RESEARCH

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

THIS WEEK'S RESEARCH QUESTIONS

459 In patients with advanced incurable cancer, can clinical and laboratory data predict survival better than clinicians' subjective estimates?

460 What are the benefits of a language promotion programme for toddlers who are slow to talk?

461 Are socioeconomic status and sex still associated with median age at death in individuals with cystic fibrosis?

462 Do systematic reviews and meta-analyses of studies testing the accuracy of screening tools for depression evaluate bias owing to inclusion of patients with known depression?

463 Do adjusted indirect comparisons of competing healthcare interventions have similar results to head to head comparisons?

Diagnostic accuracy of screening tools for depression

This week's contribution to the *BMJ*'s Research Methods and Reporting section (http://bit.ly/dW6hGA) looks at studies of diagnostic accuracy (p 464). These studies aim to compare the results of a diagnostic test or model with results of a reference standard in the same patients, yielding measures that include predictive values, post-test probabilities, ROC (receiver operating characteristics) curves, sensitivity, specificity, likelihood ratios, and odds ratios. Joris de Groot and colleagues use real examples to show how such studies can sometimes lead to biased and exaggerated estimates of diagnostic accuracy and, in turn, to "inefficiencies in diagnostic testing in practice, unnecessary costs, and physicians making incorrect treatment decisions." One cause of bias is failure to adjust analyses to allow for the inclusion in such studies of patients who already have or are very likely to have, the disease.

Brett D Thombs and colleagues take this point further in their meta-review, focusing on diagnostic tools used to screen for depression (p 462). They scrutinised 17 eligible systematic reviews and meta-analyses, covering 197 primary publications, and found none commenting on possible spectrum bias from inclusion of patients who already had depression or were being treated for it. They are concerned that only eight of the primary studies specifically excluded such patients—it's worth noting, though, that de Groot and colleagues counsel against such exclusion and argue instead for corrected analyses to account for the bias. Given these findings and the authors' comment that "no clinical trial has found better depression outcomes for screened versus unscreened patients when the same treatment and care resources are potentially available to both groups," should we be more worried about the ubiquity of depression screening in so many healthcare settings.

Predicting death from cancer: human versus tool

How long have I got left? It's a reasonable question for a patient with terminal cancer to ask. And it's important too, because patients, relatives, and doctors need to make plans. But how comfortably, or—more importantly—accurately, do doctors answer? Apparently, not well, and usually optimistically.

In an editorial Paul Glare writes that doctors tend to dodge questions of prognosis (p 429). Perhaps we can be forgiven; he goes on to say that we are rarely trained to do it, and there are no tools to predict death in widespread use. Those that do exist have methodological limitations, or have not been validated.

Gwilliam and colleagues used death markers identified in previous studies to create a tool that could predict the time to death as "days," "weeks," or "months," more accurately than professional opinion (p 459). Their population was just over 1000 patients who were new to 18 palliative care services in England. The tool can function with and without blood results, and it is accurate about 60% of the time. But as the authors and Glare write, there are cautions—for example, it will need to be validated, and it is not very user friendly, although an app is in the pipeline.

Health inequalities with cystic fibrosis

ISM/SPL

Survival among people with cystic fibrosis has improved dramatically over the past 50 years, with the median age at death in the UK rising from 6 months in 1959 to 27 years in 2008. Two long established factors associated with early death are low socioeconomic status and female sex. Helen Barr and colleagues used death registration data in England and Wales from 1959 to 2008 to investigate whether these socioeconomic and sexual inequalities in mortality from cystic fibrosis have weakened with the overall improvements in prognosis (p 461). Disappointingly, they found no substantial narrowing in the inequalities over the 50 year study period.

The cause of the sex gap in survival remains unclear, although David Taylor-Robinson and Michael Schechter suggest in their linked editorial that socially determined gender roles are probably as much to blame as biologically determined sex characteristics (p 431). The socioeconomic gradient in survival may be easier to understand.

