageal stricture was diagnosed in only 2.9% of men and 1.7% of women. At three years the rate of important diagnoses had risen to 39.4% in men and 33.6% in women, with similar rank ordering, although disorders of the stomach were diagnosed in 11.4% of men and 11.9% of women. Most important diagnoses were more commonly identified in older men and women (χ^2 test for trend: P<0.001 for men and women).

In patients with rectal bleeding, 16.7% of men and 14.5% of women received a diagnosis at 90 days, and at three years 32.3% of men and 32.4% of women had a clear diagnosis (see bmj.com). The most common diagnosis made at both times was haemorrhoids (10.0% in men and 7.8% in women at 90 days and 19.0% and 16.8% at three years), followed by anal fissure and diverticulitis. New diagnoses of Crohn's disease and ulcerative colitis were made in less than 1% of women at 90 days, with these figures rising to around 1% for Crohn's disease and 2% for ulcerative colitis in both men and women at three years.

Taking these results together, the "number needed to evaluate" to make a clinically relevant diagnosis as a result of investigation, is between four and seven for the four alarm symptoms studied.



No at risk

Haematuria	11 106	6755	5156	3845
Haemoptysis	4812	2336	1627	1215
Dysphagia	5998	3596	2745	2010
Rectal bleeding	15 289	9604	7140	5059

Proportion of patients not diagnosed with any of the selected outcomes by time since first alarm symptom

DISCUSSION

We investigated the incidence rates of non-cancer diagnoses in primary care in patients who presented with four important alarm symptoms. Within 90 days of presentation, an associated diagnosis had been made in around 15.3% with rectal bleeding, 19.3% with dysphagia, 17.8% with haematuria, and 24.4% with haemoptysis. Most of these diagnoses could have been made clinically or confirmed by performing simple investigations such as urine culture, chest radiology, and upper or lower gastrointestinal endoscopy. Over the three year follow-up these figures rose to about 28.9%, 33.0%, 35.5%, and 53.9%, respectively. After three years' follow-up, over three quarters of patients presenting with rectal bleeding did not have a definite diagnosis, with comparable figures of about 67% for dysphagia, 64% for haematuria, and 46% for haemoptysis.

Strengths and limitations

The strengths include the large representative population of patients studied, the accuracy of the data contained in the database, and the ability to identify enough patients to draw valid conclusions. Several studies have evaluated the validity of diagnoses recorded in the general practice research database with generally satisfactory results. We have noted previously that our results show some sensitivity to the scope of case definitions. Limitations include the lack of clinical contextual detail concerning the individual alarm symptoms. We rely on symptom recording rather than symptom reporting. We do not know whether these diagnoses were made on the basis of investigations. We think it unlikely that analysis of a more recent dataset, collected after the introduction of the Quality and Outcomes Framework, might have changed our results because the framework encourages accurate documentation of chronic disease management rather than of acute disease presentation. Although the 90 day follow-up is likely to reflect diagnostic outcomes of single episodes of the presentation and investigation of alarm symptoms, we do not have information about clinical events and relevant interventions taking place during the three year follow-up.

Comparison with other studies

Our figures for diagnostic rates in patients with haematuria are similar to a study from Belgium[3] but are slightly lower than another study.[4] Given the rates of diagnosis of urinary tract cancer and lung cancer, both at 8% for men at three years, it is important to pursue an infective or neoplastic cause in patients presenting in primary care with these symptoms. The diagnostic rates in our patients with haemoptysis are much lower than those emerging from studies in secondary care,[5 6] but no primary care based studies of the causes of haemoptysis in the general population have been published.

The predictive value of dysphagia for a serious organic lesion in the oesophagus has recently been called into question.[7 8] Our data suggest that dysphagia should be taken seriously, particularly in older patients.

Diagnosis rates after presentation with rectal bleeding were the lowest of the four alarm symptoms, and most patients turned out to have haemorrhoids or an anal fissure. It is important to consider the pattern of bleeding and accompanying symptoms, which might increase the likelihood of cancer.[9]

Even in our analysis of outcomes in single alarm symptoms up to one in five patients had a diagnosis at 90 days, and this proportion would almost certainly have been higher in patients with multiple symptoms.

Conclusions

We have extended the concept of "alarm symptoms" to include important non-cancer diagnoses, emphasising that these symptoms are not only red flags for malignancy but also "yellow flags" that should prompt clinicians to conduct investigations or intervene therapeutically in these benign but potentially serious disorders. For many of these patients the test of time should probably be replaced by a "timely test," although investigation will vary according to the resources available.

Data for this study were provided by EPIC, UK, a licence holder of an historical part of the general practice research database dataset.

Contributors: See bmj.com.

Funding: The study was funded by Division of Health and Social Care Research, King's College London.

Competing interests: None declared.

Ethical approval: The proposal was approved by the Scientific and Ethical Advisory Group of the GPRD (protocol number 785R).

1 Marinker M. Looking and leaping. In: Marinker M, Peckham M, eds. *Clinical futures*. London: BMJ Publishing, 1998.
2 Jones R, Latinovic R, Charlton J, Gulliford M. Alarm symptoms in early diagnosis of cancer in primary care: cohort study using general practice research database. *BMJ* 2007;334:1040-8.
3 Bruyninckx R, Buntinx F, Aertgeerts B, Van Casteren V. The diagnostic value of macroscopic haematuria for the diagnosis of urological cancer in general practice. *Br J Gen Pract* 2003;53:31-5.
4 Summerton N, Mann S, Rigby AS, Ashley J, Palmer S, Hetherington JW. Patients with new onset haematuria: assessing the discriminant value of clinical information in relation to urological malignancies. *Br J Gen Pract* 2002;52:284-9.
5 Santiago S, Tobias J, Williams AJ. A reappraisal of the causes of haemoptysis. *Arch Intern Med* 1991;151:1461-5.
6 Johnston H, Reisz G. Changing spectrum of haemoptysis: underlying causes in 148 patients undergoing diagnostic flexible fiberoptic bronchoscopy. *Arch Intern Med* 1989;149:1666-8.
7 Vakil N, Moayyedi P, Fennerty MB, Talley NJ. Limited value of alarm features in the diagnosis of upper gastrointestinal malignancy: systematic review and meta-analysis. *Gastroenterology* 2006;131:390-401.
8 Vakil N, van Zanten SV, Kahrilas P, Dent J, Jones R, the Global Consensus Group. The Montreal definition and classification of gastroesophageal reflux disease: a global evidence-based consensus. *Am J Gastroenterol* 2006;101:1900-20.
9 Robertson R, Campbell C, Weller DP, Elton R, Mant D, Primrose J, et al. Predicting colorectal cancer risk in patients with rectal bleeding. *Br J Gen Pract* 2006;56:763-7.

Accepted: 31 March 2009

Laughing gas: the best medicine?

While performing a day case laparoscopy, we asked our medical students what was being done. The first student correctly replied that gas was being inserted to insufflate the abdomen. We asked his colleague which gas was being used. After a moment's pause came the tentative reply, "Helium?"

Later on, when detaching the gas tube to finish the operation, we were amicably scolded by the scrub nurse for not turning off the gas promptly. Baffled, we asked her why this was urgent. She explained passionately that, by leaking the gas into the theatre, everyone would be killed as they would inhale the fumes.

Completely bewildered, we asked what gas we were using for insufflation. She solemnly declared, "Carbon monoxide."

