RESEARCH

Incidental findings on brain magnetic resonance imaging: systematic review and meta-analysis

Zoe Morris,[1] William N Whiteley,[1] W T Longstreth Jr,[2] Frank Weber,[3] Yi-Chung Lee,[4] Yoshito Tsushima,[5] Hannah Alphs,[6] Susanne C Ladd,[7] Charles Warlow,[1] Joanna M Wardlaw,[1 8] Rustam Al-Shahi Salman[1]

EDITORIAL by van der Lugt

[1]Division of Clinical Neurosciences, Western General Hospital, Edinburgh EH4 2XU
[2]Departments of Neurology and Epidemiology, Harborview Medical Center, Seattle, WA, USA
[3]German Air Force Institute of Aviation Medicine, Department of Neurology, Fuerstenfeldbruck, Germany
[4]Department of Neurology, Taipei Veterans General Hospital, Taiwan
[5]Department of Diagnostic Radiology and Nuclear Medicine, Gunma University Hospital, Japan
[6]Department of Radiology and Radiological Sciences, Johns Hopkins Hospital, Baltimore, MD, USA
[7]Department of Diagnostic and Interventional Radiology and Neuroradiology, University Hospital Essen, Germany
[8]SFC Brain Imaging Research Centre, University of Edinburgh
Correspondence to:
R Al-Shahi Salman
Rustam.Al-Shahi@ed.ac.uk

Cite this as: *BMJ* 2009;339:b3016
doi: 10.1136/bmj.b3016

This article is an abridged version of a paper that was published on bmj.com. Cite this article as: *BMJ* 2009;339:b3016

ABSTRACT

Objective To quantify the prevalence of incidental findings on magnetic resonance imaging (MRI) of the brain.

Design Systematic review and meta-analysis of observational studies.

Data sources Ovid Medline (1950 to May 2008), Embase (1980 to May 2008), and bibliographies of relevant articles.

Review methods Two reviewers sought and assessed studies of people without neurological symptoms who underwent MRI of the brain with or without intravenous contrast for research purposes or for occupational, clinical, or commercial screening.

Main outcome measures Overall disease specific and age specific prevalence of incidental brain findings, calculated by meta-analysis of pooled proportions using DerSimonian-Laird weights in a random effects model.

Results In 16 studies, 135 of 19 559 people had neoplastic incidental brain findings (prevalence 0.70%, 95% confidence interval 0.47% to 0.98%), and prevalence increased with age (χ^2 for linear trend, P=0.003). In 15 studies, 375 of 15 559 people had non-neoplastic incidental brain findings (prevalence 2.0%, 1.1% to 3.1%, excluding white matter hyperintensities, silent infarcts, and microbleeds). The number of asymptomatic people needed to scan to detect any incidental brain finding was 37. The prevalence of incidental brain findings was higher in studies using high resolution MRI sequences than in those using standard resolution sequences (4.3% *v* 1.7%, P<0.001). The prevalence of neoplastic incidental brain findings increased with age.

Conclusions Incidental findings on brain MRI are common, prevalence increases with age, and detection is more likely using high resolution MRI sequences than standard resolution sequences. These findings deserve to be mentioned when obtaining informed consent for brain MRI in research and clinical practice but are not sufficient to justify screening healthy asymptomatic people.

WHAT IS ALREADY KNOWN ON THIS TOPIC

Brain magnetic resonance imaging (MRI) is widely used in research and clinical practice and can be purchased for health screening purposes

Brain MRI detects incidental findings in people with asymptomatic neurological conditions

Precise estimates of the frequency of incidental findings and influences on their detection are yet to be determined

WHAT THIS STUDY ADDS

The crude prevalence of incidental findings on brain MRI is 2.7%, or one for every 37 neurologically asymptomatic people scanned

Incidental brain findings are more likely to be detected in studies using at least one high resolution MRI sequence than studies using standard sequences (4.3% *v* 1.7%)

The frequency of incidental findings should be discussed when obtaining consent for brain MRI in research and is relevant to clinical practice, but alone does not justify health screening

INTRODUCTION

The detection of incidental findings is a consequence of using magnetic resonance imaging (MRI) of the brain in clinical practice, research, and screening. Detection is potentially detrimental, partly because treatment can have harmful as well as beneficial consequences. We carried out a systematic review and meta-analysis of the published literature to provide more precise estimates of the range of incidental findings on brain MRI and to explore the influence of study design, patient characteristics, and imaging parameters on the detection of incidental brain findings.

METHODS

We searched Medline (1950-May 2008) and Embase (1980-May 2008) for reports on the use of brain MRI in healthy people, volunteers, research controls, and people undergoing commercial, clinical, or occupational screening. We also surveyed tables of contents in neurological journals and hand searched the bibliographies of pertinent articles. Two authors (ZM and WNW or RA-SS) read the titles and abstracts of identified studies and critically appraised the full text.

We defined incidental brain findings as apparently asymptomatic intracranial abnormalities that were clinically significant because of their potential to cause symptoms or influence treatment. We divided the findings into neoplastic (benign and malignant tumours) and non-neoplastic (cysts, structural vascular abnormalities, inflammatory lesions, and "other"). We did not focus on white matter hyperintensities, silent brain infarcts or lacunae, and brain microbleeds because of their known increasing prevalence with age,[1 2] largely

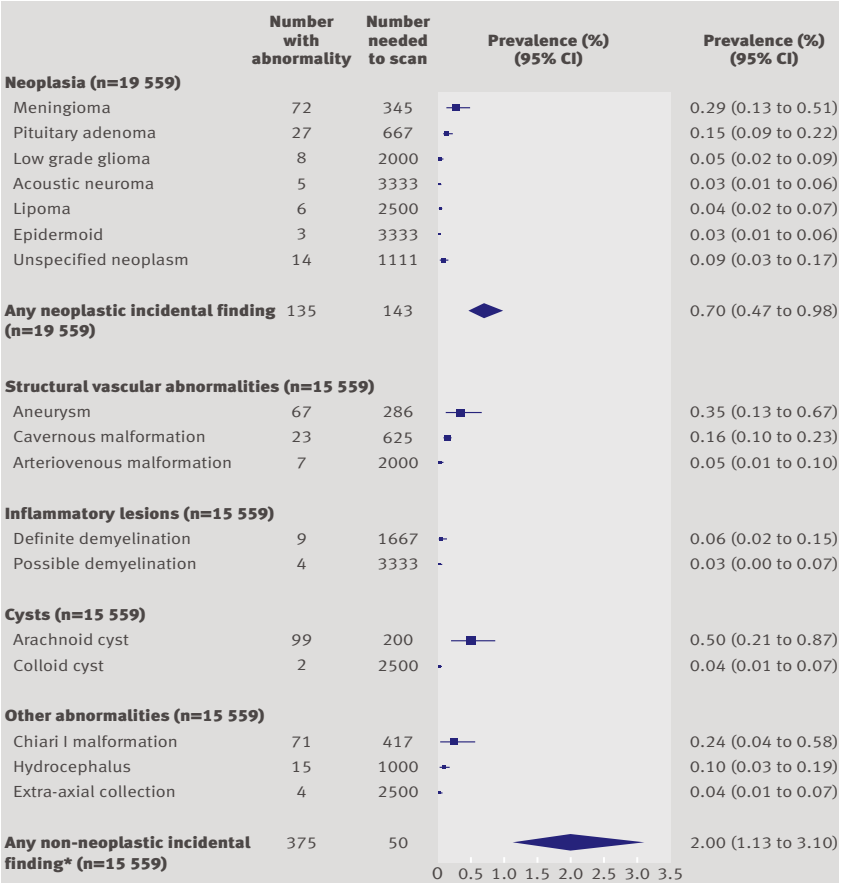
unknown role in causing symptoms, and uncertainty about whether to institute primary prevention after their detection.[2] We distinguished incidental brain findings from normal variants, which we defined as anatomical variants without the potential to cause symptoms.

We included studies that reported the prevalence of incidental brain findings in people without neurological or psychiatric symptoms, who underwent brain MRI as research cases or controls or as recipients of commercial, clinical, or occupational screening. We did not include studies restricted to markers of cerebrovascular disease because they have recently been the subject of systematic reviews.[1 2] If several publications arose from the same cohort, we included the largest study.

Two authors extracted data on study design, population characteristics, and MRI parameters from each study, and extracted the overall and age specific frequencies of each type of incidental brain finding (excluding markers of cerebrovascular disease). When age specific data on prevalence were not provided we emailed the corresponding author, who became a coauthor of this review if they extracted and supplied data.

Data analysis

We carried out a meta-analysis of prevalence data for each incidental brain finding, and all of them combined, using data from studies that enabled calculations. We used the I^2 statistic to estimate the heterogeneity of individual studies contributing to the pooled estimate. We calculated the pooled proportion (95% confidence intervals) as the back transform of the weighted mean of the transformed proportions, using DerSimonian-Laird weights in a random effects model.[3] We did subgroup analyses to explore the influence of MRI sequences, the specialty of the interpreter of the scan, and participant characteristics on the pooled prevalence of all incidental brain findings. We calculated age specific prevalence in 20 year age bands, with available data. The number of asymptomatic people needed to scan to detect one incidental brain finding (number needed to scan) was the reciprocal of the prevalence estimate.



Prevalence of some incidental findings (*excluding white matter hyperintensities, microbleeds, and silent infarcts) on brain magnetic resonance imaging

RESULTS

Of 1862 publications identified, 19 papers reporting data on 17 cohorts were eligible.[4-6 w1-w16] After exclusions we included data on 16 cohorts (see bmj.com) who had undergone brain MRI (19 559 people, 1989-2008) from Asia (n=7277),[w5 w10 w15] Europe (n=5942),[w1 w2 w9 w11 w14] the United States (n=5764),[w3 w4 w7 w8 w12 w13] and Australia (n=576).[w6 w16] The number of people in each study ranged from 60 to 4000, mean age 11 to 63 years (range 1-97 years). One study included cases (n=589) and controls (n=67),[w12] but the rest included exclusively controls (six studies, n=1702),[w2 w4 w6-w8 w13] cases (three studies, n=6739),[w1 w3 w16] or screening attendees (six studies, n=11 118).[w5 w9-w11 w14 w15]

Twenty one participants (0.1%) had preceding neurological symptoms that may have been related to abnormalities found on brain MRI.[w1 w3 w10 w16] No study prespecified the potential incidental brain findings of interest, and almost none was confirmed by pathology. Only three studies defined normal variants (see bmj.com).