LATEST RESEARCH: For this and other new research articles see www.bmj.com/research

Chocolate consumption and cardiometabolic disorders A meta-analysis by Adriana Buitrago-Lopez and colleagues suggests that increased intake of chocolate is associated with a substantial reduction in the risk of cardiometabolic disorders. The association was significant for any cardiovascular disease (37% reduction), diabetes (31%), and stroke (29%), but not for heart failure. However, the authors found no randomised trials in their systematic review—only limited observational studies in selected populations—so much more evidence is needed to confirm whether eating chocolate is good for you (doi:10.1136/bmj.d4488).

Development of Prognosis in Palliative care Study (PiPS) predictor models to improve prognostication in advanced cancer: prospective cohort study

Bridget Gwilliam,[1] Vaughan Keeley,[2] Chris Todd,[3] Matthew Gittins,[4] Chris Roberts,[4] Laura Kelly,[5] Stephen Barclay,[6] Patrick C Stone[1]

EDITORIAL by Glare

[1]Division of Population, Health Sciences and Education, St George's University of London, London SW17 ORE, UK
[2]Royal Derby Hospital, Derby, UK
[3]School of Nursing, Midwifery and Social Work, University of Manchester, Manchester, UK
[4]Health Sciences, School of Community Based Medicine, University of Manchester
[5]Macmillan Consultant in Palliative Care Team, East Surrey Hospital, Surrey and Sussex Healthcare NHS Trust, Redhill, Surrey, UK
[6]General Practice and Primary Care Research Unit, Department of Public Health and Primary Care, Institute of Public Health, Cambridge, UK
Correspondence to: P C Stone
pstone@sgul.ac.uk

Cite this as: *BMJ* 2011;343:d4920
doi: 10.1136/bmj.d4920

This is a summary of a paper that was published on bmj.com as *BMJ* 2011;343:d4920

STUDY QUESTION

In patients with advanced incurable cancer, can clinical and laboratory data predict survival better than clinicians' subjective estimates?

SUMMARY ANSWER

A prediction model was significantly more accurate than either a doctor or a nurse alone but was not significantly better than an agreed multi-professional estimate of survival.

WHAT IS KNOWN AND WHAT THIS PAPER ADDS

Prognostic information is valued by patients with advanced cancer and their carers and healthcare professionals, but clinicians' predictions are unreliable, over-optimistic, and subjective. Two prognostic scores have been created that are independent of clinicians' subjective estimates of survival; one of the scores (which requires a blood test) is significantly better than an individual doctor's or nurse's prediction, but neither score is significantly more accurate than an agreed multi-professional estimate of survival.

Participants and setting

We studied adult men and women with incurable cancer who were no longer being actively treated. We recruited patients from new referrals to 18 palliative care services (including hospices, community, and hospital support teams) between March 2006 and August 2009.

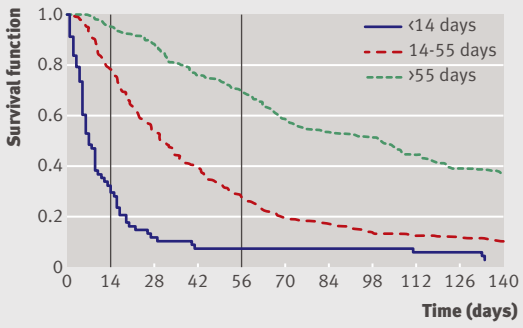
Design, size, and duration

This was a prospective, multicentre, observational, cohort study including 1018 patients (775 competent and 243 incompetent). All participants had demographic, disease, and other clinical variables recorded. Competent patients provided blood for laboratory analysis. We obtained a clinical estimate of survival independently from both a doctor and a nurse. If these estimates differed, we obtained an agreed multi-professional estimate. We asked clinicians to estimate whether the patient was likely to live for "days" (<14 days), "weeks" (14-55 days), or "months" (>55 days). We followed up patients for at least three months.