Shreelata Datta specialist registrar, Medway Maritime Hospital, London statta@doctors.org.uk
David J Houghton consultant, Medway Maritime Hospital, London

Cite this as: *BMJ* 2008;337:a570

Novel methods to deal with publication biases: secondary analysis of antidepressant trials in the FDA trial registry database and related journal publications

Santiago G Moreno,[1] Alex J Sutton,[1] Erick H Turner,[2] Keith R Abrams,[1] Nicola J Cooper,[1] Tom M Palmer,[3] A E Ades[4]

EDITORIAL by Dubben

[1]Department of Health Sciences, University of Leicester, Leicester LE1 7RH
[2]Department of Psychiatry, Oregon Health and Science University, Portland Veterans Affairs Medical Center, Portland, Oregon, USA
[3]MRC Centre for Causal Analyses in Translational Epidemiology, Department of Social Medicine, University of Bristol
[4]Department of Community Based Medicine, University of Bristol
Correspondence to: S G Moreno
sgm8@le.ac.uk

Cite this as: *BMJ* 2009;339:b2981
doi: 10.1136/bmj.b2981

ABSTRACT

Objective To assess the performance of novel contour enhanced funnel plots and a regression based adjustment method to detect and adjust for publication biases.

Design Secondary analysis of a published systematic literature review.

Data sources Placebo controlled trials of antidepressants previously submitted to the US Food and Drug Administration (FDA) and matching journal publications.

Methods Publication biases were identified using novel contour enhanced funnel plots, a regression based adjustment method, Egger's test, and the trim and fill method. Results were compared with a meta-analysis of the gold standard data submitted to the FDA.

Results Severe asymmetry was observed in the contour enhanced funnel plot that appeared to be heavily influenced by the statistical significance of results, suggesting publication biases as the cause of the asymmetry. Applying the regression based adjustment method to the journal data produced a similar pooled effect to that observed by a meta-analysis of the FDA data. Contrasting journal and FDA results suggested that, in addition to other deviations from study protocol, switching from an intention to treat analysis to a per protocol one would contribute to the observed discrepancies between the journal and FDA results.

Conclusion Novel contour enhanced funnel plots and a regression based adjustment method worked convincingly and might have an important part to play in combating publication biases.

INTRODUCTION

In 2008 Turner et al published a study showing that the journal literature on antidepressants was biased towards "favourable" results.[1] The authors compared the results in journal reports with data on the corresponding trials submitted to the US Food and Drug Administration (FDA) when applying for licensing. The discrepancies in the journal reports were due to publication biases. Although the term publication bias has been used to refer to the suppression of whole studies based on statistical significance, a range of mechanisms can distort the published literature. If such publication biases are present, any decision making based on them could be misleading,[2 3] not least through inflated clinical effects from meta-analysis.[4]

The FDA dataset on antidepressants is regarded as a gold standard source owing to the legal requirements of submitting entire evidence to the FDA and its careful monitoring for deviations from protocol.[5-7] In the absence of a gold standard dataset, meta-analysts have relied on analytical methods to detect and adjust for publication biases. While the performance of many of these has been evaluated in simulation studies, concerns remain as to whether simulations reflect real life situations. This has led to caution in the use of methods, particularly for those that adjust effect sizes for publication biases [3]; but ultimately this is what is required for rational decision making if publication biases exist.

We consider two recently described methods for identifying and adjusting for publication biases: a funnel plot enhanced by contours separating areas of statistical significance from non-significance.[8] These contours help distinguish publication biases from other factors that lead to asymmetry in the plot. The method used to adjust a meta-analysis for publication bias is based on a regression line fitted to the funnel plot.[13] We also consider established methods to deal with publication bias: the regression based Egger's test for funnel asymmetry[9] and the trim and fill method,[10] which adjusts a meta-analysis for publication bias by imputing studies to rectify asymmetry in the funnel plot. We present the results of applying the diagnostic and adjustment methods to the journal data and compare the findings with those obtained through analysis of the FDA data.

WHAT IS ALREADY KNOWN ON THIS TOPIC

Publication biases exaggerate clinical effects resulting in potentially erroneous clinical decision making

While most of the attention has focused on the non-publication of whole studies, the problem of reporting biases within published studies is receiving increased attention

WHAT THIS STUDY ADDS

Mechanisms including suppression of whole studies, selective outcome reporting, and data "massaging" (for example, selective exclusion of patients from the analysis) may act simultaneously, but may be motivated by underlying statistical significance

Contour enhanced funnel plots and a regression based adjustment method to identify and adjust for multiple publication biases using real data where a gold standard exists showed promising results

METHODS

A full description of the dataset is published elsewhere.[1] Briefly, Turner et al identified the cohort of all phase II and phase III short term double blind placebo controlled trials used for the licensing of antidepressants during 1987-2004 by the FDA. Seventy four trials involving 12 drugs and 12 564 patients were identified. To compare drug efficacy reported by the journals with that of the FDA gold standard, Turner et al collected data on the primary outcome from both. After extracting the data from the FDA trial registry,

This article is an abridged version of a paper that was published on bmj.com. Cite this article as: *BMJ* 2009;339:b2981

they searched the published literature for publications matching the same trials. When a match was identified, they extracted data on the article's apparent primary efficacy outcome. Because studies reported their outcomes on different scales, they expressed effect sizes as standardised mean differences using Hedges' *g* scores (with corresponding variances).[11] Among the 74 studies, 23 (31%), with 3449 participants, were not published. Overall, larger effects were derived from the journal data than from the FDA data. Among the 38 studies with results viewed by the FDA as statistically significant, only one was unpublished. Conversely, inconclusive studies were, with three exceptions, not published (22 studies) or published in conflict with the FDA findings (11 studies). Moreover, 94% of published studies reported a positive significant result for their primary outcome, compared with 51% according to the FDA. Data for the analysis were extracted from the previous paper,[1] in which two studies were combined, totalling 73 studies in our assessment.

Analysis

We applied two methods to the journal dataset: the contour enhanced funnel plot[8 12] to detect publication biases, and a regression based adjustment method[13] to adjust for them. We also applied the most commonly used methods to deal with publication biases—Egger's regression test[9] for detecting bias, and the trim and fill adjustment method,[10 14-16] which adjusts for publication bias by imputing studies estimated to be missing from the dataset. We use fixed effect models in our analysis.

Contour enhanced funnel plots

A funnel plot is a scatter plot of study effect sizes against their standard errors.[17] When no bias is present the plot should be symmetrical, with increasing variability in effect sizes being observed in the less precise studies towards the bottom of the plot. Asymmetry may indicate publication biases through the lack of observed data points in a region of the plot.[17] Asymmetry alone does not necessarily imply publication biases exist, however, since alternative explanations for the asymmetry may be present.[18]

To distinguish publication biases from other causes of funnel asymmetry, the plot can be enhanced by contours partitioning it into areas of statistical significance and non-significance[8 12] based on the standard Wald test, marking milestones of significance, such as the 1%, 5%, and 10% levels.[19] Thus the level of statistical significance of every study's effect estimate is identified. Since there is evidence that publication biases are related to these milestones,[20 21] this can aid interpretation of the plot—if studies seem to be missing in areas of statistical non-significance, then this adds credence to the notion that the asymmetry is due to publication biases. In such cases an attempt should be made to adjust for such biases. Conversely, if the parts of the funnel where studies are perceived to be missing are in areas of higher statistical significance, the cause of asymmetry is more likely to be due to factors other than publication biases.

Regression based adjustment

The regression based adjustment method fits a regression line of best fit to the data on a funnel plot.[22] An adjusted pooled estimate of effect is obtained by predicting from the regression line the pooled effect size for an ideal study of infinite size (with zero standard error), which would be located at the top of the plot; since it is hypothesised that there would be no bias in studies of that size. The performance of several different regression models has been considered over a range of meta-analytical and publication bias scenarios. The best models consistently outperformed the trim and fill method. One, the quadratic version of the Egger's regression test,[9] is implemented here. This assumes a linear trend between the effect size and its variance (rather than its standard error, as assumed in the original Egger's test).