None of the studies published before 2002 used sequences regarded as high resolution,[w2-w5] and most subsequent studies used lower resolution sequences (see bmj.com). Some recent studies also included magnetic resonance angiographic sequences,[w4-w7] or high resolution sequences such as three dimensional T1 spoiled or T2* gradient echo.[w1 w6-w8 w11-w13 w16] Abnormalities on scans were interpreted by neuroradiologists,[w3 w4 w6-w9 w12 w13 w16] a neuroradiologist or general radiologist,[w15] a neuroradiologist or neurologist,[w1] general radiologists,[w10 w11 w14] or unspecified observers.[w2 w5]

Disease specific and overall prevalence

Disease specific prevalence was calculable for intracranial neoplasms in all 19 559 participants, but one study of 4000 participants only described asymptomatic tumours,[w5] resulting in a denominator of 15 559 for prevalence of non-neoplastic incidental brain findings (figure). The I^2 statistic ranged from 0% to 86%, indicating variable degrees of heterogeneity among the included studies. We, therefore, used a random effects model.

The prevalence of neoplastic incidental findings was 0.70% (95% confidence 0.47% to 0.98%), but description of the prevalence of each specific tumour type was impaired by either a lack of subtyping[w12] or non-specific classifications. The prevalence of non-neoplastic incidental findings was 2.0% (1.1% to 3.1%; figure). The combined prevalence of neoplastic and non-neoplastic incidental findings was 2.7% (number needed to scan=37).

Influence of MRI sequences, reporting, and participant characteristics

The detection of incidental findings was higher in studies using at least one high resolution sequence (318/6204; 4.3%, 3.0% to 5.8%)[w1 w6-w8 w11-w13 w16] than in studies using standard sequences (176/9355; 1.7%, 1.1% to 2.4%, χ^2 P<0.001).[w2-w4 w9 w10 w14 w15] The detection of these incidental findings in studies using neuroradiologists to interpret images (272/8340; 3.5%, 1.8% to 5.7%)[w1 w3 w4 w6-w9 w12 w13 w15 w16] was not significantly higher than in studies using general radiologists (144/4954; 2.3%, 0.9% to 4.4%, χ^2 P=0.3).[w10 w11 w14]

In analyses restricted to studies using at least one high resolution MRI sequence or three dimensional time of flight magnetic resonance angiography, the prevalence of incidental brain findings was higher among research cases (198/6150; 3.4%, 0.9% to 7.5%) than among attendees of commercial screening (105/4582; 2.0, 0.9% to 3.3%) and research controls (24/1635; 1.6%, 1.0% to 2.2%, χ^2 P<0.001).

Age specific prevalence

Of the 16 included studies, the original data were no longer available for two (5000 participants),[w4 w5] one (n=2000) declined to provide age specific data,[w1] and five (n=1582) failed to contribute data on request,[w2 w6-w8 w14] leaving age specific grouped summary data on 10 977 people, provided by six studies[w3 w9-w12 w15] and extracted from the reports of two others with participants in just one 20 year age band.[w13 w16] After omissions, four 20 year age bands were left for analysis of age specific prevalence (see bmj.com).

The prevalence of neoplastic incidental brain findings increased with age (χ^2 for linear trend=8.8, P=0.003), whereas the prevalence of non-neoplastic incidental brain findings seemed to decline (χ^2 for linear trend=6.9, P=0.008; see bmj.com). This trend was, however, reversed in a sensitivity analysis restricted to studies with age specific data that used at least one high resolution sequence (χ^2 for linear trend=66, P<0.001; see bmj.com).[w11 w12 w16]

DISCUSSION

In this systematic review and meta-analysis of 16 studies totalling 19 559 participants, the overall prevalence of incidental brain findings on brain MRI was 2.7% (number needed to scan=37). In studies where participants underwent at least one high resolution MRI sequence (common in brain imaging research) the prevalence of incidental brain findings was 4.3% (number needed to scan=23) compared with 1.7% (number needed to scan=59) in studies using only low resolution sequences (most commonly used in clinical practice). We found an increasing prevalence of all neoplastic incidental brain findings with age (see bmj.com), probably driven by the increasing prevalence of meningiomas,[w1] the most common neoplastic incidental brain finding (figure).

Strengths and weaknesses of this review

By synthesising all the data on incidental brain findings and adding unpublished data where possible, we increased the precision of estimates of prevalence across the whole age range (figure). The influence of variations in study design was diluted by pooling the data, and we used subgroup analyses to explore this heterogeneity in study characteristics and imaging parameters. The provision of some unpublished grouped summary data also enabled us to examine age specific prevalence.

Using only grouped summary data prevented us from exploring the influence of sex on the prevalence of incidental brain findings found in other studies.[7 w7 w8] The proportion of participants who actually had neurological symptoms referable to incidental brain findings was low (0.1%); this may be unavoidable because some participants may attend for investigation of undeclared symptoms,[8] whereas others may be serial attenders checking on undeclared underlying disease.[9] Because people with incidental findings are not eligible for some research studies, we may have underestimated their true prevalence.

Comparison with other studies

Our pooled estimate provides a more precise summary of the existing data. The prevalence of incidental brain findings described by other studies has varied, in part because of the factors we have explored in sensitivity analyses. Others have found an increasing prevalence of some incidental brain findings with age,[w1 w6 w8 w12] but we were able to classify them into neoplastic and non-neoplastic incidental findings (see bmj.com) and confirm the robustness of the trend in the prevalence of neoplastic incidental findings in a sensitivity analysis.

Implications for clinical practice, research, and screening

The evidence on what to do with most incidental brain findings is insufficient, partly because of the lack of controlled trials of their treatments and partly because MRI has been available for only 20-30 years.

Some have suggested a subdivision of incidental brain findings by the perceived need and urgency of referral to a specialist.[10] But urgency is difficult to gauge given the paucity of robust evidence on the treatment of asymptomatic incidental findings. The clinical urgency will vary according to the age and healthiness of the patient, and the perceived urgency may change over time as knowledge about the effects of treatment changes.[11]

Apart from the harm that may arise from lack of evidence, the detection of incidental brain findings

can provoke considerable anxiety about a "possible abnormality"[w13]; involve a costly cascade of investigations, with risks of complications; lead to costly medical opinions; and worry patients about the consequences of forgoing treatment. For the patient, the discovery of an incidental brain finding may result in loss of their driving licence, life insurance, and even employment.

At the very least clinicians should counsel patients about the chance of incidental findings with brain MRI. Volunteers for research studies using brain MRI should be informed about incidental findings, and research centres need to have mechanisms in place to deal with these once found.[12] Furthermore, the increasing number of screening companies that provide "health check-ups"[13 14] has attracted caution from only a few regulatory bodies.[15 16] In such screening the actual objective is the discovery of incidental brain findings, which may be regarded by the client as fortuitous.[8 13] Although true negative results from brain MRI may be reassuring, many of the requirements of a screening test are not fulfilled; most of all, the overall benefit of such screening on quality adjusted life years is unproved.

Conclusions

Doctors who request scans in clinical practice or who recommend screening for health check-ups, and researchers who obtain consent from volunteers, should provide information about the prevalence of incidental brain findings on brain MRI, the higher prevalence with high resolution MRI sequences, and the shortage of evidence to inform their management.

We thank the Edinburgh Stroke Research Group for its comments.

Contributors: See bmj.com.

Funding: The Chief Scientist Office of the Scottish Government supports WW with a Clinical Academic Training Fellowship, and ZM with a research and development infrastructure award for radiology to Lothian Universities Hospitals NHS Trust Research and Development Office. The UK Medical Research Council supports RA-SS with a clinician scientist fellowship. JMW is partly supported by the Scottish Funding Council and Chief Scientist Office through the SINAPSE (Scottish Imaging Network. A Platform for Scientific Excellence) Collaboration (www.sinapse.ac.uk). The Cardiovascular Health Study is funded by the National Heart, Lung, and Blood Institute. The researchers are independent of the funders, and the study sponsors had no role in the study design; the collection, analysis, and interpretation of data; the writing of the report; and the decision to submit the article for publication.

Competing interests: None declared.

Ethical approval: Not required.

1 Cordonnier C, Al-Shahi Salman R, Wardlaw J. Spontaneous brain microbleeds: systematic review, subgroup analyses and standards for study design and reporting. *Brain* 2007;130:1988-2003.
2 Vermeer SE, Longstreth Jr WT, Koudstaal PJ. Silent brain infarcts: a systematic review. *Lancet Neurol* 2007;6:611-9.
3 DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986;7:177-88.
4 Weber F, Knopf H. Cranial MRI as a screening tool: findings in 1,772 military pilot applicants. *Aviation Space Environment Med* 2004;75:158-61.
5 Göhde SC, Goyen M, Forsting M, Debatin JF. [Prevention without radiation-a strategy for comprehensive early detection using magnetic resonance tomography]. [German]. *Radiologe* 2002;42:622-9.
6 Meadows J, Kraut M, Guarnieri M, Haroun RI, Carson BS. Asymptomatic Chiari type I malformations identified on magnetic resonance imaging. *J Neurosurg* 2000;92:920-6.
7 Sawamura Y, Ikeda J, Ozawa M, Minoshima Y, Saito H, Abe H. Magnetic resonance images reveal a high incidence of asymptomatic pineal cysts in young women. *Neurosurgery* 1995;37:11-5.
8 Mamourian A. Incidental findings on research functional MR images: should we look? *AJNR Am J Neuroradiol* 2004;25:520-2.
9 Pickard JD, Gillard JH. Guidelines reduce the risk of brain-scan shock. *Nature* 2005;435:17.
10 Bryan RN, Manolio TA, Schertz LD, Jungreis C, Poirier VC, Elster AD, et al. A method for using MR to evaluate the effects of cardiovascular disease on the brain: the cardiovascular health study. *AJNR Am J Neuroradiol* 1994;15:1625-33.
11 Hoggard N, Darwent G, Capener D, Wilkinson ID, Griffiths PD. The high incidence and bioethics of findings on magnetic resonance brain imaging of normal volunteers for neuroscience research. *J Med Ethics* 2009;35:194-9.
12 Illes J, Kirschen MP, Edwards E, Bandettini P, Cho MK, Ford PJ, et al. Practical approaches to incidental findings in brain imaging research. *Neurology* 2008;70:384-90.
13 Al-Shahi Salman R, Whiteley WN, Warlow C. Screening using whole-body magnetic resonance imaging scanning: who wants an incidentaloma? *J Med Screen* 2007;14:2-4.
14 Komotar RJ, Starke RM, Connolly ES. Brain magnetic resonance imaging scans for asymptomatic patients: role in medical screening. *Mayo Clin Proc* 2008;83:563-5.
15 Health Canada. *Whole body screening using MRI or CT technology*. Ottawa, ON: Health Canada, 2003.
16 Royal Australian and New Zealand College of Radiologists. *The use of diagnostic imaging for screening purposes and non-referred investigations: RANZCR statement of principles*. Sydney, Australia: RNAZCR, 2005.