Main results and the role of chance

On multivariate analysis, 11 variables (pulse rate, general health status, mental test score, performance status, presence of anorexia, presence of any site of metastatic disease, presence of liver metastases, C reactive protein, white blood count, platelet count, and urea) independently predicted both two week and two month survival.

KAPLAN-MEIER SURVIVAL CURVES FOR PiPS-B MODEL (WHICH INCLUDES BLOOD VARIABLES)



Graph shows survival curves for three prognostic groups identified by PiPS-B scores. Vertical lines indicate survival at specific "cut-off points" of 14 and 56 days

Four variables had prognostic significance only for two week survival (dyspnoea, dysphagia, bone metastases, and alanine transaminase). Eight variables had prognostic significance only for two month survival (primary breast cancer, male genital cancer, tiredness, weight loss, lymphocyte count, neutrophil count, alkaline phosphatase, and albumin). We created separate prognostic models for patients without (PiPS-A) or with (PiPS-B) blood results. The area under the curve for all models varied between 0.79 and 0.86. Agreement between actual survival and PiPS predictions was 57.3% (after correction for over-optimism). The median survival across the PiPS-A categories was 5, 33, and 92 days, and across PiPS-B categories it was 7, 32, and 100.5 days. Both models performed as well as, or better than, clinicians' estimates of survival.

Bias, confounding, and other reasons for caution

Owing to the nature of the study population, we could approach only 34% (2401/7017) of eligible patients, of whom 42% (1018/2401) agreed to participate. The study sample may not be representative of the wider palliative care population. Although we did cross validation of the models (using a bootstrap technique), further external validation is needed.

Generalisability to other populations

We developed the models in a sample of patients with end stage, incurable cancer. They should not be used in other patients with advanced cancer (such as those still receiving palliative chemotherapy or those in whom such treatment is planned).

Study funding/potential competing interests

This study was funded by Cancer Research UK.

bmj.com

- News: Per patient funding would end inequalities in palliative care provision, says review (*BMJ* 2011;343:d4242)
- News: Palliative care is "neglected" worldwide, report says (*BMJ* 2011;342:d3510)
- News: Italy sets up national palliative care service (*BMJ* 2010;340:c1481)

Outcomes of population based language promotion for slow to talk toddlers at ages 2 and 3 years: Let's Learn Language cluster randomised controlled trial

Melissa Wake,[1] Sherryn Tobin,[1] Luigi Girolametto,[2] Obioha C Ukoumunne,[3] Lisa Gold,[4] Penny Levickis,[1] Jane Sheehan,[1] Sharon Goldfeld,[1] Sheena Reilly[1]

EDITORIAL by Boyle

[1]Royal Children's Hospital, Murdoch Childrens Research Institute and University of Melbourne, Parkville, VIC 3052, Australia
[2]Department of Speech-Language Pathology, University of Toronto, Toronto, ON, Canada, M5G 1V7
[3]PenCLAHRC, Peninsula College of Medicine and Dentistry, University of Exeter, Exeter, UK
[4]Deakin Health Economics, Deakin University, Burwood, VIC 3125, Australia
Correspondence to: M Wake, Centre for Community Child Health, Royal Children's Hospital, Flemington Road, Parkville, VIC 3052, Australia
melissa.wake@rch.org.au

Cite this as: *BMJ* 2011;343:d4741
doi: 10.1136/bmj.d4741

This is a summary of a paper that was published on bmj.com as *BMJ* 2011;343:d4741

STUDY QUESTION

What, if any, are the benefits of a six session parent-toddler language promotion programme delivered to toddlers with low spoken vocabulary on screening at 18 months in universal services?

SUMMARY ANSWER

Although parents reported improved communication, it did not improve language or behaviour either immediately or at age 3 years.