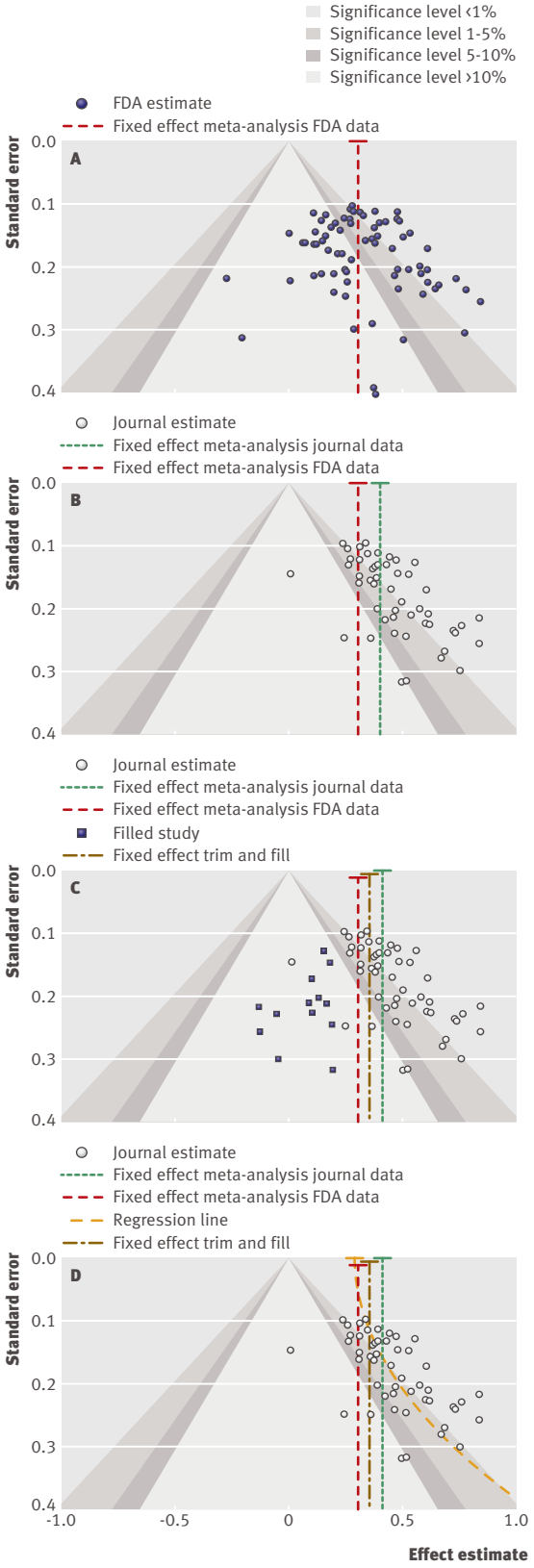
RESULTS

Figure A displays a contour enhanced funnel plot of the FDA studies, with the corresponding fixed effect meta-analysis pooled estimate providing a weighted average of effect sizes across trials (*g* score 0.31, 95% confidence interval 0.27 to 0.35). This plot is reasonably symmetrical (Egger's test P=0.10), which is consistent with the hypothesis that the FDA is an unbiased data source.

The contour enhanced funnel plot for the journal data (fig B) is highly asymmetrical (Egger's test P<0.001). A meta-analysis of these data results in a higher average effect size (*g* score 0.41, 0.37 to 0.45). Most of the estimates now lie above the right contour, indicating a statistically significant benefit at the 5% level. The area where studies seem to be "missing" is within the area where non-significant studies would be located; inside the triangle defined by P=0.10 contour boundaries. This adds further credence to the asymmetry being caused by publication biases. Hence, even without the availability of the corresponding plot for the FDA data (fig A), a contour enhanced one has identified publication biases as a major problem for the journal data.

For the journal dataset, the trim and fill method imputed a total of 18 "missing" studies, all in the region of non-statistical significance (squares in figure C). This agrees reasonably well with the truth, as 23 studies identified through the FDA registry were not identified in the journal literature. The application of the trim and fill method reduced the average effect size to 0.35 (95% confidence interval 0.31 to 0.39), about halfway between the FDA and journal estimates (fig C).

The fitted line corresponding to the regression based adjustment method is plotted in figure 1D. The adjusted estimate is obtained by extrapolating the line to where the standard error is 0 (top of figure). This produces an adjusted average effect size of 0.29 (95% confidence interval 0.23 to 0.35), close to the estimate produced by the meta-analysis of the FDA data (0.31, 0.27 to 0.35).



Contour enhanced funnel plots (95% confidence intervals at top of plots). (A) Studies submitted to US Food and Drug Administration (FDA). (B) Studies published in medical journals. (C) Implementation of trim and fill method on journal dataset. (D) Implementation of regression adjustment model on journal dataset (adjusted effect at top where standard error is 0)

DISCUSSION

The application of two approaches to identify and adjust for publication biases in a dataset derived from a journal publication, where a gold standard dataset exists, produced encouraging results. Detection of publication biases was convincing using a contour enhanced funnel plot, and the regression based method produced a corrected average effect size close to that obtained from the FDA dataset. We think the analysis presented here provides strong evidence that these novel methods have a useful role.

This assessment does have limitations. Firstly, the findings relate to one dataset and are not necessarily generalisable to other examples. All the trials were sponsored by the pharmaceutical industry and we assume that the FDA data are completely unbiased. Furthermore, the evaluated methods were designed to assess efficacy outcomes and might not be appropriate for safety outcomes.

Recently there has been a lot of research into refining tests for funnel plot asymmetry,[9 23-26] and while we support the formalisation of such an assessment, none of the tests (nor trim and fill or the regression adjustment method) considers the statistical significance of the study estimates. For this reason we think the consideration of the contours on the funnel plot to be an essential component of distinguishing publication biases from other causes of funnel plot asymmetry. We make no claim that the contours can distinguish between the different mechanisms for publication bias—for example, missing whole studies, selectively reported outcomes, or "massaged" data. But we do not think this is an important limitation because all these biases have the same effect in a meta-analysis. There is empirical evidence to support this notion for the effect of reporting biases within published clinical trials in general[27-29] and for trials on antidepressants in particular.[1 30 31] Potential mechanisms that are known to induce this include: (*a*) selectivity in which outcomes are reported or labelled as primary in journal publications; (*b*) post hoc searches for statistical significance using numerous hypothesis tests (data dredging or fishing); and (*c*) selectivity in the analysis methods applied to the journal data. Regarding the last point, the FDA makes its recommendations based on the intention to treat principle,[32 33] whereas only half the journal publications are analysed and reported using this approach.[34-37] The usual alternative—the per protocol approach—excludes dropouts and non-adherents (or patients with protocol deviations in general) and aims to estimate drug efficacy, which will tend to inflate effect sizes compared with the intention to treat approach, which estimates effectiveness.[38-41] An estimate from a per protocol analysis will generally have less precision than for the associated intention to treat analyses owing to the removal of patients with protocol deviations,[42 43] which would result in a shift downwards along the y axis of a funnel plot.

Few methods for specifically addressing outcomes[44 45] and subgroup reporting biases[46] exist. It is reassuring that the methods used in this article to address publication and related biases generally seem to work well in

the presence of multiple types of publication biases. We no longer advocate the use of the trim and fill method because of problems identified through simulation studies.[13 47 48] The regression adjustment method[49] consistently outperformed the trim and fill method in an extensive simulation study[13] and in this dataset.