Accepted: 27 May 2009

The doctor who became a shepherd

I used to start my day as a GP by logging on and checking my patients' latest laboratory and imaging results. I work in Ashdod, a town of a quarter of a million inhabitants in the southern part of Israel, about 40 km from the Gaza strip. When the recent crisis started, before I asked a patient to undress or prepare for an electrocardiogram, I had to consider whether we would have time to run to the bomb shelter if the siren wails. What would I do about elderly or infirm patients who couldn't move quickly enough to the shelter? A new ethical dilemma.

The type of work changed as well. Although it was the flu season, only patients with urgent problems were willing to leave their homes to visit me, and they were reluctant to undergo tests. The sound of the siren brought back dark memories for many patients who came from Europe after the second world war to find a peaceful haven. I also reflected with great respect on my parents and grandparents, who spoke of the days and nights they spent sheltering from Nazi bombs in London during the Blitz. When the siren did wail I gathered together the patients and staff and shepherded them out and down the stairs to the bomb shelter. Despite this, spirits were high and a new closeness was forged between us—neighbours who hadn't spoken for ages were now friends in this time of adversity.

After minutes of tense waiting and listening for the sound of falling rockets, I shepherded my patients and staff back to the surgery and tried to carry on the examination where I left off.

Anthony S Oberman general practitioner, Ashdod, Israel
anthonyo@doctors.org.uk

Cite this as: *BMJ* 2009;339:b449

Disagreements in meta-analyses using outcomes measured on continuous or rating scales: observer agreement study

Britta Tendal,[1] Julian P T Higgins,[4] Peter Jüni,[2 3] Asbjørn Hróbjartsson,[1] Sven Trelle,[2 3] Eveline Nüesch,[2 3] Simon Wandel,[2 3] Anders W Jørgensen,[1] Katarina Gesser,[5] Søren Ilsøe-Kristensen,[5] Peter C Gøtzsche[1]

[1]Nordic Cochrane Centre, Rigshospitalet, Dept 3343, Blegdamsvej 9, DK-2100 Copenhagen, Denmark
[2]Institute of Social and Preventive Medicine, University of Bern, Switzerland
[3]CTU Bern, Bern University Hospital, Switzerland
[4]MRC Biostatistics Unit, Institute of Public Health, University of Cambridge, Cambridge
[5]Faculty of Pharmaceutical Sciences, University of Copenhagen, Denmark
Correspondence to: B Tendal
bt@cochrane.dk

Cite this as: *BMJ* 2009;339:b3128
doi: 10.1136/bmj.b3128

ABSTRACT

Objective To study the inter-observer variation related to extraction of continuous and numerical rating scale data from trial reports for use in meta-analyses.

Design Observer agreement study.

Data sources A random sample of 10 Cochrane reviews that presented a result as a standardised mean difference (SMD), the protocols for the reviews and the trial reports (n=45) were retrieved.

Data extraction Five experienced methodologists and five PhD students independently extracted data from the trial reports for calculation of the first SMD result in each review. The observers did not have access to the reviews but to the protocols, where the relevant outcome was highlighted. The agreement was analysed at both trial and meta-analysis level, pairing the observers in all possible ways (45 pairs, yielding 2025 pairs of trials and 450 pairs of meta-analyses). Agreement was defined as SMDs that differed less than 0.1 in their point estimates or confidence intervals.

Results The agreement was 53% at trial level and 31% at meta-analysis level. Including all pairs, the median disagreement was SMD=0.22 (interquartile range 0.07-0.61). The experts agreed somewhat more than the PhD students at trial level (61% *v* 46%), but not at meta-analysis level. Important reasons for disagreement were differences in selection of time points, scales, control groups, and type of calculations; whether to include a trial in the meta-analysis; and data extraction errors made by the observers. In 14 out of the 100 SMDs calculated at the meta-analysis level, individual observers reached different conclusions than the originally published review.

Conclusions Disagreements were common and often larger than the effect of commonly used treatments. Meta-analyses using SMDs are prone to observer variation and should be interpreted with caution. The reliability of meta-analyses might be improved by having more detailed review protocols, more than one observer, and statistical expertise.

WHAT IS ALREADY KNOWN ON THIS TOPIC

Incorrect data extraction in meta-analyses can lead to false results

Multiplicity in trial reports invites variation in data extraction, as different judgments will lead to different choices about which data to extract

The impact of these different errors and choices on meta-analysis results is not clear

WHAT THIS STUDY ADDS

There is considerable observer variation in data extraction and decisions on which trials to include

The reasons for disagreement are different choices and errors

The impact on meta-analyses is potentially large

This article is an abridged version of a paper that was published on bmj.com. Cite this article as: *BMJ* 2009;339:b3128

INTRODUCTION

Systematic reviews of clinical trials, with meta-analyses if possible, are regarded as the most reliable resource for decisions about prevention and treatment. They should be based on a detailed protocol that aims to reduce bias by pre-specifying methods and selection of studies and data.[1] However, as meta-analyses are usually based on data that have already been processed, interpreted, and summarised by other researchers, data extraction can be complicated and can lead to important errors.[2]

There is often a multiplicity of data in trial reports that makes it difficult to decide which ones to use in a meta-analysis. Furthermore, data are often incompletely reported,[2 3] which makes it necessary to perform calculations or impute missing data. Different observers may get different results, but previous studies on observer variation have not been informative, because of few observers, few trials, or few data.[4 5] We report here a detailed study of observer variation that explores the sources of disagreement when extracting data for calculation of standardised mean differences.

METHODS

We selected a random sample of 10 Cochrane reviews published in the *Cochrane Library* in 2006-7.[6-15] We retrieved the reports of the randomised trials that were included in the reviews and the protocols for each of the reviews. We included reviews that reported at least one result as a standardised mean difference (SMD). The SMD is used when trial authors have used different scales for measuring the same underlying outcome. In such cases, it is necessary to standardise the measurements on a uniform scale before they can be pooled in a meta-analysis. The SMD for each trial is calculated as the difference in means between the two groups, divided by the pooled standard deviation of the measurements.[1] The first SMD result in each review was selected as our index result. The index result had to be based on two to 10 trials and on published data only.

Five methodologists and five PhD students independently extracted data from the trial reports for calculation of the SMDs. The observers had access to the review protocols but not the completed Cochrane reviews. An additional researcher highlighted the relevant outcome in the protocols, along with other important issues such as pre-specified time points of interest, which intervention was the experimental one and which was the

Table 1 | Levels of overall agreement between observer pairs in the calculated standardised mean differences (SMDs)* from 10 meta-analyses (which comprised a total of 45 trials)

Observer pairs	No (%) of pairs in agreement
Trial level	
All pairs (n=2025):	1068 (53)
Methodologists (n=450)	273 (61)
PhD students (n=450)	209 (46)
Mixed pairs (n=1125)	586 (52)
Meta-analysis level	
All pairs (n=450):	138 (31)
Methodologists (n=100)	33 (33)
PhD students (n=100)	27 (27)
Mixed pairs (n=250)	78 (31)

*Agreement defined as SMDs that differed less than 0.1 in their point estimates and in their 95% confidence intervals.

control. If information was missing the observers decided themselves what to select from the trial reports.

If the data were available the observers extracted means, standard deviations, and number of patients for each group; otherwise, they could calculate missing data, such as from an exact P value. The observers interpreted the sign of the SMD results—that is, whether a negative or a positive result indicated superiority of the experimental intervention. Observers could exclude trials. Based on the extracted data, the additional researcher calculated trial and meta-analysis SMDs for each observer.

Agreement between pairs of observers was assessed at both meta-analysis and trial level, pairing the 10 observers in all possible ways (45 pairs). Agreement was defined as SMDs that differed less than 0.1 in their point estimates and in their confidence intervals. To determine the variation in meta-analysis results that could be obtained from the multiplicity of different SMD estimates across observers, we conducted a Monte Carlo simulation for each meta-analysis.

RESULTS

The 10 meta-analyses comprised 45 trials, which yielded 450 pairs of observers at the meta-analysis level and 2025 pairs at the trial level.

None of the review protocols contained information on which scales should be preferred. Three protocols gave information about which time point to select and four mentioned whether change from baseline or values after treatment should be preferred. Nine described which type of control group to select. The outcomes analysed in the 10 meta-analyses were diverse: in six, the outcome was a clinician reported score; in one, it was objective; and in three, it was self reported.

Table 2 | Levels of agreement at the meta-analysis level between observer pairs in the calculated standardised mean differences (SMDs) from 10 meta-analyses*

Meta-analysis	No (%) of pairs in agreement			
	All pairs (n=45)	Methodologist (n=10)	Students (n=10)	Mixed pairs (n=25)
Gava et al[6]	6 (13)	1 (10)	0 (0)	5 (20)
Woodford et al[7]	11 (24)	2 (20)	1 (10)	8 (32)
Martinez et al[8]	7 (16)	3 (30)	1 (10)	3 (12)
Orlando et al[9]	5 (11)	1 (10)	2 (20)	2 (8)
Buckley et al[10]	6 (13)	1 (10)	1 (10)	4 (16)
Ipser et al[11]	13 (29)	4 (40)	2 (20)	7 (28)
Mistiaen et al[12]	16 (36)	6 (60)	2 (20)	8 (32)
Afolabi et al[13]	28 (62)	6 (60)	6 (60)	16 (64)
Uman et al[14]	36 (80)	6 (60)	10 (100)	20 (80)
Moore et al[15]	10 (22)	3 (30)	2 (20)	5 (20)

*Agreement defined as SMDs that differed less than 0.1 in their point estimates and in their 95% confidence intervals.

Agreement at trial level

Across trials, the agreement was 53% for the 2025 pairs (61% for the 450 pairs of methodologists, 46% for the 450 pairs of PhD students, and 52% for the 1125 mixed pairs) (table 1). The agreement rates for the individual trials ranged from 4% to 100%. Agreement between all observers was found for four of the 45 trials.

The reasons for disagreement fell into three broad categories: different choices, exclusion of a trial, and data extraction errors. The different choices mainly concerned selection of experimental or the control groups (15 trials), which time point to select (nine trials), which scale to use (six trials), and different ways of calculating or imputing missing numbers (six trials). The most common reasons for deciding to exclude a trial was that the trial did not meet the inclusion criteria described in the protocol (14 trials) and that the reporting was so unclear that data extraction was not possible (14 trials). Data extraction errors were less common but involved misinterpretation of the direction of the effect in four trials.

Agreement at meta-analysis level

Across the meta-analyses, the agreement was 31% for the 450 pairs (33% for the 100 pairs of methodologists, 27% for the 100 pairs of PhD students, and 31% for the 250 mixed pairs) (table 1). The agreement rates for the individual meta-analyses ranged from 11% to 80% (table 2). Agreement between all observers was not found for any of the 10 meta-analyses.

Of the 450 pairs, 10% agreed completely, 21% had a disagreement below our cut point of 0.1, 38% had a disagreement between 0.1 and 0.49, and 28% disagreed by at least 0.50 (including 10% that had disagreements of ≥1). The last 18 pairs (4%) were not quantifiable since one observer excluded all the trials from two meta-analyses. The median disagreement was SMD=0.22 for the 432 quantifiable pairs with an interquartile range from 0.07 to 0.61. There were no differences between the methodologists and the PhD students (table 1).