WHAT IS KNOWN AND WHAT THIS PAPER ADDS

Clinical intervention to tackle preschool language delay typically starts late and has limited effect; selective prevention could start earlier and reach more children, but its benefits are uncertain. Population based screening for slow to talk toddlers followed by a language promotion programme was feasible and acceptable but did not improve language or behaviour immediately or at age 3 years.

Design

This trial was nested within a population based survey. An independent statistician stratified maternal and child health centres (clusters) and randomly allocated the clusters to intervention or control ("usual care"). Randomisation and measurement of outcomes were blinded. The intervention was a modified "You Make the Difference" parent-toddler language promotion programme, delivered over six weeks in weekly sessions each lasting two hours.

Participants and setting

We recruited 1217 parents consecutively attending 12 month well child checks over six months; 1138 (93.5%) completed an expressive vocabulary checklist (UK Sure Start expressive language screen) at 18 months. The 301 (26.4%) children scoring at or below the 20th centile entered the trial (158 intervention, 143 control).

Primary outcome(s)

The primary outcome at ages 2 and 3 years was the child's language score, measured by the Preschool Language Scale-4 Expressive Communication and Auditory Comprehension scales. Other measures were vocabulary checklist raw scores, Expressive Vocabulary Test (age 3 years only) and behaviour scores.

Main results and the role of chance

Retention was 94% (148/158) for intervention children and 96% (137/143) for controls at 2 years and 89% (140/158; 127/143) for both groups at 3 years. Of intervention parents, 115 (73%) attended at least one session (mean 4.5 sessions). In both unadjusted and adjusted analyses, the intervention and control arms had similar means for all outcomes at ages 2 and 3 years (table). In the intervention group, 86/100 parents said that the programme had benefited how they communicated with their child, and 72/100 reported that it had benefited their child's communication.

Harms

We did not identify any harms of the intervention.

Bias, confounding, and other reasons for caution

Children entered the trial with no or very few spoken words at age 18 months. This is an accurate marker of spoken language at this age but was a poor predictor of subsequent language delay, as language scores in both groups were close to population means by age 3. If screening for very early language delay is to be effective, population based studies must first develop markers with adequate prognostic sensitivity and specificity.

Generalisability to other populations

Our results are likely to be applicable to many Western populations.

Study funding/potential competing interests

The trial and several authors were funded by the Australian National Health and Medical Research Council. Research at the Murdoch Childrens Research Institute is supported by the Victorian government's Operational Infrastructure Support Program.

Trial registration number

Current Controlled Trials ISRCTN20953675.

EFFECTS AT AGES 2 AND 3 YEARS OF SCREENING AND INTERVENTION FOR SLOW TO TALK TODDLERS

Outcomes	Mean (SD) for trial arms		Adjusted difference (I–C)	
	Intervention (I)	Control (C)	Mean (95% CI)	P value
2 years				
MCDI vocabulary raw score	34.5 (22.4)	34.4 (23.4)	2.1 (−3.0 to 7.2)	0.42
PLS expressive standard score	90.4 (12.9)	90.1 (11.2)	1.2 (−1.6 to 4.0)	0.41
PLS comprehension standard score	88.8 (15.2)	88.9 (14.3)	1.4 (−2.2 to 5.0)	0.44
CBCL externalising behaviour raw score	12.3 (7.8)	12.0 (7.3)	−0.3 (−1.6 to 1.1)	0.71
CBCL internalising behaviour raw score	5.7 (5.2)	5.4 (3.9)	0.1 (−0.9 to 1.1)	0.78
3 years				
MCDI vocabulary raw score	53.5 (27.9)	51.4 (25.2)	4.1 (−2.3 to 10.6)	0.21
EVT expressive vocabulary standard score	100.5 (15.6)	101.6 (12.0)	−0.5 (−4.4 to 3.4)	0.80
PLS expressive standard score	97.7 (16.1)	100.7 (14.0)	−2.4 (−6.2 to 1.4)	0.21
PLS comprehension standard score	96.1 (17.5)	97.0 (14.7)	−0.3 (−4.2 to 3.7)	0.90
CBCL externalising behaviour raw score	10.8 (7.9)	10.7 (6.9)	−0.1 (−1.6 to 1.4)	0.86
CBCL internalising behaviour raw score	6.3 (5.7)	6.0 (4.6)	−0.1 (−1.3 to 1.2)	0.92