We consider technical issues relating the influence of choice of outcome metric on the robustness of the results, and analyses methods used within the assessments. Firstly, the Hedges' *g* score outcome metric was used throughout the analysis. This includes a correction for small sample size. An alternative metric, without the correction, is the Cohen's *d* score, which could also have been used. However this would have negligible influence on the funnel plots presented here since the correction is still modest even for the smallest trials (n=25). An additional consideration is that the contours on the funnels are constructed assuming normality of the effect size since they are based on the Wald test. This may not be exactly the statistical test used in the original analyses for some of the trials. For example, for trials with small sample sizes, a *t* test may have been used. However, as the Wald and *t* test statistics converge as the sample size increases, this is only going to affect the assessment of the most imprecise trials at the bottom of the funnel, and all our findings are clearly robust to this.

The 73 randomised controlled trials considered here correspond to 12 different antidepressants. Despite this, there was little statistical heterogeneity in both datasets and so we carried out fixed effect analyses for simplicity. There is an ever present tension in meta-analysis between "lumping and splitting" studies, and an argument could be made for allowing for specific differences in drug treatment by stratifying them and carrying out 12 separate analyses. Challenges would arise if attempting to detect and adjust for publication biases in each of the analyses independently owing to the difficulty of interpreting funnel plots with small numbers of studies and the limited power of statistical methods.[23]

Undoubtedly the best solution to publication biases is prevention.[50] Using a gold standard data source is one way of achieving this. As this is a long way off from becoming a reality for many analyses we often have to rely on analytical methods to deal with the problem. We believe that the contour enhanced funnel plot and the regression based adjustment method are important developments to combat publication biases.

Contributors: See bmj.com.

Funding: SGM was supported by a Medical Research Council Health Services Research Collaboration studentship in the UK. AEA was funded by the Medical Research Council Health Services Research Collaboration. The funding agreement ensured the authors' independence in designing the study, interpreting the data, and writing and publishing the report. The corresponding author as well as the other authors had access to all the data and take responsibility for the integrity of the data and the accuracy of the data analysis.

Competing interests: None declared.

Ethical approval: Not required.

Data sharing: Data are available on request from the first author.

1 Turner EH, Matthews AM, Linardatos E, Tell RA, Rosenthal R. Selective publication of antidepressant trials and its influence on apparent efficacy. *N Engl J Med* 2008;358:252-60.
2 Egger M, Smith GD. Misleading meta-analysis [editorial]. *BMJ* 1995;310:752-4.
3 Sterne JAC, Egger M, Davey Smith G. Systematic reviews in health care: investigating and dealing with publication and other biases in meta-analysis. *BMJ* 2001;323:101-5.
4 Egger M, Davey Smith G, Phillips AN. Meta-analysis: principles and procedures. *BMJ* 1997;315:1533-7.
5 Ioannidis JPA. Effectiveness of antidepressants: an evidence myth constructed from a thousand randomized trials? *Philos Ethics Humanit Med* 2008;3:14.
6 Chan A-W. Bias, spin, and misreporting: time for full access to trial protocols and results. *PLoS Med* 2008;5:e230.
7 Turner EH. A taxpayer-funded clinical trials registry and results database. *PLoS Med* 2004;1:e60.
8 Peters J, Sutton AJ, Jones DR, Abrams KR, Rushton L. Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry. *J Clin Epidemiol* 2008:991-6.
9 Egger M, Smith GD, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997;315:629-34.
10 Duval S, Tweedie RL. Trim and fill: a simple funnel plot based method of testing and adjusting for publication bias in meta-analysis. *Biometrics* 2000;56:455-63.
11 Hedges LV. Estimating effect size from a series of independent experiments. *Psychol Bull* 1982;92:490-9.
12 Palmer TM, Peters JL, Sutton AJ, Moreno SG. Contour-enhanced funnel plots for meta-analysis. *Stata J* 2008;8:242-54.
13 Moreno SG, Sutton AJ, Ades AE, Stanley TD, Abrams KR, Peters JL, et al. Assessment of regression-based methods to adjust for publication bias through a comprehensive simulation study. *BMC Med Res Methodol* 2009;9:2.
14 Sutton AJ, Duval SJ, Tweedie RL, Abrams KR, Jones DR. Empirical assessment of effect of publication bias on meta-analyses. *BMJ* 2000;320:1574-7.
15 Duval S, Tweedie RL. A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. *J Am Stat Assoc* 2000;95:89-98.
16 Steichen TJ. METATRIM: stata module to perform nonparametric analysis of publication bias. *Stata Tech Bull* 2000;61:8-14.
17 Sterne JAC, Egger M. Funnel plots for detecting bias in meta-analysis: guidelines on choice of axis. *J Clin Epidemiol* 2001;54:1046-55.
18 McMahon B, Holly L, Harrington R, Roberts C, Green J. Do larger studies find smaller effects? The example of studies for the prevention of conduct disorder. *Eur Child Adolesc Psychiatry* 2008;17:432-7.
19 Gerber AS, Malhotra N. Publication bias in empirical sociological research: do arbitrary significance levels distort published results? *Sociol Methods Res* 2008;37:3-30.
20 Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR. Publication bias in clinical research. *Lancet* 1991;337:867-72.
21 Ioannidis JPA. Effect of the statistical significance of results on the time to completion and publication of randomized efficacy trials. *JAMA* 1998;279:281-6.
22 Thompson SG, Higgins JPT. How should meta-regression analyses be undertaken and interpreted? *Stat Med* 2002;21:1559-73.
23 Sterne JAC, Gavaghan D, Egger M. Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature. *J Clin Epidemiol* 2000;53:1119-29.
24 Harbord RM, Egger M, Sterne JAC. A modified test for small-study effects in meta-analyses of controlled trials with binary endpoints. *Stat Med* 2006;25:3443-57.
25 Peters JL, Sutton AJ, Jones DR, Abrams KR, Rushton L. Comparison of two methods to detect publication bias in meta-analysis. *JAMA* 2006;295:676-80.
26 Rücker G, Schwarzer G, Carpenter J. Arcsine test for publication bias in meta-analyses with binary outcomes. *Stat Med* 2008;27:746-63.
27 Chan A-W, 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.
28 Chan A-W, Krleza-Jeric K, Schmid I, Altman DG. Outcome reporting bias in randomized trials funded by the Canadian Institutes of Health Research. *CMAJ* 2004;17:735-40.
29 Dwan K, Altman DG, Arnaiz JA, Bloom J, Chan A, Cronin E, et al. Systematic review of the empirical evidence of study publication bias and outcome reporting bias. *PLoS ONE* 2008;3:e3081.
30 Hotopf M, Barbui C. Bias in the evaluation of antidepressants. *Epidemiol Psichiatr Soc* 2005;14:55-7.
31 Furukawa TA, Watanabe N, Omori IM, Montori VM, Guyatt GH. Association between unreported outcomes and effect size estimates in Cochrane meta-analyses. *JAMA* 2007;297:468-70.
32 Heritier SR, Gebski VJ, Keech AC. Inclusion of patients in clinical trial analysis: the intention-to-treat principle. *Med J Aust* 2003;179:438-40.
33 Lewis JA. Statistical principles for clinical trials (ICH E9): an introductory note on an international guideline. *Stat Med* 1999;18:1903-42.
34 Gravel J, Opatrny L, Shapiro S. The intention-to-treat approach in randomized controlled trials: are authors saying what they do and doing what they say? *Clin Trials* 2007;4:350-6.