We compared the SMDs calculated by each of the 10 observers for the 10 meta-analyses, and the results from the originally published meta-analyses. Out of the total of 100 calculated SMDs, seven values corresponding to significant results in the originally published meta-analyses were now non-significant, three values corresponding to non-significant results were now significant, and four values, which were related to the same published meta-analysis, showed a significantly beneficial effect for the control group whereas the original publication reported a significantly beneficial effect for the experimental group.[10]

The Monte Carlo investigation showed that four of the 10 meta-analyses[6 10 12 13] had considerable variation in the potential SMDs, allowing for differences in SMDs of up to 3 (see bmj.com).

DISCUSSION

We found that disagreements between observers were common and often large. Ten per cent of the disagreements at the meta-analysis level amounted to an SMD of at least 1, which is far greater than the effect of most of the treatments we use compared with no treatment. Important reasons for disagreement were differences in selection of time points, scales, control groups, and type of calculations, whether to include a trial in the meta-analysis, and data extraction errors made by the observers.

The disagreement depended on the reporting of data in the trial reports and on how much room was left for decision in the review protocols.

Strengths and weaknesses

We took a broad approach and showed that there are other important sources of variation in meta-analysis results than simple errors. We included a number of experienced as well as inexperienced observers and a large number of trials.

The experimental setting had limitations. Single data extraction produces more errors than double data extraction.[5] In real life, some of the errors we made would therefore probably have been detected before the data were used for meta-analyses, as it is recommended for Cochrane reviews that any disagreement should be resolved by discussion and arbitration.[1] We did not perform a consensus step. However, given the amount of multiplicity in the trial reports and the uncertainties in the protocols, it is likely that even pairs of observers would disagree considerably with other pairs.

The observers were presented with protocols they had not developed themselves, based on research questions they had not asked, and in disease areas where they were mostly not experts. Another limitation is that some of the trial reports did not contain the data needed for the calculation of an SMD. It would therefore have been helpful to contact trial authors.

Other similar research

The SMD is intended to give clinicians and policymakers the most reliable summary of the available trial evidence when the outcomes have been measured on different continuous or numeric rating scales. Surprisingly, the method has not previously been examined in any detail for its own reliability. Previous research has been sparse and has focused on errors in data extraction.[2 4 5] In one study, the authors found errors in 20 of 34 Cochrane reviews, but, as they gave no numerical data, it is not possible to judge how often these were important.[4] The results of our study apply more broadly than to meta-analyses using the SMD, as many of the reasons for disagreement were not related to the SMD method.

Conclusions

Meta-analyses using SMDs are prone to observer variation and should be interpreted with caution. The reliability of meta-analyses might be improved by having more detailed review protocols, more than one observer, and statistical expertise.

Contributors: See bmj.com.

Funding: This study is part of a PhD funded by IMK Charitable Fund and the Nordic Cochrane Centre. The sponsors had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review, or approval of the manuscript. The researchers were independent from the funders.

Competing interests: None declared.

Ethical approval: Not required

1 Higgins JPT, Green S, eds. *Cochrane handbook for systematic reviews of interventions*. Version 5.0.0. 2008. www.cochrane-handbook.org
2 Gotzsche PC, Hrobjartsson A, Maric K, Tendal B. Data extraction errors in meta-analyses that use standardized mean differences. *JAMA* 2007;298:430-7.
3 Chan AW, Hrobjartsson A, Haahr MT, Gotzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *JAMA* 2004;291:2457-65.
4 Jones AP, Remmington T, Williamson PR, Ashby D, Smyth RL. High prevalence but low impact of data extraction and reporting errors were found in Cochrane systematic reviews. *J Clin Epidemiol* 2005;58:741-2.
5 Buscemi N, Hartling L, Vandermeer B, Tjosvold L, Klassen TP. Single data extraction generated more errors than double data extraction in systematic reviews. *J Clin Epidemiol* 2006;59:697-703.
6 Gava I, Barbui C, Aguglia E, Carlino D, Churchill R, De VM, et al. Psychological treatments versus treatment as usual for obsessive compulsive disorder (OCD). *Cochrane Database Syst Rev* 2007;(2):CD005333.
7 Woodford H, Price C. EMG biofeedback for the recovery of motor function after stroke. *Cochrane Database Syst Rev* 2007;(2):CD004585.
8 Martinez DP, Waddell A, Perera R, Theodoulou M. Cognitive behavioural therapy for tinnitus. *Cochrane Database Syst Rev*: 2007;(1):CD005233.
9 Orlando R, Azzalini L, Orando S, Lirussi F. Bile acids for non-alcoholic fatty liver disease and/or steatohepatitis. *Cochrane Database Syst Rev* 2007;(1):CD005160.
10 Buckley LA, Pettit T, Adams CE. Supportive therapy for schizophrenia. *Cochrane Database Syst Rev* 2007;(3):CD004716.
11 Ipser JC, Carey P, Dhansay Y, Fakier N, Seedat S, Stein DJ. Pharmacotherapy augmentation strategies in treatment-resistant anxiety disorders. *Cochrane Database Syst Rev* 2006;(4):CD005473.
12 Mistiaen P, Poot E. Telephone follow-up, initiated by a hospital-based health professional, for postdischarge problems in patients discharged from hospital to home. *Cochrane Database Syst Rev* 2006;(4):CD004510.
13 Afolabi BB, Lesi FE, Merah NA. Regional versus general anaesthesia for caesarean section. *Cochrane Database Syst Rev* 2006;(4):CD004350.
14 Uman LS, Chambers CT, McGrath PJ, Kisely S. Psychological interventions for needle-related procedural pain and distress in children and adolescents. *Cochrane Database Syst Rev* 2006;(4):CD005179.
15 Moore M, Little P. Humidified air inhalation for treating croup. *Cochrane Database Syst Rev* 2006;(3):CD002870.

Accepted: 11 March 2009

Endpiece

The belly of gourmets

The belly of gourmets has reached such daintiness that they cannot taste a fish unless they see it swimming and palpitating in the very dining room. What a lot is being added to the ingenuity of excessive extravagance! And how much more delicately and elegantly does our madness invent something while despising anything ordinary!

Seneca (4 BC-AD 62). *Natural Questions*. 3:18.3.

Submitted by Jeremy Hugh Baron, honorary professorial lecturer, Mount Sinai School of Medicine, New York

Cite this as: *BMJ* 2008;337:a555

Has payment by results affected the way that English hospitals provide care? Difference-in-differences analysis

Shelley Farrar,[1] Deokhee Yi,[1] Matt Sutton,[2] Martin Chalkley,[3] Jon Sussex,[4] Anthony Scott[5]

EDITORIAL by Anderson

[1]Health Economics Research Unit, University of Aberdeen, Aberdeen AB25 2ZD
[2]Health Methodology Research Group, University of Manchester, Manchester M13 9PL
[3]Department of Economic Studies, University of Dundee, Dundee DD1 4HN
[4]Office of Health Economics, London SW1A 2DY
[5]Melbourne Institute of Applied Economic and Social Research, University of Melbourne, Parkville, VIC 3010, Australia
Correspondence to: S Farrar
s.farrar@abdn.ac.uk

Cite this as: *BMJ* 2009;339:b3047
doi: 10.1136/bmj.b3047

ABSTRACT

Objective To examine whether the introduction of payment by results (a fixed tariff case mix based payment system) was associated with changes in key outcome variables measuring volume, cost, and quality of care between 2003/4 and 2005/6.

Setting Acute care hospitals in England.

Design Difference-in-differences analysis (using a control group created from trusts in England and providers in Scotland not implementing payment by results in the relevant years); retrospective analysis of patient level secondary data with fixed effects models.

Data sources English hospital episode statistics and Scottish morbidity records for 2002/3 to 2005/6.

Main outcome measures Changes in length of stay and proportion of day case admissions as a proxy for unit cost; growth in number of spells to measure increases in output; and changes in in-hospital mortality, 30 day post-surgical mortality, and emergency readmission after treatment for hip fracture as measures of impact on quality of care.

Results Length of stay fell more quickly and the proportion of day cases increased more quickly where payment by results was implemented, suggesting a reduction in the unit costs of care associated with payment by results. Some evidence of an association between the introduction of payment by results and growth in acute hospital activity was found. Little measurable change occurred in the quality of care indicators used in this study that can be attributed to the introduction of payment by results.

Conclusion Reductions in unit costs may have been achieved without detrimental impact on the quality of care, at least in as far as these are measured by the proxy variables used in this study.

WHAT IS ALREADY KNOWN ON THIS TOPIC

In April 2002 the Department of Health in England outlined plans to introduce a new system of financing hospitals, called payment by results

The system directly links the income hospitals receive with the number and case mix of patients treated

Similar systems adopted in other countries have been shown to have effects on the cost, quality, and volume of patients' care

WHAT THIS STUDY ADDS

The results show a reduction in unit costs (with average length of stay and proportion of day cases as proxies) associated with the introduction of payment by results

The evidence on volume of care is more equivocal but suggests that the volume of spells of care increased in response to the new payment system

Payment by results had no effect on the outcomes used as proxies for the quality of care

This article is an abridged version of a paper that was published on bmj.com. Cite this article as: *BMJ* 2009;339:b3047

INTRODUCTION

In April 2002 the Department of Health in England outlined plans to introduce a new system of financing hospitals, called "payment by results," a fixed price system that makes a direct link between a hospital's income and the number and case mix of patients treated.[1 2] Under payment by results, prices (or tariffs) for hospital care are defined in terms of healthcare resource group (HRG) spells of stay in hospital. A spell of activity is a hospital stay from admission to discharge and is a measure of the hospital's output. An HRG code is assigned to each spell of activity.

The tariff system has various characteristics that shape the incentives of the system. The payment the hospital receives for providing an HRG spell is determined by whether that spell is elective.[3] A single tariff exists to reimburse trusts for each HRG for day case and inpatient elective care.[1] Payment by results removes the option for hospitals to use their own cost circumstances to negotiate for higher payment. Any surplus earned by a hospital because it reduces its unit costs can be retained by the hospital.

Providers can act to reduce costs in various ways: by increasing efficiency, by selecting patients who need less resource intensive care, or by reducing the level of resources in the provision of care, which may compromise the quality of care.[4] Payment by results makes a link between the number of patients treated and the payment to a hospital, creating an incentive to provide more of those treatments. However, this will be the case only if the payment for a treatment is higher than the costs of providing the treatment, such that a surplus can be made.

Our objective was to examine whether changes in key outcome variables measuring the volume, cost, and quality of care during 2004/5 and 2005/6 were associated with tariff funding introduced for NHS hospitals in England under the payment by results policy.

