

Research article

Open Access

## Evaluating heterogeneity in cumulative meta-analyses

Elmer V Villanueva\*<sup>1,2</sup> and Silva Zavarsek<sup>3</sup>

Address: <sup>1</sup>Monash Institute of Health Services Research, Monash University, Clayton, Victoria, Australia, <sup>2</sup>National Breast Cancer Centre, Camperdown, New South Wales, Australia and <sup>3</sup>Centre for Clinical Effectiveness, Monash Institute of Health Services Research, Monash University, Clayton, Victoria, Australia

Email: Elmer V Villanueva\* - [Elmer.Villanueva@nbcc.org.au](mailto:Elmer.Villanueva@nbcc.org.au); Silva Zavarsek - [Silva.Zavarsek@med.monash.edu.au](mailto:Silva.Zavarsek@med.monash.edu.au)

\* Corresponding author

Published: 13 July 2004

Received: 24 March 2004

*BMC Medical Research Methodology* 2004, **4**:18 doi:10.1186/1471-2288-4-18

Accepted: 13 July 2004

This article is available from: <http://www.biomedcentral.com/1471-2288/4/18>

© 2004 Villanueva and Zavarsek; licensee BioMed Central Ltd. This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's original URL.

### Abstract

**Background:** Recently developed measures such as  $I^2$  and  $H$  allow the evaluation of the impact of heterogeneity in conventional meta-analyses. There has been no examination of the development of heterogeneity in the context of a cumulative meta-analysis.

**Methods:** Cumulative meta-analyses of five smoking cessation interventions (clonidine, nicotine replacement therapy using gum and patch, physician advice and acupuncture) were used to calculate  $I^2$  and  $H$ . These values were plotted by year of publication, control event rate and sample size to trace the development of heterogeneity over these covariates.

**Results:** The cumulative evaluation of heterogeneity varied according to the measure of heterogeneity used and the basis of cumulation. Plots produced from the calculations revealed areas of heterogeneity useful in the consideration of potential sources for further study.

**Conclusion:** The examination of heterogeneity in conjunction with summary effect estimates in a cumulative meta-analysis offered valuable insight into the evolution of variation. Such information is not available in the context of conventional meta-analysis and has the potential to lead to the development of a richer picture of the effectiveness of interventions.

### Background

As predicted by Mulrow [1,2] (among others), reports of meta-analyses – the suite of statistical techniques used to arrive at pooled estimates of effects across a series of studies (often but not always) during the course of a systematic review – have ballooned in parallel with the rapid and sustained pace at which information becomes available about the efficacy of interventions. Often, meta-analyses are conducted after a collection of studies have been identified; statistical pooling occurs at one point in time. The sequential pooling of the effect estimate in a "cumulative" manner as studies are published or according to other specific variables of interest (study quality or control event rate, for instance) were described and developed by Lau

and colleagues [3-5]. These developments have focused on describing the evolution of the point estimate and its confidence intervals.

Conventional meta-analyses are usually reported in conjunction with a test for heterogeneity. A popular statistic, Cochran's  $Q$ , is the sum of the squared differences between each study's effect estimate and the overall effect estimate, weighted for the information provided by the particular study [6,7]. Traditionally,  $Q$  has been used as a formal test of homogeneity as, under the null hypothesis, it follows a chi-squared distribution with degrees of freedom equal to the total number of studies less one. In spite of its problems [8],  $Q$  is widely used as a means of

determining whether statistically significant heterogeneity is present.

Moving forward from this binary consideration, Higgins and colleagues [9,10] proposed measures to quantify the impact (as opposed to the extent) of heterogeneity in meta-analyses. Two measures have particular intuitive appeal.  $H^2$ , the ratio of  $Q$  to its degrees of freedom, may be roughly interpreted as the ratio of confidence interval widths for single summary estimates from random effect and fixed effect meta-analyses.  $I^2$  describes the amount of heterogeneity among studies relative to the total variability among the effect estimates.

Cumulative meta-analytic techniques have revealed how estimates of effect evolve across time. How is the temporal progress of heterogeneity characterised? We are not aware of studies that have examined measures of heterogeneity in the context of a cumulative meta-analysis. To this end, we were interested in describing  $H^2$  and  $I^2$  in sequentially pooled effect estimates when cumulative meta-analysis is performed.

**Methods**

**Measures of heterogeneity**

Higgins and colleagues [9] derived three candidate measures of the extent of heterogeneity in a meta-analysis, two of which are considered in this paper. The measures were created to meet three specific criteria. First, the measure was to increase as the variance in the underlying treatment effects increases. Second, the measure was not to be affected by the scale of measurement or the type of outcome being considered. Lastly, the measure was not to depend on the number of studies.