35 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.
36 Hotopf M, Lewis G, Normand C. Putting trials on trial—the costs and consequences of small trials in depression: a systematic review of methodology. *J Epidemiol Community Health* 1997;51:354-8.
37 Hollis S, Campbell F. What is meant by intention to treat analysis? Survey of published randomised controlled trials. *BMJ* 1999;319:670-4.
38 Gartlehner G, Hansen RA, Nissman D, Lohr KN, Carey TS. A simple and valid tool distinguished efficacy from effectiveness studies. *J Clin Epidemiol* 2006;59:1040-8.
39 Bollini P, Pampallona S, Tibaldi G, Kupelnick B, Munizza C. Effectiveness of antidepressants. Meta-analysis of dose-effect relationships in randomised clinical trials. *Br J Psychiatry* 1999;174:297-303.
40 Revicki DA, Frank L. Pharmacoeconomic evaluation in the real world. Effectiveness versus efficacy studies. *Pharmacoeconomics* 1999;15:423-34.
41 Schulz KF, Grimes DA. Sample size slippages in randomised trials: exclusions and the lost and wayward. *Lancet* 2002;359:781-5.
42 Porta N, Bonet C, Cobo E. Discordance between reported intention-to-treat and per protocol analyses. *J Clin Epidemiol* 2007;60:663-9.
43 Fergusson D, Aaron SD, Guyatt G, Hebert P. Post-randomisation exclusions: the intention to treat principle and excluding patients from analysis. *BMJ* 2002;325:652-4.
44 Williamson PR, Gamble C. Application and investigation of a bound for outcome reporting bias. *Trials* 2007;8:9.
45 Hutton JL, Williamson PR. Bias in meta-analysis due to outcome variable selection within studies. *Appl Stat* 2000;49:359-70.
46 Hahn S, Williamson PR, Hutton JL, Garner P, Flynn EV. Assessing the potential for bias in meta-analysis due to selective reporting of subgroup analyses within studies. *Stat Med* 2000;19:3325-36.
47 Terrin N, Schmid CH, Lau J, Olkin I. Adjusting for publication bias in the presence of heterogeneity. *Stat Med* 2003;22:2113-26.
48 Peters JL, Sutton JA, Jones DR, Abrams KR, Rushton L. Performance of the trim and fill method in the presence of publication bias and between-study heterogeneity. *Stat Med* 2007;26:4544-62.
49 Thompson SG, Sharp SJ. Explaining heterogeneity in meta-analysis: a comparison of methods. *Stat Med* 1999;18:2693-708.
50 Rothstein HR, Sutton AJ, Borenstein M. *Publication bias in meta-analysis. Prevention, assessment and adjustments*. Chichester: Wiley, 2005.

Accepted: 10 May 2009

Association between radiographic features of knee osteoarthritis and pain: results from two cohort studies

Tuhina Neogi,[1] David Felson,[1] Jingbo Niu,[1] Michael Nevitt,[2] Cora E Lewis,[3] Piran Aliabadi,[4] Burt Sack,[1] James Torner,[5] Lawrence Bradley,[3] Yuqing Zhang[1]

EDITORIAL by Pincus and Block

[1]Clinical Epidemiology Research and Training Unit, Boston University School of Medicine, 650 Albany Street, Boston, MA 02118, USA
[2]Department of Epidemiology and Biostatistics, University of California at San Francisco, San Francisco, CA 94107-1762, USA
[3]Division of Preventive Medicine, University of Alabama at Birmingham, Birmingham, AL 35205, USA
[4]Department of Radiology, Brigham and Women's Hospital, Boston, MA 02115, USA
[5]University of Iowa, College of Public Health, E107 General Hospital, Iowa City, IA 52242, USA
Correspondence to: T Neogi
tneogi@bu.edu

Cite this as: *BMJ* 2009;339:b2844
doi: 10.1136/bmj.b2844

This article is an abridged version of a paper that was published on bmj.com. Cite this article as: *BMJ* 2009;339:b2844

ABSTRACT

Objective To examine the relation of radiographic features of osteoarthritis to knee pain in people with knees discordant for knee pain in two cohorts.

Design Within person, knee matched, case-control study.

Setting and participants Participants in the Multicenter Osteoarthritis (MOST) and Framingham Osteoarthritis studies who had knee radiographs and assessments of knee pain.

Main outcome measures Association of each pain measure (frequency, consistency, and severity) with radiographic osteoarthritis, as assessed by Kellgren and Lawrence grade (0-4) and osteophyte and joint space narrowing grades (0-3) among matched sets of two knees within individual participants whose knees were discordant for pain status.

Results 696 people from MOST and 336 people from Framingham were included. Kellgren and Lawrence grades were strongly associated with frequent knee pain—for example, for Kellgren and Lawrence grade 4 *v* grade 0 the odds ratio for pain was 151 (95% confidence interval 43 to 526) in MOST and 73 (16 to 331) in Framingham (both P<0.001 for trend). Similar results were also seen for the relation of Kellgren and Lawrence scores to consistency and severity of knee pain. Joint space narrowing was more strongly associated with each pain measure than were osteophytes.

Conclusions Using a method that minimises between person confounding, this study found that radiographic osteoarthritis and individual radiographic features of osteoarthritis were strongly associated with knee pain.

WHAT IS ALREADY KNOWN ON THIS TOPIC

Little is understood about the causes of knee pain, as the association between radiographic osteoarthritis and knee pain has generally been accepted to be weak or modest
Previous studies comparing pain and structural abnormalities across patients may not have found a strong association because of between person confounding

WHAT THIS STUDY ADDS

Comparison of radiographic abnormalities between two knees within a person in whom pain measures in the two knees were discordant eliminated between person confounding
A strong association existed between radiographic osteoarthritis and knee pain, supporting a causative role for structural abnormalities in the presence of knee pain

INTRODUCTION

The general opinion is that only a modest association exists between radiographic features of osteoarthritis and knee pain.[1-3] Several investigators have shown discordance between these two features of osteoarthritis: people with abnormal joint radiographs may have no or only mild pain,[3 4] whereas others with pain may not have radiographic osteoarthritis.[5] Furthermore, although pain has been associated with osteophytes on plain radiographs,[6-9] it has generally not been associated with joint space narrowing.[6 8-10]

A factor can be strongly causally associated with an outcome and yet not be a strong predictor of the outcome on its own if several other factors contribute to the outcome. This is particularly relevant to the study of pain, which is a subjective experience. Genetic, psychological, and sociocultural factors all contribute to a person's response to pain.[11-21] These factors are often neither measured nor controlled for in studies examining the relation of pain to radiographic osteoarthritis across individual patients. Consequently, residual confounding may have diluted the association between radiographic knee osteoarthritis and knee pain.

We examined the relation of radiographic osteoarthritis to knee pain among participants who had knees that were discordant for pain (that is, one knee had pain but the other did not). We compared the presence of radiographic features between the naturally paired knees, eliminating confounders at the level of the participant.

METHODS

Study populations—The Multicenter Osteoarthritis (MOST) study is a prospective cohort study of 3026 people aged 50 to 79 years either with or at high risk of osteoarthritis, recruited from two communities in Iowa and Alabama, USA. The Framingham Osteoarthritis Study included members of the original cohort from the Framingham Heart Study, the Framingham Offspring Study, and a new cohort recruited from Framingham, Massachusetts.[22] Selection of participants was not based on the presence or absence of knee osteoarthritis or knee pain.