METHODS

Study design

We constructed a quasi-experiment using various naturally occurring control groups. The tariff system under payment by results was first applied to marginal changes in output for 15 HRGs in 2003/4 and extended to a further 33 HRGs in 2004/5.[1 3] For a subset of NHS trusts—foundation trusts and three early implementing non-foundation trusts—payment by results was applied to most inpatient, day case, and outpatient output activity in 2004/5. For the remaining non-foundation trusts, it was applied to most elective admissions in 2005/6.[5] Throughout the period

2003/4 to 2005/6 the tariff system was not adopted in Scotland, and for most trusts in England it was not adopted extensively until 2005/6. This phased and partial introduction provides a series of treatment and control groups.

Difference-in-differences analysis is commonly used in the evaluation of impacts of policy. It uses the assumption that unobserved differences between groups are the same over time.[6 7] Non-foundation trusts and foundation trusts were subject to virtually all the same changes in policy over the period of study. Although differences exist in aspects of the healthcare systems and policies of England and Scotland, they are broadly comparable.

We used differences between the phasing in of payment by results by foundation trusts, non-foundation trusts, and Scotland to estimate the effects of the introduction of the policy. For each of the outcome measures used in the study, we made three main comparisons: between foundation trusts (treatment group) and non-foundation trusts (control) for changes from 2003/4 to 2004/5; between foundation trusts (treatment group) and providers in Scotland (control) for changes from 2003/4 to 2004/5; and between non-foundation trusts (treatment group) and providers in Scotland (control) for changes from 2004/5 to 2005/6. We made a fourth comparison to analyse changes in foundation trusts over the full first two years of implementation of payment by results: between foundation trusts (treatment group) and providers in Scotland (control) for changes from 2003/4 to 2005/6.

Data sources

We used data from the hospital episode statistics for 2002/3 to 2005/6 for England and from the Scottish morbidity records for Scotland. The English data represent 248 acute care trusts, and the Scottish data come from 49 hospitals. Thirty four of the English trusts gained foundation status or were early implementers of payment by results during the period of analysis.

We used log length of stay and day cases as a proportion of elective admissions as measures of unit costs or efficiency for hospital admissions. For quality of care we used in-hospital mortality, 30 day post-surgical mortality, and emergency readmission after treatment for hip fracture.

Econometric methods

We used fixed effects to control for differences between the characteristics of HRGs and trusts that were unobserved and did not change over time. Some unobserved factors are likely to vary both within trusts and within HRGs. In addition, some trusts may be more efficient at providing particular HRGs, and some HRGs will be provided by particular types of trust. We therefore interacted the two variables to create fixed effects for each combination of HRG and trust. A total of 81 820 fixed effects exist.

Table 1 | Effects of payment by results on measures of unit costs: length of stay (days) and proportion of day cases (change in percentage points)

Treatment group	Control group	Years	Length of stay		Day case proportion	
			Change	No*	Change	No*
Foundation trusts	Non-foundation trusts	2003/4-2004/5	0.02†	1091	0.4†	8266
Foundation trusts	Scotland	2003/4-2004/5	−0.08†	2724	0.4†	2810
Non-foundation trusts‡	Scotland‡	2004/5-2005/6	−0.03†	1704	0.8†	6842
Foundation trusts	Scotland	2003/4-2005/6	−0.18†	1248	1.5†	1178

*Number of observations in 1000s. †Significant at P<0.01. ‡Elective only.

Table 2 | Effects of payment by results on growth in volume of care (change in percentage points)

Treatment group	Control group	Years	Growth in volume	
			Change	No*
Foundation trusts	Non-foundation trusts	2003/4-2004/5	−0.25	82 816
Foundation trusts	Scotland	2003/4-2004/5	1.33†	20 431
Non-foundation trusts‡	Scotland‡	2004/5-2005/6	2.57†	51 249
Foundation trusts	Scotland	2003/4-2005/6	4.95†	21 598

*Number of observations in 1000s. †Significant at P<0.01. ‡Elective only.

RESULTS

Impact on unit costs

The results for proxies of unit costs were consistent across most of the difference-in-differences analyses: they suggest that unit costs fell more quickly where payment by results was implemented (table 1). In all but one of the difference-in-differences comparisons used, foundation trusts and non-foundation trusts responded in the expected way to the incentives associated with payment by results.

Length of stay fell more quickly where payment by results was implemented. In 2004/5 the average length of stay in foundation trusts fell by 0.08 of a day more than it did in Scotland. This equates to eight inpatient days saved for every 100 inpatient admissions. For non-foundation trusts that implemented payment by results in 2005/6 the difference was in the same direction, falling by 0.03 days or a saving of three days per 100 admissions.

The proportion of elective care provided as day cases increased more quickly where payment by results was implemented. This change was seen for both foundation trusts and non-foundation trusts. In 2004/5 day cases as a proportion of elective care grew by 0.4 percentage points more in foundation trusts than in other providers, and in 2005/6 the proportion of elective care provided as day cases grew by 0.8 percentage points more in non-foundation trusts than in Scotland.

One difference-in-differences analysis using non-foundation trusts as the control for foundation trusts in 2004/5 did not support expectations. The non-foundation trusts reduced length of stay more quickly than did the foundation trusts.

Impact on volume of spells

Using Scotland as the control group, we found that both foundation trusts and non-foundation trusts experienced growth in volume associated with payment by results.

The number of foundation trusts' spells grew by 1.33 percentage points more than for providers in Scotland in 2004/5, and non-foundation trusts' spells grew by 2.57 percentage points over and above growth in Scotland. However, the comparison of the tariffed response of foundation trusts with the non-tariffed response of non-foundation trusts, in 2004/5, shows that foundation trusts' spells did not increase relative to the non-foundation trusts. Table 2 summarises the difference-in-differences results for the growth in volume of elective and non-elective spells.

Impact on quality of care

We found little evidence of an association between the introduction of payment by results and a change in the quality of care. The only year on year result with statistical significance was the difference in the change in in-hospital mortality for foundation trusts compared with Scotland (see bmj.com).

DISCUSSION

Most of our tests on mean length of stay and the proportion of day case activity in total elective admissions are consistent with a reduction in unit costs associated with introduction of a fixed price tariff under payment by results. Of the five tests on the effect of the tariff on the volume of care, four provided evidence that the tariff was associated with growth in activity. During their first full year of payment by results, foundation trusts did not experience the greater reductions in average length of stay and higher growth expected relative to non-foundation trusts. Only when they were compared with Scotland was a difference apparent. Foundation trusts by definition are the more efficient and well managed trusts in England, with shorter lengths of stay, and may have experienced less pressure from the tariff to reduce costs. In addition, the low absolute length of stay may have made it more difficult for foundation trusts to produce efficiency savings in this area. However, these foundation trusts should have been in a better position to have benefited from increases in the volume of patients treated, but we did not see this in the data. The results for the effects on quality of care were generally not statistically significant and may be taken as evidence that payment by results did not have an adverse impact on the quality of care.

Limitations

Mortality has been criticised as an insufficiently sensitive measure of change in the quality of care.[8] However, it is widely used in the absence of other routine data.[4] In-hospital mortality specifically is open to criticism as a means of measuring quality of care in a system that is also reducing length of stay. Dimensions of quality of care that we have not captured could have been adversely affected by payment by results.

Ideal conditions for applying difference-in-differences analysis require that in the absence of the policy intervention the average outcomes for the treatment and control groups would be parallel over time. Waiting time targets of the same level were used in England and Scotland throughout our study with the exception of the six months target in England for the end of 2004/5 compared with nine months in Scotland. However, stronger incentive mechanisms were associated with these targets in England.[9] The additional incentive in England to increase the throughput of patients may have had a confounding effect on the results for the growth in the volume of spells in 2004/5.

The results of this study refer specifically to the effects of payment by results in its early years. As hospitals become more familiar with the effects on their own costs and revenues, and more confident about the permanence of payment by results, they may be less cautious in their responses. Our study covers a period when most hospitals affected by payment by results were on a transition pathway that partially protected them from financial losses associated with the tariff. As this protection reduces, the incentives of payment by results may be strengthened.

Implications

A report by the Audit Commission on the effects of the introduction of payment by results in England draws on a similar time period to the research reported here, and the findings mirror those of our evaluation.[10] Our analyses suggest that payment by results is capable of achieving real changes in delivery of health care in hospitals in England. Our evaluation suggests a potentially rich set of further research questions. Our approach has of necessity been a general one; much remains to be learnt about the impact of payment by results at a more disaggregate level.

Contributors: See bmj.com.
Funding: This research was funded by the Department of Health through the Policy Research Programme. The researchers are independent of the funders. The views expressed in this paper are those of the authors and not the funding body.
Competing interests: None declared.
Ethical approval: Not needed.

1 Department of Health. *Reforming NHS financial flows: introducing payment by results*. London: Department of Health, 2002.
2 Sussex J, Street A. *Activity-based financing for hospitals: English policy and international experience*. London: Office of Health Economics, 2005.
3 Department of Health. *NHS reference costs 2003 and national tariff 2004 (payment by results core tools 2004)*. London: Department of Health, 2004.
4 Chalkley M, Malcolmson JM. Government purchasing of health services. In: Culyer AJ, Newhouse JP, eds. *Handbook of health economics*. Vol 1A. Amsterdam: North-Holland, 2000:847-90.
5 Department of Health. *Implementing payment by results. technical guidance 2005/06 (version 1.0)*. London: Department of Health, 2004.
6 Blundell R, Costa Dias M. Evaluation methods for non-experimental data. *Fiscal Studies* 2000;21:427-68.
7 Qin J, Zhang B. Empirical likelihood based on difference-in-differences estimators. *J R Stat Soc* 2008;70:329-49.
8 Pitches DW, Mohammed MA, Lilford RJ. What is the empirical evidence that hospitals with higher-risk adjusted mortality rates provide poorer quality of care? A systematic review of the literature. *BMC Health Serv Res* 2007;7:91.
9 Propper C, Sutton M, Whitnall C, Windmeijer F. *Did 'targets and terror' reduce waiting times in England for hospital care?* Bristol: University of Bristol, 2007. (Centre for Market and Public Organisation working paper series number 07/179.)
10 Audit Commission. *The right result? Payment by results 2003-07*. London: Audit Commission, 2008.

Accepted: 17 May 2009

Hormonal contraception and risk of venous thromboembolism: national follow-up study

Øjvind Lidegaard,[1] Ellen Løkkegaard,[2] Anne Louise Svendsen,[3] Carsten Agger[4]

EDITORIAL by Dunn
RESEARCH, p 561
CLINICAL REVIEW, p 563

[1]Gynaecological Clinic, Rigshospitalet, Copenhagen University, DK-2100 Copenhagen, Denmark
[2]Department of Obstetrics and Gynaecology, Hillerød Hospital, Copenhagen University
[3]Department of Biostatistics, University of Copenhagen
[4]Research Centre for Prevention and Health, Glostrup Hospital, University of Copenhagen
Correspondence to: Ø Lidegaard
Lidegaard@rh.regionh.dk

Cite this as: *BMJ* 2009;339:b2890
doi: 10.1136/bmj.b2890

ABSTRACT

Objective To assess the risk of venous thrombosis in current users of different types of hormonal contraception, focusing on regimen, oestrogen dose, type of progestogen, and route of administration.