CBCL=Child Behavior Checklist; EVT=Expressive Vocabulary Test; MCDI=MacArthur-Bates Communicative Development Inventory; PLS=Preschool Language Scale.
Sample sizes 119-125 (I) and 121-122 (C) at 2 years; 89-116 (I) and 91-112 (C) at 3 years.

Association between socioeconomic status, sex, and age at death from cystic fibrosis in England and Wales (1959 to 2008): cross sectional study

Helen L Barr,[1] John Britton,[2] Alan R Smyth,[3] Andrew W Fogarty[2]

EDITORIAL by Taylor-Robertson and Schechter

[1]Nottingham Respiratory Biomedical Research Unit, Department of Respiratory Medicine and Department of Epidemiology and Public Health, City Hospital Campus, Nottingham NG5 1PB, UK
[2]Nottingham Respiratory Biomedical Research Unit, Department of Epidemiology and Public Health
[3]Nottingham Respiratory Biomedical Research Unit, School of Clinical Sciences, Nottingham University Hospitals NHS Trust, Queen's Medical Centre, Nottingham
Correspondence to: H L Barr
helen.barr@nottingham.ac.uk

Cite this as: *BMJ* 2011;343:d4662
doi: 10.1136/bmj.d4662

This is a summary of a paper that was published on bmj.com as *BMJ* 2011;343:d4662

bmj.com
Domhnall MacAuley: The silence next door
http://bit.ly/pxb3x5

STUDY QUESTION

In the 21st century, are socioeconomic status and sex associated with median age at death in individuals with a diagnosis of cystic fibrosis?

SUMMARY ANSWER

Socioeconomic status and sex remain strong determinants of median age of death from cystic fibrosis in England and Wales, and the magnitude of these associations does not appear to have changed over the past 50 years.

WHAT IS KNOWN AND WHAT THIS PAPER ADDS

Survival in individuals with a diagnosis of cystic fibrosis has improved greatly over the past 50 years, with the median age of death increasing from 6 months to 27 years. Despite overall improved survival in the 21st century, females and socioeconomically disadvantaged individuals continue to be more likely to die below the median age at death than males and socioeconomically advantaged individuals.

Participants and setting

All registered deaths with a diagnosis of cystic fibrosis in England and Wales, from 1959 to 2008.

Design

Series of annual cross sectional studies of mortality data.

Primary outcomes

Mutually adjusted odds ratios for death above the annual median age at death by socioeconomic status and sex, calculated with logistical regression.

Main results

Between 1959 and 2008, 6750 deaths recorded in England and Wales were attributed to cystic fibrosis. Males were more likely to die above the annual median age at death than females (from 1959 to 1999, adjusted odds ratio for socioeconomic status 1.28, 95% confidence intervals 1.13 to 1.45; from 2000 to 2008, 1.57, 1.18 to 2.08). Individuals in the highest socioeconomic group were also more likely to die above the median age of death than those in manual occupations (from 1959 to 2000, adjusted odds ratio for sex 2.50, 2.16 to 2.90; from 2001 to 2008, 1.89, 1.20 to 2.97) (table).

Bias, confounding, and other reasons for caution

Socioeconomic status classification changed during the study period, making longitudinal trend comparisons challenging. Overall, 1825 (32%) of 5759 individuals with available socioeconomic data were coded as unclassified, which included a broad cross section of society and which could introduce bias to our results. In addition, the association between low socioeconomic status and increased mortality may be confounded by reverse causation.