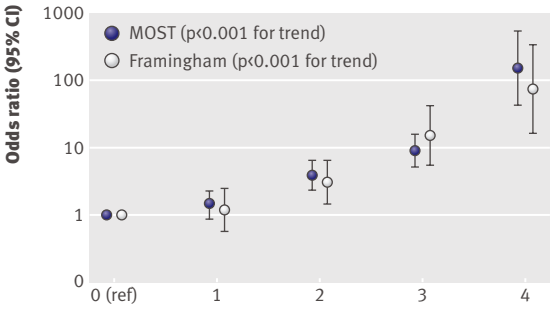
Radiographic assessment—All participants had radiographic evaluation of the knee. Radiographs were scored by a musculoskeletal radiologist and a rheumatologist blinded to pain status. Each knee joint was scored for Kellgren and Lawrence grade (0-4), maximal osteophyte grade (0-3), and maximal joint space narrowing grade (0-3).[23]

Assessment of pain—We used three measures to characterise knee pain: presence of frequent knee pain, consistency of frequent knee pain, and severity of knee pain, reported at the clinic visit. Frequent knee pain was defined as pain, aching, or stiffness in either knee on most of the previous 30 days. Consistency of frequent knee pain was assessed from the presence of frequent knee pain in a telephone screen before the clinic visit and during the clinic visit. Severity of knee pain was determined from the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) pain questionnaire (0-20 scale).[24]

Statistical analyses—We did separate analyses for MOST and Framingham for analysis of frequent knee pain. The other two pain measures were only assessed in MOST. We identified people who had knees that were discordant for each pain measure separately. Two such knees within a person then formed a matched set for a specific pain measurement. For severity of knee pain, we classified knees into three categories: severe to extreme pain, mild to moderate pain, and no pain. In separate analyses, we identified participants in whom both knees had some pain but one knee had greater pain severity than the other knee. We examined the relation of each radiographic measure of osteoarthritis to the prevalence of each of the pain measurements using conditional logistic regression.

RESULTS

We identified 696 people from MOST and 336 people from Framingham as having knees discordant for presence of frequent knee pain. For these participants in MOST, 418 (60%) were female, the mean age was 62 (SD 8, range 50-79) years, and the mean body mass index was 31 (6, 18-56). In Framingham, 208 (62%) of these participants were female, the mean age was 68 (SD 10, range 49-93), and the mean body mass index was 29 (5, 17-58).



MOST (n=696)					
Case	210	96	95	177	108
Control	300	162	96	119	19
Framingham (n=336)					
Case	162	28	51	61	34
Control	205	40	54	32	5

Associations of frequent knee pain with Kellgren and Lawrence (KL) grade among people with two knees discordant for frequent knee pain status. No of case knees (those with frequent knee pain) and control knees (those without frequent knee pain) shown beneath graph for each KL grade. Note that y axis is logarithmically scaled

The distribution of Kellgren and Lawrence grades showed that knees with frequent pain were more likely to have higher Kellgren and Lawrence grades than were the contralateral knees without frequent pain. In MOST, knees with Kellgren and Lawrence grades 1, 2, 3, and 4 had 1.5 (95% confidence interval 0.9 to 2.3), 3.9 (2.4 to 6.5), 9.0 (5.2 to 15.6), and 151 (43 to 526) times higher odds of frequent knee pain, respectively, than knees with Kellgren and Lawrence grade 0 ($P<0.001$ for trend) (figure). The corresponding odds ratios for Framingham participants were 1.2 (0.6 to 2.5), 3.1 (1.5 to 6.5), 15.1 (5.6 to 41.2), and 73 (16.2 to 331) ($P<0.001$ for trend). Both osteophytes and joint space narrowing were associated with presence of frequent knee pain in a dose-response manner, although the magnitude of association with joint space narrowing was stronger than that for osteophytes (see bmj.com).

The severity of radiographic knee osteoarthritis and individual radiographic features were also strongly associated with consistency of frequent knee pain. Compared with knees with a Kellgren and Lawrence grade of 0, the odds ratios for consistent frequent knee pain versus no frequent knee pain were 1.3, 5.5, 10.0, and 317 for knees with Kellgren and Lawrence grades of 1, 2, 3, and 4, respectively ($P<0.001$ for trend) (table 1). The odds of inconsistent frequent knee pain versus no frequent knee pain also increased as severity of radiographic osteoarthritis increased; the magnitude of association, however, was smaller than that for consistent frequent knee pain. Our findings on the relation of severity of radiographic osteoarthritis to consistent versus inconsistent frequent knee pain were similar.

Table 1 | Associations of consistency of knee pain with Kellgren and Lawrence grade, maximal osteophyte grade, and maximal joint space narrowing grade among people with two knees discordant for consistency of knee pain in MOST

	Odds ratio (95% CI)		
	Consistent pain *v* no pain (n=429 people)	Inconsistent pain *v* no pain (n=383 people)	Consistent *v* inconsistent knee pain (n=249 people)
Kellgren and Lawrence grade			
0	1.0 (referent)	1.0 (referent)	1.0 (referent)
1	1.3 (0.7 to 2.4)	1.4 (0.8 to 2.3)	2.4 (1.1 to 5.6)
2	5.5 (2.7 to 11.1)	3.0 (1.6 to 5.5)	4.1 (1.6 to 10.6)
3	10.0 (4.8 to 20.4)	8.6 (3.7 to 20.2)	10.2 (3.7 to 28.2)
4	317 (40 to 2523)	42.7 (10.3 to 177)	56.0 (13.5 to 232)
P for trend	P<0.001	P<0.001	P<0.001
Maximal osteophyte grade*			
0	1.0 (referent)	1.0 (referent)	1.0 (referent)
1	0.8 (0.5 to 1.4)	1.8 (1.1 to 3.1)	1.8 (0.8 to 4.3)
2	1.6(0.7 to 3.4)	3.1 (1.4 to 6.9)	2.9 (1.0 to 8.8)
3	2.0 (0.8 to 5.5)	2.3 (0.8 to 6.8)	3.2 (1.0 to 10.8)
P for trend	P=0.3	P=0.03	P=0.1
Maximal joint space narrowing grade*			
0	1.0 (referent)	1.0 (referent)	1.0 (referent)
1	3.0 (1.7 to 5.5)	1.7 (1.0 to 3.0)	2.2 (0.9 to 5.5)
2	6.4 (2.8 to 14.2)	4.3 (1.7 to 10.7)	4.4 (1.5 to 13.0)
3	103 (20.4 to 518)	26.5 (5.8 to 121)	21.1 (4.8 to 92.7)
P for trend	P<0.001	P<0.001	P<0.001

*Mutually adjusted for one another.

As shown in table 2, severity of radiographic osteoarthritis and severity of knee pain were also positively associated. Compared with knees with Kellgren and Lawrence grade 0, the odds ratio of severe to extreme pain versus no pain was 129 for knees with Kellgren and Lawrence grade 4 (P<0.001 for trend). We found similar results for the relation of severity of radiographic knee osteoarthritis to mild to moderate pain versus no pain (P<0.001 for trend) as well as to severe to extreme versus mild to moderate knee pain (P<0.001 for trend).

Among people in whom both knees were painful, when one knee had more severe pain than the other (difference of ≥20% and absolute difference of ≥2 on a 0-20 scale), we saw similar associations. For example, increasing radiographic severity by Kellgren and Lawrence grade was associated with odds ratios of 1.0 (referent), 1.3, 2.4, 6.3, and 30.8 (P<0.001 for trend) for having more severe compared with less severe knee pain. We also found similar associations for the relation of osteophytes and joint space narrowing to severity of pain, although the magnitudes of effect for osteophyte grades were not as large as those for joint space narrowing grades.