Design National cohort study.

Setting Denmark, 1995-2005.

Participants Danish women aged 15-49 with no history of cardiovascular or malignant disease.

Main outcome measures Adjusted rate ratios for all first time deep venous thrombosis, portal thrombosis, thrombosis of caval vein, thrombosis of renal vein, unspecified deep vein thrombosis, and pulmonary embolism during the study period.

Results 10.4 million woman years were recorded, 3.3 million woman years in receipt of oral contraceptives. In total, 4213 venous thrombotic events were observed, 2045 in current users of oral contraceptives. The overall absolute risk of venous thrombosis per 10 000 woman years in non-users of oral contraceptives was 3.01 and in current users was 6.29. Compared with non-users of combined oral contraceptives the rate ratio of venous thrombembolism in current users decreased with duration of use (<1 year 4.17, 95% confidence interval 3.73 to 4.66, 1-4 years 2.98, 2.73 to 3.26, and >4 years 2.76, 2.53 to 3.02; P<0.001) and with decreasing dose of oestrogen. Compared with oral contraceptives containing levonorgestrel and with the same dose of oestrogen and length of use, the rate ratio for oral contraceptives with norethisterone was 0.98 (0.71 to 1.37), with norgestimate 1.19 (0.96 to 1.47), with desogestrel 1.82 (1.49 to 2.22), with gestodene 1.86 (1.59 to 2.18), with drospirenone 1.64 (1.27 to 2.10), and with cyproterone 1.88 (1.47 to 2.42). Compared with non-users of oral contraceptives, the rate ratio for venous thromboembolism in users of progestogen only oral contraceptives with levonorgestrel or norethisterone was 0.59 (0.33 to 1.03) or with 75 µg desogestrel was 1.12 (0.36 to 3.49), and for hormone releasing intrauterine devices was 0.90 (0.64 to 1.26).

Conclusion The risk of venous thrombosis in current users of combined oral contraceptives decreases with duration of use and decreasing oestrogen dose. For the same dose of oestrogen and the same length of use, oral contraceptives with desogestrel, gestodene, or drospirenone were associated with a significantly higher risk of venous thrombosis than oral contraceptives with levonorgestrel. Progestogen only pills and hormone releasing intrauterine devices were not associated with any increased risk of venous thrombosis.

WHAT IS ALREADY KNOWN ON THIS TOPIC

Previous studies have shown an increased risk of venous thrombosis with use of combined oral contraceptives and a higher risk with use of combined pills containing the progestogens desogestrel or gestodene than those containing levonorgestrel

WHAT THIS STUDY ADDS

The risk of venous thrombosis in users of combined oral contraceptives decreases with decreasing dose of oestrogen
The risk of venous thrombosis from pills containing drospirenone corresponds to those containing desogestrel or gestodene and is higher than those with levonorgestrel
Progestogen only pills and hormone releasing intrauterine devices did not confer any increased risk of venous thrombosis
The absolute risk of venous thrombosis with use of any types of combined oral contraceptives in young women is less than one in 1000 user years

This article is an abridged version of a paper that was published on bmj.com. Cite this article as: *BMJ* 2009;339:b2890

INTRODUCTION

Studies that have shown an increased risk of venous thrombosis with combined oral contraceptives have generally found a higher risk during the first year of use and with pills containing desogestrel or gestodene.[1-9]

With the shift from combined pills containing 50 µg oestrogen to those containing 30-40 µg, a decrease in the risk of venous thrombosis would be expected. Results have, however, been conflicting,[9 10] and evidence of a further decrease in risk with oestrogen reduced to 20 µg is lacking.[9] Evidence is also sparse on the risk of venous thrombosis with oral contraceptives containing the new progestogen drospirenone, progestogen only pills with 75 µg desogestrel, and hormone releasing intrauterine devices.

We assessed the risk of venous thrombosis in current users of hormonal contraception, focusing on duration of use, regimen (combined pills versus progestogen only pills), and the effect of oestrogen dose, progestogen type, and route of administration.

METHODS

This study was designed as a historical cohort study, with linkage between four Danish registries: the National Registry of Medicinal Products Statistics (prescriptions), the National Registry of Patients (discharge diagnoses and surgical codes from all Danish hospitals), Statistics of Denmark (education), and the Central Person Registry (addresses and vital status).

We identified Danish women aged 15-49 from 1 January 1995 to 31 December 2005. After exclusions the study population comprised non-pregnant women with no previous cancer or cardiovascular diseases (new cases during the study period were censored at the date of diagnosis). Women who emigrated were censored at the time they left the country.

Our end points were first time deep venous thrombosis, portal thrombosis, thrombosis of caval vein, thrombosis of renal vein, unspecified deep vein thrombosis, and pulmonary embolism during the study period.

Current use of hormonal contraception was defined as having a valid prescription when admitted to hospital. Previous use was defined as any previous recorded use during the study period, and never use as no recorded prescription for hormonal contraception during the study period. Length of use was defined as the sum of valid prescriptions, with periods of non-use subtracted if they occurred between periods of use.

Hormone releasing intrauterine devices were assumed to be used for an average of three years. If oral contraception was prescribed before the three years expired, the device was considered to have been removed at the time the pill was prescribed.

Hormonal contraception was categorised according to usage (current, previous, never), regimen (combined pill, progestogen only pill, hormone releasing intrauterine device), oestrogen dose (50 µg, 30-40 µg, 20 µg), progestogen type (norethisterone, levonorgestrel, norgestimate, desogestrel, gestodene, drospirenone, cyproterone), and length of use of combined pills in current users (<1 year, 1-4 years, >4 years).

Crude incidence rates and adjusted rate ratios of venous thrombosis in women using different types of hormonal contraception

Characteristics	Woman years	% of woman years	No of women with venous thrombosis	Rate per 10 000 woman years	Adjusted rate ratio (95% CI)
Age group					
15-19	1 359 821	13.0	250	1.84	0.39 (0.33 to 0.45)*
20-24	1 491 764	14.3	444	2.98	0.62 (0.54 to 0.70)*
25-29	1 491 959	14.3	537	3.60	0.86 (0.76 to 0.96)*
30-34	1 587 896	15.2	598	3.77	Reference
35-39	1 628 852	15.6	685	4.21	1.18 (1.05 to 1.32)*
40-44	1 518 172	14.5	797	5.25	1.57 (1.41 to 1.74)*
45-49	1 368 909	13.1	902	6.59	2.09 (1.88 to 2.32)*
Total	10 447 373	100.0	4213	4.03	—
Previous use					
Never	4 813 053	46.1	1467	3.05	Reference
Former	2 278 576	21.8	667	2.93	1.08 (0.98 to 1.18)†
Current use					
Non-use (never or former use of oral contraceptives)	7 194 242	67.9	2168	3.01	Reference
Current use of oral contraceptives	3 253 131	31.1	2045	6.29	2.83 (2.65 to 3.01)†
Use of combined pill:					
<1 year	684 061	21.6	443	6.48	4.17 (3.73 to 4.66)†
1-4 years	1 449 000	45.8	787	5.43	2.98 (2.73 to 3.26)†
>4 years	1 031 953	32.6	793	7.68	2.76 (2.53 to 3.02)†
Oral contraceptives with 50 µg oestrogen	82 902	2.5	65	7.84	2.67 (2.09 to 3.42)†
Oral contraceptives with 20-40 µg oestrogen and:					
Levonorgestrel	367 408	10.9	201	5.47	2.02 (1.75 to 2.34)†
Desogestrel or gestodene	2 008 262	59.8	1370	6.82	3.55 (3.30 to 3.83)†
Drospirenone	131 541	3.9	103	7.83	4.00 (3.26 to 4.91)†
Progestogen only:					
Levonorgestrel 30 µg or norethisterone 350 µg	65 820	0.6	12	1.82	0.59 (0.33 to 1.04)†
Desogestrel 75 µg	9044	0.1	3	3.32	1.10 (0.35 to 3.41)†
Hormone releasing intrauterine device	101 351	1.0	34	3.35	0.89 (0.64 to 1.26)

*Adjusted for current use of oral contraceptives, calendar year, and educational level.
†Adjusted for age, calendar year, and educational level.

Progestogen only pills were subdivided into those with 30 µg levonorgestrel or 350 µg norethisterone and those with 75 µg desogestrel. We chose non-users of oral contraceptives (never users plus former users) as our reference group.

We obtained information on redeemed drugs for confounding factors such as diabetes and heart disease (see bmj.com). Educational level was categorised as primary school only, secondary school only, any school with three or four years of further education, and secondary school with five or six years of further education.

Statistical analysis

Data were analysed using Poisson regression. The data consisted of time at risk (woman years) and number of venous thrombotic events for each combination of contraception, length of use, age band, and educational level. Age was the timescale in analyses and divided into five year bands, assuming a linear trend in risk of venous thromboembolism within each band. Confounders were retained in the multivariate analysis if they changed the estimates by more than 5%.

Absolute crude risk estimates and adjusted rate ratios (95% confidence intervals) were calculated for combinations of oestrogen dose, progestogen type, and length of use. In addition we calculated the influence of progestogen types after adjustment for length of use.

RESULTS

The analysis included 3.4 million woman years of current use, 2.3 million woman years of former use, 4.8 million woman years of never use, or 10.4 million woman years of observation (table). A total of 4213 first time venous thrombotic events were recorded during the study period and of these 2045 were among current users of hormonal contraception.

Age, calendar year, and education were significant confounders. There was no interaction between age and the rate ratios.

Young women more often used newer pills than older women, who more often used hormone releasing intrauterine devices (see bmj.com). The incidence of venous thromboembolism increased with age (15-19 years, 1.84 per 10 000 woman years; 45-49 years, 6.59 per 10 000 woman years; table). The incidence also increased during the study period, on average by 1.05 (95% confidence interval 1.04 to1.06) per calendar year. Finally, the risk of venous thromboembolism increased with decreasing education. Using the least educated women (primary school only) as the reference group, the rate ratios of venous thromboembolism for those with secondary school education only was 0.52 (0.46 to 0.59), with any schooling and three or four years of further education was 0.58 (0.54 to 0.63), and with secondary school education with five or six years of further education was 0.43 (0.39 to 0.47).

The crude incidence of venous thromboembolism among non-users of hormonal contraceptives was 3.01 per 10 000 woman years, and among current users of oral contraceptives was on average 6.29 per 10 000 woman years (table).