Generalisability to other populations

National differences between healthcare systems could limit the generalisability of these data. However, other studies from the United States report similar socioeconomic and sex health gaps in individuals with a diagnosis of cystic fibrosis.

Study funding/potential competing interests

This research was funded by the Medical Research Council and the University of Nottingham.

SOCIOECONOMIC STATUS AND RISK OF DEATH FROM CYSTIC FIBROSIS AT AGE GREATER THAN MEDIAN AGE AT DEATH

		Risk of death at age greater than median age at death, by socioeconomic status				
Study year	Total no of deaths	Manual	Non-manual*	Routine and manual	Intermediate*	Professional and managerial*
1959-63†	969	1.00	1.90 (1.35 to 2.67)	–	–	–
1970-79	1425	1.00	2.05 (1.57 to 2.69)	–	–	–
1980-89	1215	1.00	3.04 (2.26 to 4.10)	–	–	–
1990-2000	1277	1.00	2.06 (1.52 to 2.80)	–	–	–
2001-08‡	873	–	–	1.00	1.07 (0.66 to 1.72)	1.89 (1.20 to 2.97)

Data are odds ratio or adjusted odds ratio (95% confidence intervals) unless stated otherwise.
*Odds ratios adjusted for sex.
†Socioeconomic status not classified in 1964-69.
‡Socioeconomic status coding changed after 2000.

Risk of bias from inclusion of patients who already have diagnosis of or are undergoing treatment for depression in diagnostic accuracy studies of screening tools for depression: systematic review

Brett D Thombs,[1] Erin Arthurs,[1] Ghassan El-Baalbaki,[1] Anna Meijer,[2] Roy C Ziegelstein,[3] Russell J Steele[4]

[1]Lady Davis Institute for Medical Research, Jewish General Hospital and McGill University, Montreal, Quebec, Canada H3T 1E4
[2]Interdisciplinary Centre for Psychiatric Epidemiology, University Medical Centre Groningen, University of Groningen, 9713 GZ, Netherlands
[3]Johns Hopkins University School of Medicine, Baltimore, Maryland 21224, USA
[4]Department of Mathematics and Statistics, McGill University, 805 Sherbrooke Ouest, Montreal, Quebec, H3A 2K6
Correspondence to: B D Thombs
brett.thombs@mcgill.ca

Cite this as: *BMJ* 2011;343:d4825
doi: 10.1136/bmj.d4825

This is a summary of a paper that was published on bmj.com as *BMJ* 2011;343:d4825

STUDY QUESTIONS

Do systematic reviews and meta-analyses of the accuracy of these screening tools evaluate possible bias that results from the inclusion in original studies of patients who are already known to have depression?

SUMMARY ANSWER

Fewer than 5% of studies on the diagnostic accuracy of screening tools for depression appropriately excluded patients who already had a diagnosis of or were being treated for depression. No systematic reviews or meta-analyses commented on possible bias from the inclusion of such patients in original studies, even though many reviews used quality assessment tools with items designed to rate risk of bias from composition of the sample of patients.

WHAT IS KNOWN AND WHAT THIS PAPER ADDS

The inclusion of patients who are known to have depression in studies that examine the accuracy of screening tools for depression almost certainly results in widespread overestimates of accuracy of screening and case yield compared with what would be achieved if these screening tools were used as designed—namely, to identify new cases.

Selection criteria for studies

We searched the Medline, PsycINFO, CINAHL, Embase, ISI, SCOPUS, and Cochrane databases from 1 January 2005 to 29 October 2009 to identify systematic reviews and meta-analyses of the diagnostic accuracy of screening tools for depression. Systematic reviews and meta-analyses in any language were eligible if they reviewed the accuracy of self reported depression screening tools compared with a diagnosis of depression. We restricted the search to this period to obtain recent systematic reviews and meta-analyses that reflect relatively current practices.