Table 2 | Association of severity of knee pain with Kellgren and Lawrence grade among people with two knees discordant for knee pain severity in MOST

Kellgren and Lawrence grade	Odds ratio (95% CI)		
	Severe to extreme *v* mild to moderate pain (n=257 people)	Severe to extreme *v* no pain (n=64 people)	Mild to moderate *v* no pain (n=533 people)
0	1.0 (referent)	1.0 (referent)	1.0 (referent)
1	1.5 (0.6 to 3.4)	1.2 (0.2 to 9.1)	2.1 (1.3 to 3.4)
2	2.6 (1.1 to 6.3)	7.6 (1.1 to 50.9)	6.2 (3.9 to 11.5)
3	6.0 (2.6 to 14.0)	16.1 (2.2 to 115.1)	12.6 (6.3 to 25.3)
4	15.7 (4.8 to 51.5)	129 (8.7 to 1908)	66.1 (20.4 to 214)
P for trend	P<0.001	P<0.001	P<0.001

DISCUSSION

We found a strong dose-response relation between severity of radiographic knee osteoarthritis and knee pain. Moreover, we were able to show these associations even for mild stages of osteoarthritis of the knee. Our findings for the association with frequent knee pain were consistent across two large cohorts, one with or at high risk of knee osteoarthritis and the other unselected for knee osteoarthritis. In contrast to previous findings, our study suggests that the magnitude of association between joint space narrowing and knee pain is larger than that for osteophytes and knee pain.

Strengths and limitations

Several studies have concluded that only modest associations exist between knee pain and radiographic osteoarthritis.[3-5] These conclusions need to be scrutinised. Firstly, confounding is a central concern in epidemiological studies.[25] Pain is a subjective phenomenon with many factors contributing to its occurrence. Most studies do not collect sufficient data on all of the domains that contribute to pain to allow proper assessment of the natural variability of pain among individual people. Such studies are susceptible to confounding. We compared two knees within a person in whom the two knees had different levels of pain, an approach in which all person level factors influencing pain would contribute equally to both knees, thereby eliminating between person confounding.

Secondly, as a substantial proportion of people with knee osteoarthritis have intermittent pain,[26 27] this temporal variability further complicates observational studies of knee pain. A person can be misclassified as being pain-free when pain status has been ascertained at only a single time point. Our study design is also susceptible to unadjusted potential time varying factors. Such misclassification is likely to be non-differential in our study design and would dilute the true estimates towards the null. In MOST, we assessed participants' pain status at two proximate time points and found stronger associations between radiographic osteoarthritis features among those with consistent than those with inconsistent knee pain.

Comparison with previous studies

In addition to the widely held belief that only a modest association exists between radiographic severity and pain symptoms, another difference compared with previous studies is that we found joint space narrowing to be more strongly associated than osteophytes with knee pain.[8 9] This finding was consistent irrespective of which measure we used to assess knee pain. This suggests that joint space narrowing grades adequately reflect the underlying pathological changes occurring in advanced stages of osteoarthritis. Previous studies examining radiographic narrowing and its relation to pain may have been limited by radiographic techniques that either were not standardised or did not optimally assess joint space narrowing.

Clinical implications

The lack of co-occurrence of knee pain and radiographic knee osteoarthritis may suggest that radiographic osteoarthritis has limited discriminating potential for knee pain. It does not imply, however, that the association between those two factors is weak. Knowing that a person has radiographic changes of osteoarthritis may not allow one to accurately predict the presence of pain but does contribute to our understanding of pain in osteoarthritis. Understanding the pathophysiology of pain in osteoarthritis will ultimately lead to rational therapeutic targets for this disease.

Conclusions

A strong structure-symptom association exists in osteoarthritis of the knee. Radiographic severity is a strong risk factor for the presence, consistency, and severity of knee pain and accurately reflects the presence of painful pathology. Our findings add credence to ongoing efforts to use magnetic resonance imaging studies to better understand underlying pathological structures that may be contributing to the pain of osteoarthritis.

Contributors: See bmj.com.
Funding: MOST was funded by NIH U01 AG18820, U01 AG18832, U01 AG18947, and U01 AG19069. Framingham was funded by NIH AR47785, AG18393, and NHLBI, Framingham Heart Study contract N01-HC-25195. TN was supported by NIAMS K23 AR055127, an Arthritis Foundation arthritis investigator award, and a Boston OAIC Claude Pepper research career development core grant. The study sponsors had no role in the study design; collection, analysis, and interpretation of data; writing of the article; or the decision to submit it for publication. The researchers work independently of their funders.
Competing interests: None declared.
Ethical approval: The MOST study protocol was approved by the institutional review boards at the University of Iowa; University of Alabama, Birmingham; University of California, San Francisco; and Boston University Medical Center. The Framingham Osteoarthritis Study was approved by the Boston University Medical Center institutional review board.

1 Dieppe PA. Relationship between symptoms and structural change in osteoarthritis: what are the important targets for osteoarthritis therapy? *J Rheum Suppl* 2004;70:50-3.
2 Hadler NM. Knee pain is the malady—not osteoarthritis. *Ann Intern Med* 1992;116:598-9.
3 Lawrence JS, Bremner JM, Bier F. Osteo-arthrosis: prevalence in the population and relationship between symptoms and x-ray changes. *Ann Rheum Dis* 1966;25:1-24.
4 Davis MA, Ettinger WH, Neuhaus JM, Barclay JD, Segal MR. Correlates of knee pain among US adults with and without radiographic knee osteoarthritis. *J Rheumatol* 1992;19:1943-9.
5 Hannan MT, Felson DT, Pincus T. Analysis of the discordance between radiographic changes and knee pain in osteoarthritis of the knee. *J Rheumatol* 2000;27:1513-7.
6 Lanyon P, O'Reilly S, Jones A, Doherty M. Radiographic assessment of symptomatic knee osteoarthritis in the community: definitions and normal joint space. *Ann Rheum Dis* 1998;57:595-601.
7 O'Reilly SC, Muir KR, Doherty M. Screening for pain in knee osteoarthritis: which question? *Ann Rheum Dis* 1996;55:931-3.
8 Creamer P, Lethbridge-Cejku M, Hochberg MC. Determinants of pain severity in knee osteoarthritis: effect of demographic and psychosocial variables using 3 pain measures. *J Rheumatol* 1999;26:1785-92.
9 Spector TD, Hart DJ, Byrne J, Harris PA, Dacre JE, Doyle DV. Definition of osteoarthritis of the knee for epidemiological studies. *Ann Rheum Dis* 1993;52:790-4.
10 Lethbridge-Cejku M, Scott WW Jr, Reichle R, Ettinger WH, Zonderman A, Costa P, et al. Association of radiographic features of osteoarthritis of the knee with knee pain: data from the Baltimore longitudinal study of aging. *Arthritis Care Res* 1995;8:182-8.
11 Zubieta JK, Smith YR, Bueller JA, Xu Y, Kilbourn MR, Jewett DM, et al. Regional mu opioid receptor regulation of sensory and affective dimensions of pain. *Science* 2001;293:311-5.
12 Mogil JS. The genetic mediation of individual differences in sensitivity to pain and its inhibition. *Proc Natl Acad Sci U S A* 1999;96:7744-51.
13 Vase L, Riley JL 3rd, Price DD. A comparison of placebo effects in clinical analgesic trials versus studies of placebo analgesia. *Pain* 2002;99:443-52.
14 Colloca L, Benedetti F. How prior experience shapes placebo analgesia. *Pain* 2006;124:126-33.
15 Fields HL. Pain modulation: expectation, opioid analgesia and virtual pain. *Prog Brain Res* 2000;122:245-53.
16 Wager TD. Expectations and anxiety as mediators of placebo effects in pain. *Pain* 2005;115:225-6.
17 Villemure C, Slotnick BM, Bushnell MC. Effects of odors on pain perception: deciphering the roles of emotion and attention. *Pain* 2003;106:101-8.
18 Bradley LA. Recent approaches to understanding osteoarthritis pain. *J Rheumatol Suppl* 2004;70:54-60.
19 Giardino ND, Jensen MP, Turner JA, Ehde DM, Cardenas DD. Social environment moderates the association between catastrophizing and pain among persons with a spinal cord injury. *Pain* 2003;106:19-25.
20 Gamsa A. Is emotional disturbance a precipitator or a consequence of chronic pain? *Pain* 1990;42:183-95.
21 Deshields TL, Tait RC, Gfeller JD, Chibnall JT. Relationship between social desirability and self-report in chronic pain patients. *Clin J Pain* 1995;11:189-93.
22 Englund M, Guermazi A, Gale D, Hunter DJ, Aliabadi P, Clancy M, et al. Incidental meniscal findings on knee MRI in middle-aged and elderly persons. *N Engl J Med* 2008;359:1108-15.
23 Felson DT, Nevitt MC, Yang M, Clancy M, Niu J, Torner JC, et al. A new approach yields high rates of radiographic progression in knee osteoarthritis. *J Rheumatol* 2008;35:2047-54.
24 Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. *J Rheumatol* 1988;15:1833-40.
25 Rothman KJ. *Epidemiology: an introduction*. New York: Oxford University Press, 2002.
26 Creamer P. Current perspectives on the clinical presentation of joint pain in human OA. *Novartis Found Symp* 2004;260:64-74; discussion 74-8, 100-4, 277-9.
27 Hawker GA, Stewart L, French MR, Cibere J, Jordan JM, March L, et al. Understanding the pain experience in hip and knee osteoarthritis—an OARSI/OMERACT initiative. *Osteoarthr Cartil* 2008;16:415-22.