Combined oral contraceptives

The risk among women using combined pills decreased with duration of use (adjusted rate ratio for first year, 4.17, 95% confidence interval 3.73 to 4.66; >4 years 2.76, 2.53 to 3.02; table).

The risk among current users of combined pills was also influenced by oestrogen dose and progestogen type (see bmj.com). For a given progestogen type and after adjustment for length of use, the risk of venous thromboembolism decreased with decreasing dose of oestrogen (see bmj.com). A reduction in dose from 50 µg to 30-40 µg in pills containing levonorgestrel reduced the risk by 17% (not significant), and for those containing norethisterone by 32% (not significant). Furthermore, a reduction in oestrogen dose from 30-40 µg to 20 µg for pills containing desogestrel or gestodene reduced the risk of venous thromboembolism by 18% (7% to 27%).

Compared with current users of oral contraceptives with levonorgestrel, using the same dose of oestrogen and after adjustment for duration of use, the rate ratios of venous thromboembolism in women using pills containing norethisterone was 0.98 (0.71 to 1.37), norgestimate 1.19 (0.96 to 1.47), desogestrel 1.82 (1.49 to 2.22), gestodene 1.86 (1.59 to 2.18), drospirenone 1.64 (1.27 to 2.10), and cyproterone 1.88 (1.47 to 2.42; see bmj.com).

Progestogen only pills

Progestogen only pills containing levonorgestrel 30 µg or norethisterone 350 µg, as well as desogestrel 75 µg did not confer any increased risk of venous thromboembolism when compared with non-users of oral contraceptives. The adjusted rate ratio for venous thromboembolism in women using hormone releasing intrauterine devices was 0.89 (0.64 to 1.26; table).

DISCUSSION

The risk of venous thromboembolism in current users of combined oral contraceptives decreases with duration of use and decreasing oestrogen dose. For the same dose of oestrogen and the same length of use, oral contraceptives containing desogestrel, gestodene, or drospirenone were associated with a higher risk of venous thromboembolism than pills containing levonorgestrel. Progestogen only pills and hormone releasing intrauterine devices did not confer any increased risk.

The extent of an overall risk estimate of venous thromboembolism in current users of oral contraceptives depends on several factors. Exclusion of women with previous thrombosis and cancer from the reference group would increase the overall risk estimate because of the decreased risk in this group. The estimate would also be increased by the inclusion of relatively more new users or short term users of oral contraceptives, or if many women were using oral contraceptives that contained desogestrel, gestodene, or drospirenone than those containing levonorgestrel. The inclusion of pregnant women in the reference group or women using progestogen only pills in the oral contraceptives group would, however, decrease the overall estimate.

Newer absolute risk estimates would be expected to be higher than older estimates because of improvements in diagnosing venous thromboembolism. We controlled for this by including calendar year in the multivariate analyses.

Reducing the dose of oestrogen from 50 µg to 30-40 µg non-significantly reduced the risk of venous thromboembolism by 17-32%. Reducing the dose from 30-40 µg to 20 µg in users of pills containing desogestrel or gestodene significantly reduced the risk by 18% (95% confidence interval 7% to 27%), after adjustment for duration of use of oral contraceptives. Without this adjustment the association was confounded and not significant.

DISCUSSION

The higher risk of venous thromboembolism in users of pills containing desogestrel or gestodene compared with levonorgestrel is in line with several studies,[1 6-9 11] although not all[11 12] (see bmj.com). In our study, oral contraceptives with norgestimate were associated with about the same risk of venous thromboembolism as those with levonorgestrel.

Studies have shown a threefold increased risk of venous thromboembolism in women using oral contraceptives with cyproterone compared with non-users,[9 13 14] results similar to ours. The European active surveillance study found 9.1 venous thrombotic events per 10 000 user years (26 events) in women using oral contraceptives with drospirenone compared with 8.0 per 10 000 woman years (n=25) in those using pills that contained levonorgestrel, and 2.3 per 10 000 woman years in non-pregnant non-users (n=5).[12]

In users of pills containing drospirenone we found an incidence of venous thromboembolism of 7.9 per 10 000 woman years and an adjusted rate ratio of 1.64 (1.27 to 2.10) when compared with users of pills containing levonorgestrel, and of 4.00 (3.26 to 4.91) when compared with non-users. These estimates were based on 103 venous thrombotic events in users of pills with drospirenone. The four times increased risk of venous thromboembolism compared with non-users is in accordance with the surveillance study.[12] The higher risk of venous thromboembolism when compared with users of pills with levonorgestrel is a new finding. The reason for this rate ratio was primarily a lower risk in users of pills containing levonorgestrel in our study compared with the same estimates in the surveillance study.[12] The risk estimate was of the same magnitude as for pills containing desogestrel or gestodene with the same dose of oestrogen and same length of use, results in accordance with two other studies.[12 15]

Finally, the lack of increased risk of venous thrombosis by use of progestogen only pills is in line with previous findings.[9] The reduced risk of venous thrombosis with increasing length of education could be attributed to the higher prevalence of obesity in the least educated women.

Strengths and limitations of the study

The registry linkage design of this study has advantages and limitations. One strength is the high external validity, as we included all Danish women aged 15-49 who fulfilled the inclusion criteria. Recall bias was eliminated,

as the national prescription registry provided precise data on use of hormonal contraception. The national approach ensured a relatively high statistical power by including 4213 venous thrombotic events. Consequently we were able to assess the risk for subtypes of oral contraceptives and to consider separately oestrogen dose, type of progestogen, and length of use in the risk of venous thrombosis. Finally, the cohort design allowed the calculation of absolute risk estimates as well as rate ratios between different types of hormonal contraception.

One limitation of our study was the lack of data on family predisposition and body mass index, two potential confounders. When oral contraceptives containing desogestrel or gestodene were introduced in the late 1980s they were considered safer than the older types of oral contraceptives. Therefore, women with a family predisposition for venous thromboembolism were preferentially prescribed these new pills.[16] However, this preferential prescribing stopped after new studies were published in the 1990s.[9]

Being overweight predisposes to venous thromboembolism. If some oral contraceptives are preferentially prescribed to women with an increased body mass index, then the risk of these pills could be overestimated. Controlling for these two potential confounders in a previous study did not change the risk estimates of venous thromboembolism.[9]

When the new pill containing drospirenone was introduced in Denmark in 2001, it was not considered as safer than the older pills. Therefore preferential prescribing to women at increased risk of venous thromboembolism is not expected. In our study we were able to investigate the proportion of women taking different oral contraceptives. We found the same or lower prevalence in users of pills containing drospirenone compared with those containing levonorgestrel, suggesting the same baseline health status. Therefore bias from failing to control for body mass index and family predisposition was probably small.

Another limitation was that the registry approach did not permit us to evaluate the validity of each included diagnosis of venous thromboembolism. They were identified as the final discharge diagnosis as reported to the National Registry of Patients. The inclusion of about 10% uncertain diagnoses may have biased our results only if the misclassification was differential, implying fewer or more women with an uncertain diagnosis among current users of oral contraceptives compared with non-users. In this study the slightly lower risk estimates among current users of oral contraceptives compared with our previous study in which these uncertain cases were excluded [9] might suggest fewer users of oral contraceptives among women with an uncertain diagnosis than among those with a validated and confirmed diagnosis. The reduction in oestrogen dose of oral contraceptives through the study period, however, could also have contributed to the reduced risk estimates. Finally, registry data do not include information on lifestyle such as being sedentary, long distance flights, and limited mobility at home.

Clinical implications

For women of normal weight and without known genetic predispositions, we recommend a low dose combined pill as first choice for contraception. For women genetically predisposed to venous thrombosis who want hormonal contraception, however, a progestogen only pill or hormone releasing intrauterine device seems to be the appropriate first choice.

We thank the National Registry of Medicinal Products Statistics for providing the data on use of hormonal contraception and the National Board of Health for providing the data from the National Registry of Patients.

Contributors: See bmj.com.

Funding: The Gynaecological Clinic, Rigshospitalet covered the expenses for this study.

Competing interests: ØL has received grants from pharmaceutical companies for cover of research expenses and has received fees for speeches on pharmacoepidemiological topics.

Ethical approval: This study was approved by the Danish Data Protection Agency (J No 2003-41-2872). Ethical approval is not requested for registry based studies in Denmark.

1 Bloemenkamp KWM, Rosendaal FR, Helmerhorst FM, Büller HR, Vandenbroucke JP. Enhancement by factor V Leiden mutation of risk of deep-vein thrombosis associated with oral contraceptives containing a third-generation progestagen. *Lancet* 1995;346:1593-6.
2 Jick H, Jick SS, Gurewich V, Myers MW, Vasilakis C. Risk of idiopathic cardiovascular death and nonfatal venous thromboembolism in women using oral contraceptives with differing progestagen components. *Lancet* 1995;346:1589-93.
3 World Health Organization Collaborative Study on Cardiovascular Disease and Steroid Hormone Contraception. Effect of different progestagens in low oestrogen oral contraceptives on venous thromboembolic disease. *Lancet* 1995;346:1582-8.
4 Spitzer WO, Lewis MA, Heinemann LAJ, Thorogood M, MacRae KD. Third generation oral contraceptives and risk of venous thromboembolic disorders: an international case-control study. *BMJ* 1996;312:83-8.
5 Farmer RDT, Lawrenson RA, Thompson CR, Kennedy JG, Hambleton IR. Population-based study of risk of venous thromboembolism associated with various oral contraceptives. *Lancet* 1997;349:83-8.
6 Bloemenkamp KWM, Rosendaal FR, Büller HR, Helmerhorst FM, Colly LP, Vandenbroucke JP. Risk of venous thrombosis with use of current low-dose oral contraceptives is not explained by diagnostic suspicion and referral bias. *Arch Intern Med* 1999;159:65-70.
7 Todd J-C, Lawrenson R, Farmer RDT, Williams TJ, Leydon GM. Venous thromboembolic disease and combined oral contraceptives: a re-analysis of the MediPlus database. *Hum Reprod* 1999;14:1500-5.
8 Jick H, Kaye JA, Vasilakis-Scaramozza C, Jick SS. Risk of venous thromboembolism among users of third generation oral contraceptives compared with users of oral contraceptives with levonorgestrel before and after 1995: cohort and case-control analysis. *BMJ* 2000;321:1190-5.
9 Lidegaard Ø, Edström B, Kreiner S. Oral contraceptives and venous thromboembolism. A five-year national case-control study. *Contraception* 2002;65:187-96.
10 World Health Organization Collaborative Study of Cardiovascular Disease and Steroid Hormone Contraception. Venous thromboembolic disease and combined oral contraceptives: results of international multicentre case-control study. *Lancet* 1995;346:1575-82.
11 Lewis MA, MacRae KD, Kühl-Habich D, Bruppacher R, Heinemann LAJ, Spitzer WO. The differential risk of oral contraceptives: the impact of full exposure history. *Hum Reprod* 1999;14:1493-9.
12 Dinger JC, Heinemann LAJ, Kühl-Habich D. The safety of a drospirenone-containing oral contraceptive: final results from the European Active Surveillance study on oral contraceptives based on 142,475 women-years of observation. *Contraception* 2007;75:344-54.
13 Parkin L, Skegg DCG, Wilson M, Herbison GP, Paul C. Oral contraceptives and fatal pulmonary embolism. *Lancet* 2000;355:2133-4.
14 Vasilakis-Scaramozza C, Jick H. Risk of venous thromboembolism with cyproterone and levonorgestrel contraceptives. *Lancet* 2001;358:1427-9.
15 Seeger JD, Loughlin J, Eng PM, Clifford CR, Cutone J, Walker AM. Risk of thromboembolism in women taking ethinylestradiol/drospirenone and other oral contraceptives. *Obstet Gynecol* 2007;110:587-93.
16 Lidegaard Ø. The influence of thrombotic risk factors when oral contraceptives are prescribed. *Acta Obstet Gynaecol Scand* 1997;76:252-60.