Primary outcomes

For each systematic review or meta-analysis, we noted whether or not they mentioned possible bias in original studies from inclusion of patients who already had a diagnosis of or were being treated for depression. For original diagnostic accuracy studies included in systematic reviews and meta-analyses, we noted whether or not such patients were excluded.

Main results and role of chance

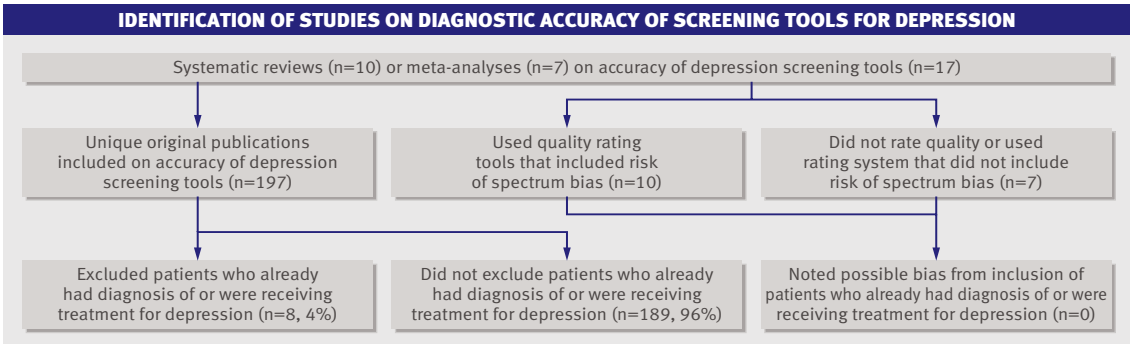
There were 17 eligible systematic reviews and meta-analyses that included a total of 197 unique publications. Only eight of 197 unique publications (4%) specifically excluded patients who already had a diagnosis of or were being treated for depression. None of the 17 systematic reviews or meta-analyses commented on possible spectrum bias from inclusion in studies of such patients, even though 10 of 17 systematic reviews or meta-analyses used quality assessment methods that included an assessment of possible bias from the composition of the sample of patients.

Bias, confounding, and other reasons for caution

We searched for systematic reviews and meta-analyses, rather than for original studies, and there are probably many original studies on the diagnostic accuracy of depression screening tools that were not included. This is because our purpose was both to assess whether original studies appropriately excluded patients who already had a diagnosis or were being treated and to determine whether systematic reviews and meta-analyses reflected potential bias from the failure to do this, which required a review of reviews.

Study funding/potential competing interests

BDT is supported by a New Investigator Award from the Canadian Institutes of Health Research and an Établissement de Jeunes Chercheurs award from the Fonds de la Recherche en Santé Québec. RCZ is supported by the National Center for Complementary and Alternative Medicine (grant No R24AT004641) and the Miller Family Scholar Program of the Johns Hopkins Center for Innovative Medicine.



Inconsistency between direct and indirect comparisons of competing interventions: meta-epidemiological study

Fujian Song,[1] Tengbin Xiong,[12] Sheetal Parekh-Bhurke,[13] Yoon K Loke,[1] Alex J Sutton,[4] Alison J Eastwood,[5] Richard Holland,[1] Yen-Fu Chen,[6] Anne-Marie Glenny,[7] Jonathan J Deeks,[6] Doug G Altman[8]

[1]Norwich Medical School, Faculty of Medicine and Health Science, University of East Anglia, Norwich NR4 7TJ, UK
[2]Department of Oncology, University of Cambridge, Cambridge, UK
[3]NIHR Trials and Studies Coordinating Centre, University of Southampton, Southampton, UK
[4]Department of Health Science, University of Leicester, Leicester, UK
[5]Centre for Reviews and Dissemination, University of York, York, UK
[6]Public Health, Epidemiology and Biostatistics, University of Birmingham, Birmingham, UK
[7]School of Dentistry, University of Manchester, Manchester, UK
[8]Centre for Statistics in Medicine, University of Oxford, Oxford, UK

Correspondence to: F Song
fujian.song@uea.ac.uk

Cite this as: *BMJ* 2011;343:d4909
doi: 10.1136/bmj.d4909

This is a summary of a paper that was published on bmj.com as *BMJ* 2011;343:d4909

STUDY QUESTION

Are the results of adjusted indirect comparison consistent with the results of head to head comparison of competing healthcare interventions?