Accepted: 29 April 2009

A lengthy illness

In 1916 a 16 year old girl was admitted in a depressive stupor to Bloomfield Hospital, a small Quaker asylum in Dublin. She was neither eating nor speaking. My grandfather (a general physician and lecturer in materia medica at Trinity College Dublin) looked after her and tube fed her for a month until she started to recover. After leaving hospital, she remained well for more than 20 years, but then she became seriously depressed again and was readmitted to Bloomfield, where my father (a consultant physician and professor of jurisprudence and hygiene at the Royal College of Surgeons Medical School) was now in charge. He treated her with a course of four straight ECT (electroconvulsive therapy without anaesthetic or muscle relaxant), and she recovered again. Twelve years later, when I was a senior house officer at St Patrick's Hospital, she was admitted with a third episode of this severe depressive illness. This time her treatment was a course of modified ECT (with an anaesthetist and the use of a muscle relaxant). When she had recovered and was leaving, she said to me, "I'll tell you something, Dr Bewley, there has been a steady improvement in psychiatric treatment over my lifetime."

She would have found more improvements with the advent of effective drugs to treat any further relapses. It is unusual to be treated by three generations of the same family for a single illness, but the past was a different time when all physicians were generalists.

Thomas Bewley retired consultant psychiatrist, London
thomasbewley@waitrose.com
Patient consent not required (patient anonymised, dead, or hypothetical).
Cite this as: *BMJ* 2009;339:b2934

pico

Use of smokeless tobacco and risk of myocardial infarction and stroke: systematic review with meta-analysis

Paolo Boffetta, Kurt Straif

International Agency for Research on Cancer, Lyon, France

Correspondence to: P Boffetta, Genetics and Epidemiology Cluster, International Agency for Research on Cancer, 150 cours Albert Thomas, 69008 Lyon, France
boffetta@iarc.fr

Cite this as: *BMJ* 2009;339:b3060
doi: 10.1136/bmj.b3060

STUDY QUESTION Are users of smokeless tobacco products in Europe and North America at increased risk of developing and dying from myocardial infarction and stroke compared with non-users?

SUMMARY ANSWER An association was detected between use of smokeless tobacco products and risk of fatal myocardial infarction and stroke, which does not seem to be explained by chance.

Selection criteria for studies

We carried out a meta-analysis of observational studies from Europe and North America on use of smokeless tobacco products and risk of myocardial infarction and stroke. Electronic databases and lists of references were used to identify studies providing a quantitative estimate of the association between use of smokeless tobacco and risk of myocardial infarction or stroke among never smokers. Eleven studies, mainly in men, were retained. Both authors independently abstracted risk estimates, together with the characteristics of each study. Summary relative risks were estimated on the basis of random effects models.

Primary outcome(s)

The primary outcomes were the occurrence of and death from myocardial infarction and stroke.

Main results and role of chance

Eight risk estimates were available for fatal myocardial infarction; the relative risk for ever use of smokeless tobacco products was 1.13 (95% confidence 1.06 to 1.21). The relative risk for fatal stroke, on the basis of five risk estimates, was 1.40 (95% confidence interval 1.28 to 1.54). For both causes of death the excess risk was restricted to current users of smokeless tobacco products. For both diseases the increased risk of death was present in studies from the United States as well as from Sweden. The inclusion of non-fatal myocardial infarction and non-fatal stroke lowered the summary risk estimates. Data on dose-response were limited but did not suggest a strong relation between risk of dying from either disease and frequency or duration of use of smokeless tobacco products. On the basis of prevalence of use of 4.4% among US men and 23% among Swedish men, the proportion of deaths from myocardial infarction attributable to use of smokeless tobacco products was 0.5% in the United States and 5.6% in Sweden; the corresponding figures for deaths from stroke were 1.7% and 5.4%.

Bias, confounding and other reasons for caution

The small magnitude of the excess risk is a reason for caution. Additional limitations were the heterogeneity of results on non-fatal myocardial infarction and non-fatal stroke. Although reporting bias might have affected case-control studies, the results were consistent with those of prospective cohort studies. Other possible forms of bias, such as changes in tobacco use during the follow-up of cohort studies, are not likely to give false positive results.

Study funding/Potential competing interests

This study received no funding. The authors declare no conflicts of interest.

This is a summary of a paper that was published on bmj.com as *BMJ* 2009;339:b3060

RESULTS OF META-ANALYSIS ON USE OF SMOKELESS TOBACCO AND RISK OF MYOCARDIAL INFARCTION AND STROKE

Outcome	No of risk estimates	P for heterogeneity	Relative risk (95% CI)
Any myocardial infarction	9	0.05	0.99 (0.89 to 1.10)
Fatal myocardial infarction	8	0.9	1.13 (1.06 to 1.21)
Any stroke	6	<0.001	1.19 (0.97 to 1.47)
Fatal stroke	5	0.5	1.40 (1.28 to 1.54)

BMJ pico: advice to authors

The full text versions of all accepted *BMJ* research articles are published online, with open access and no word limits, on bmj.com as soon as they are ready. In the print *BMJ* each research article is abridged, as a one page BMJ pico, with the aim of making research more inviting and useful to readers. Starting in August 2009, authors are writing their own BMJ picos.

We have designed BMJ pico with evidence based medicine experts to succinctly present the key evidence from each study, to help minimise delay between online and print publication, and to enable us to publish more research in each week's print *BMJ*.

There is no need for authors to prepare a BMJ pico to submit along with the full research article. Authors produce their own BMJ pico, using a template from us, only after the full article has been accepted.

Because a research article on bmj.com is the definitive, indexed, citable version (not an interim "epublication ahead of print"), authors who do not wish to abridge their articles using BMJ pico will be able to opt for online only publication.

For templates and more information, please visit http://resources.bmj.com/bmj/authors/article-submission/bmj-pico-abridged-research-articles. You may also find our frequently asked questions helpful at http://resources.bmj.com/bmj/authors/article-submission/faqs-about-bmj.pico