Accepted: 27 May 2009

pico

The venous thrombotic risk of oral contraceptives, effects of oestrogen dose and progestogen type: results of the MEGA case-control study

A van Hylckama Vlieg,[1] F M Helmerhorst,[1 2] J P Vandenbroucke,[1] C J M Doggen,[1] F R Rosendaal[1 3 4]

EDITORIAL by Dunn
RESEARCH, p 557
CLINICAL REVIEW, p 563

[1]Department of Clinical Epidemiology, Leiden University Medical Center, C7-P, PO Box 9600, NL-2300 RC Leiden, Netherlands
[2]Department of Gynaecology and Reproductive Medicine, Leiden University Medical Center
[3]Department of Thrombosis and Haemostasis, Leiden University Medical Center
[4]Einthoven Laboratory for Experimental Vascular Medicine, Leiden University Medical Center
Correspondence to: F R Rosendaal
F.R.Rosendaal@lumc.nl

Cite this as: *BMJ* 2009;339:b2921
doi: 10.1136/bmj.b2921

STUDY QUESTION What is the risk of venous thrombosis from oral contraceptives currently available in the Netherlands, and how much does the dose of oestrogen and type of progestogen matter?

SUMMARY ANSWER Oral contraceptives increased the risk of venous thrombosis fivefold. The risk clearly differed by type of progestogen and increased with dose of oestrogen. The safest prescription remains one containing levonorgestrel with a low dose of oestrogen.

Participants and setting

Premenopausal women aged <50 years who were not pregnant, not within four weeks postpartum, and not using a hormonal intrauterine device or depot contraceptive were eligible.

Design, size, and duration

Analyses were performed on data from the MEGA study, a large, population based, case-control study on risk factors for venous thrombosis. From the total MEGA study, 1524 patients and 1760 controls were included.

Primary outcome, risks, exposures

The primary outcome was deep venous thrombosis of the leg or arm or pulmonary embolism. Exposure was the use of different types of oral contraceptives—different doses of oestrogen and chemically different types of progestogens—in the previous year. Information on contraceptive use was obtained using a standardised questionnaire on risk factors for venous thrombosis.

Main results and the role of chance

Currently available oral contraceptives were associated with a fivefold increased risk of venous thrombosis or pulmonary embolism (odds ratio 5.0, 95% CI 4.2 to 5.8). The risk clearly differed by type of progestogen (table) and increased with increasing dose of oestrogen.

The risk of thrombosis was higher for deep venous thrombosis of the leg (odds ratio 6.6, 5.4 to 8.0) than for pulmonary embolism (odds ratio 3.9, 3.2 to 4.8) and was highest in the first year of use, with a peak in the first three months (odds ratio 12.6, 7.1 to 22.4).

Bias, confounding, and other reasons for caution

Recall bias may have occurred in our study. However, the short time between the thrombotic event and completing the questionnaire and the fact that the questionnaire was sent to the participants' homes, where the package of the oral contraceptive was readily available, makes such bias unlikely.

Generalisability to other populations

These results apply to women in industrialised countries where a range of oral contraceptives is available.

Study funding/potential competing interests

This research was supported by the Netherlands Heart Foundation, Dutch Cancer Foundation, and Netherlands Organisation for Scientific Research. We have no competing interests.

THROMBOSIS RISK ASSOCIATED WITH COMBINED ORAL CONTRACEPTIVES BY TYPE OF PROGESTOGEN

Progestogen	Odds ratio (95% CI)*
Levonorgestrel	3.6 (2.9 to 4.6)
Gestodene	5.6 (3.7 to 8.4)
Desogestrel	7.3 (5.3 to 10.0)
Lynestrenol	5.6 (3.0 to 10.2)
Norethisterone	3.9 (1.4 to 10.6)
Cyproterone acetate	6.8 (4.7 to 10.0)
Norgestimate	5.9 (1.7 to 21.0)
Drospirenone	6.3 (2.9 to 13.7)
No oral contraceptive (reference)	1

*Odds ratio adjusted for age and period of inclusion (categorical; divide per 6 calendar months)
If more than one dose of oestrogen was available, analysis was restricted to the preparation with most commonly used dose of oestrogen (usually 30 µg)

BMJ pico: advice to authors

The full text of all accepted *BMJ* research articles is published online in full, with open access and no word limit, on bmj.com as soon as it is 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*. For more details, see http://tinyurl.com/kp5c7o/.

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 publication of research on bmj.com is definitive, rather than 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.

This is a summary of a paper that was published on bmj.com as *BMJ* 2009;339:b2921

pico

Adverse cardiovascular events during treatment with pioglitazone and rosiglitazone: population based cohort study

David N Juurlink,[1 2 3 4 5] Tara Gomes,[5] Lorraine L Lipscombe,[5 6] Peter C Austin,[4 5 7] Janet E Hux,[1 2 4 5] Muhammad M Mamdani[2 4 5 8]

EDITORIAL by de Vries and Russell-Jones

[1]Division of Clinical Pharmacology and Toxicology, Department of Medicine, Sunnybrook Health Sciences Centre, 2075 Bayview Avenue, Toronto, ON, Canada M4N 3M5
[2]Department of Medicine, University of Toronto
[3]Department of Pediatrics, University of Toronto
[4]Department of Health Policy, Management, and Evaluation, University of Toronto
[5]Institute for Clinical Evaluative Sciences, Toronto
[6]Women's College Hospital, Toronto
[7]Department of Public Health Sciences, University of Toronto
[8]Applied Health Research Centre, Li Ka Shing Knowledge Institute, St Michael's Hospital, Toronto
Correspondence to: D Juurlink dnj@ices.on.ca

Cite this as: *BMJ* 2009;339:b2942
doi: 10.1136/bmj.b2942

This is a summary of a paper that was published on bmj.com as *BMJ* 2009;339:b2942

STUDY QUESTION In older patients with diabetes, are rosiglitazone and pioglitazone associated with different risks of heart failure, myocardial infarction, and death?

SUMMARY ANSWER Patients treated with pioglitazone were at lower risk of admission to hospital for heart failure and death from any cause compared with those treated with rosiglitazone, but the risk of myocardial infarction did not differ significantly between the two groups.

Participants and setting
Residents of Ontario, Canada, aged 66 years and older were included in the study.

Design, size, and duration
This was a retrospective cohort study of 22 785 patients treated with rosiglitazone and 16 951 highly comparable patients treated with pioglitazone between 1 April 2002 and 31 March 2008. The analysis included extensive adjustment for a variety of clinical and demographic characteristics.

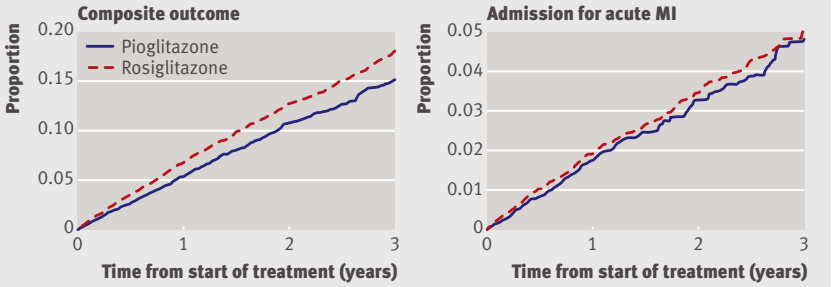
Main results and the role of chance
In total, 1563 (6.9%) patients receiving rosiglitazone and 895 (5.3%) receiving pioglitazone reached the primary composite outcome of death or admission to hospital for either acute myocardial infarction or congestive heart failure. The adjusted hazard ratio was 0.83 (95% confidence interval 0.76 to 0.90). Compared with rosiglitazone, treatment with pioglitazone was associated with a lower risk of heart failure (adjusted hazard ratio 0.77, 0.60 to 0.87) and death (0.86, 0.75 to 0.98) but no significant difference in the risk of acute myocardial infarction (0.95, 0.81 to 1.11). The results suggest that one additional hospital admission for heart failure each year would be expected for every 120 patients treated with rosiglitazone rather than pioglitazone, and one additional death would be expected for every 269 patients treated with rosiglitazone rather than pioglitazone.

Bias, confounding, and other reasons for caution
Patients who received rosiglitazone had a marginally higher burden of cardiovascular disease at baseline compared with those who received pioglitazone. However, the risk of admission to hospital for myocardial infarction was similar in both groups, suggesting that the main results are not explained by baseline differences in cardiac risk.

Generalisability to other populations
Whether the observations can be generalised to younger patients with diabetes is not known.

Study funding/potential competing interests
The study was supported by a grant from the Ontario Ministry of Health and Long Term Care. We have no competing interests.

SURVIVAL CURVES FOR COMPOSITE OUTCOME* AND HOSPITAL ADMISSION FOR ACUTE MYOCARDIAL INFARCTION FROM START OF TREATMENT



*Composite of death or hospital admission for acute myocardial infarction or heart failure

What types of article does the *BMJ* consider?

We are delighted to receive articles for publication—from doctors and others—on the clinical, scientific, social, political, and economic factors affecting health. We give priority to articles that will help doctors to make better decisions. Please see our advice to authors at http://resources.bmj.com/bmj/authors, and if you would like to submit an article do so via our online editorial office at http://submit.bmj.com.

All original research articles are submitted, although we may invite submission (without promising acceptance) if we come across research being presented at conferences, if we see it in abstract form, or if the authors make an inquiry about the suitability of their work before submission.

We are also pleased to consider submitted articles for sections which carry a mix of commissioned and submitted articles—editorials, analysis, clinical review, practice, fillers, and Career Focus. Please follow the specific advice on each of these article types (see http://resources.bmj.com/bmj/authors/types-of-article) before submitting your article. Some types of article—news, features, observations, head to head, views and reviews—are commissioned by the editors.

Cite this as: *BMJ* 2009;339:b3112