SUMMARY ANSWER

Based on data from a sample of meta-analyses, significant inconsistency between direct and indirect comparisons may be more prevalent than previously observed.

WHAT IS KNOWN AND WHAT THIS PAPER ADDS

Limited empirical evidence indicated that differences between direct and adjusted indirect comparisons were only occasionally statistically significant. The risk of statistically significant inconsistency is associated with fewer trials included in analyses, subjectively assessed outcomes, and statistically significant estimates of treatment effects by either direct or indirect comparisons.

Selection criteria for studies

We searched the Cochrane Database of Systematic Reviews and PubMed to identify systematic reviews that provided sufficient data for both direct comparison of two interventions and independent indirect comparisons on the basis of a common comparator, and in which the odds ratio could be used as the outcome statistic.

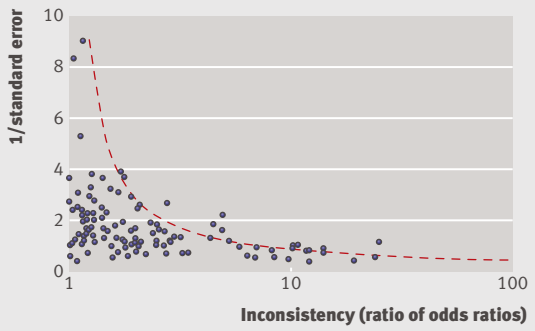
Primary outcome

The indirect comparison involved a comparison of two interventions by using data from separate trials, on the basis of a common comparator. We measured inconsistency by the difference in the log odds ratio between the direct and indirect methods. The inconsistency between direct and indirect estimates can also be expressed as a ratio of odds ratios by an antilog transformation.

Main results and role of chance

The study included 112 independent trial networks (including 1552 trials with 478 775 patients in total) that allowed both direct and indirect comparison of two interventions. Indirect comparison had already been explicitly done in only 13 of the 85 Cochrane reviews included. The funnel plot shows the effect of random error, in which the points of inconsistency (ratio of odds ratios) with less precise estimates were much more widely scattered. The prevalence of statistically significant inconsistency between the direct and indirect comparison was 14% (95% confidence interval 9% to 22%), which is higher than that expected (5%) and higher than that found in a previous study (7%). The statistically significant inconsistency was associated with fewer trials, subjectively assessed outcomes, and statistically significant treatment effects in either direct or indirect comparisons. Owing to considerable inconsistency, many (14/39) of the statistically significant effects by direct comparison became non-significant when we combined the direct and indirect estimates.

ONE SIDED FUNNEL PLOT OF (ABSOLUTE) ESTIMATED INCONSISTENCY BETWEEN DIRECT AND INDIRECT COMPARISON



Dotted line indicates statistical significance at P=0.05

Bias, confounding, and other reasons for caution

An indirect comparison had not been done in most (72/85) Cochrane reviews that provided sufficient data for both direct and indirect comparisons. If authors of systematic reviews did not do or report the indirect comparison because of perceived inconsistency, we may have overestimated the prevalence of statistical inconsistency in our study. The number of included trials and patients within individual networks was rather small in most cases, and the available evidence for exploratory subgroup analyses was limited. Wider 95% confidence intervals of inconsistencies indicated the possibility of insufficient statistical power in many of the included cases. Therefore, statistically non-significant inconsistency does not necessarily imply clinical consistency.

Study funding/potential competing interests

The study was funded by the UK Medical Research Council (G0701607).