$H^2$  represents the proportional surfeit of  $Q$  relative to its degrees of freedom (or the number of studies,  $k$ , less one),

$$H^2 = \frac{Q}{k-1}. \tag{1}$$

Given that  $E [Q] = k - 1$  in the absence of heterogeneity, a value of 1 for  $H^2$  indicates an absence of heterogeneity. Higgins and colleagues make several recommendations about the use of  $H^2$ , including the use of the square root of  $H^2$  (or  $H$ ) in the same way that discussions about the standard deviation are more familiar to clinicians than are discussions about the variance; the use of a test-based standard error for the natural logarithm of  $H$

$$SE[\ln(H)] = \begin{cases} \frac{1}{2} \frac{\ln(Q) - \ln(k-1)}{\sqrt{2Q} - \sqrt{2k-3}} & \text{if } Q > k; \\ \sqrt{\frac{1}{2(k-2)} \left(1 - \frac{1}{3(k-2)^2}\right)} & \text{if } Q \leq k \end{cases} \tag{2}$$

leading to the calculation of the  $100(1-2\alpha)\%$  confidence interval (i.e., the estimated range of values which has a probability  $100(1-2\alpha)\%$  of including the unknown population parameter) for  $H$  as

$$\exp\left(\ln(H) \pm Z_{\alpha/2} \times SE[\ln(H)]\right); \tag{3}$$

and taking the maximum out of  $H$  and 1 [9].

First examined by Takkouche, Cadarso-Suarez and Spiegelman [11] in the context of meta-analyses of observational studies,  $I^2$  represents the proportion of the total variability (i.e., the sum of between-study variance  $\tau^2$  and residual variability arising from sampling errors averaged across all studies  $\sigma^2$ ) that is explained by the variability in the underlying treatment effects

$$I^2 = \frac{\tau^2}{\tau^2 + \sigma^2}. \tag{4}$$

In the absence of heterogeneity,  $I^2 = 1$ . The two measures are related through

$$I^2 = \frac{H^2 - 1}{H^2} \tag{5}$$

and leads to the calculation of confidence intervals for  $I^2$  using equations 2 and 3.

**Motivating systematic reviews**

We chose four systematic reviews of randomised controlled trials relating to different interventions designed to promote smoking cessation: acupuncture [12], clonidine [13], nicotine replacement therapy (reported separately for gum and patch) [14], and physician advice [15]. The reviews were published in the third issue of 2003 of the Cochrane Database of Systematic Reviews, although the reviews themselves were last updated in 2001 [13,15] or 2002 [12,14]. As the focus of the present paper is heterogeneity as a methodological phenomenon, we invite readers to examine the latest version of the Cochrane Library for a current exposition of the effectiveness of these interventions.

**Clonidine: No heterogeneity in final analysis, presence of statistically significant effect**

Clonidine, traditionally used to lower blood pressure, has been used as a therapy for smoking cessation as it may reduce symptoms of withdrawal via its action on the central nervous system. The meta-analysis performed by Gourlay and colleagues analysed the effectiveness of clonidine therapy (oral or transdermal) versus placebo in smoking cessation [13]. While only one of the six studies included in the meta-analysis showed that clonidine was statistically significantly more effective than placebo, the

resulting meta-analysis showed that clonidine was more effective overall. The authors concluded that it was reasonable to consider clonidine as second-line therapy for smoking cessation [13]. The value of  $I^2$  in a meta-analysis containing all studies is zero.

**Nicotine replacement: Minimal heterogeneity in final analysis, presence of statistically significant effect**

Nicotine replacement therapy has been used to aid in smoking cessation by replacing nicotine from cigarettes, thereby reducing withdrawal symptoms. Different forms of nicotine replacement therapies exist and these include nicotine gum, transdermal patches, nasal spray, inhalers and tablets. The results of the meta-analysis performed by Silagy and colleagues indicated that all of the commercially available forms of nicotine replacement therapy were effective in promoting smoking cessation and that nicotine replacement was more effective than placebo in achieving smoking cessation independent of therapy duration and the level of additional support and advice [14]. In this paper, we focus on the effects of nicotine replacement in gum and patch form. The value of  $I^2$  in a meta-analysis containing all studies is 19.41% for gum and 28.57% for patch.

**Physician advice: Moderate heterogeneity in final analysis, presence of statistically significant effect**

Physicians may play an important role in facilitating smoking cessation by providing advice (ranging from brief to intensive) to patients on how their health can improve by quitting. Silagy and Stead [15] examined the effectiveness of advice from physicians in promoting smoking cessation. Results indicated that brief advice was more effective than no advice on smoking cessation. The value of  $I^2$  in a meta-analysis containing all studies is 35.51%.

**Acupuncture: Moderate heterogeneity in final analysis, no statistically significant effect**

Treatment with acupuncture and related therapies has been used to aid patients cease dependence on addictive drugs by reducing withdrawal symptoms. White and colleagues performed a meta-analysis to assess the effectiveness of acupuncture, acupressure, laser therapy and electrostimulation on smoking cessation compared with sham treatment, other interventions or no intervention [12]. The authors concluded that the evidence did not suggest that any of these techniques were more effective than placebo for smoking cessation. The value of  $I^2$  in a meta-analysis containing all studies is 49.98%.

**Cumulative meta-analysis**

Using each of the five smoking cessation interventions, we conducted a cumulative meta-analysis based on repeated pooling of individual studies according to publication

year, control group event rate, and study size. The fixed-effects model described by Mantel and Haenszel [16] was used to arrive at summary estimates and confidence intervals for the odds ratio (OR) of the primary intervention in the treatment group versus the control group. Measures of heterogeneity were calculated at each pooling event and plotted alongside the summary statistic in a forest plot. All calculations were performed on Stata 8.2 SE (StataCorp Inc, College Station, Texas, USA).

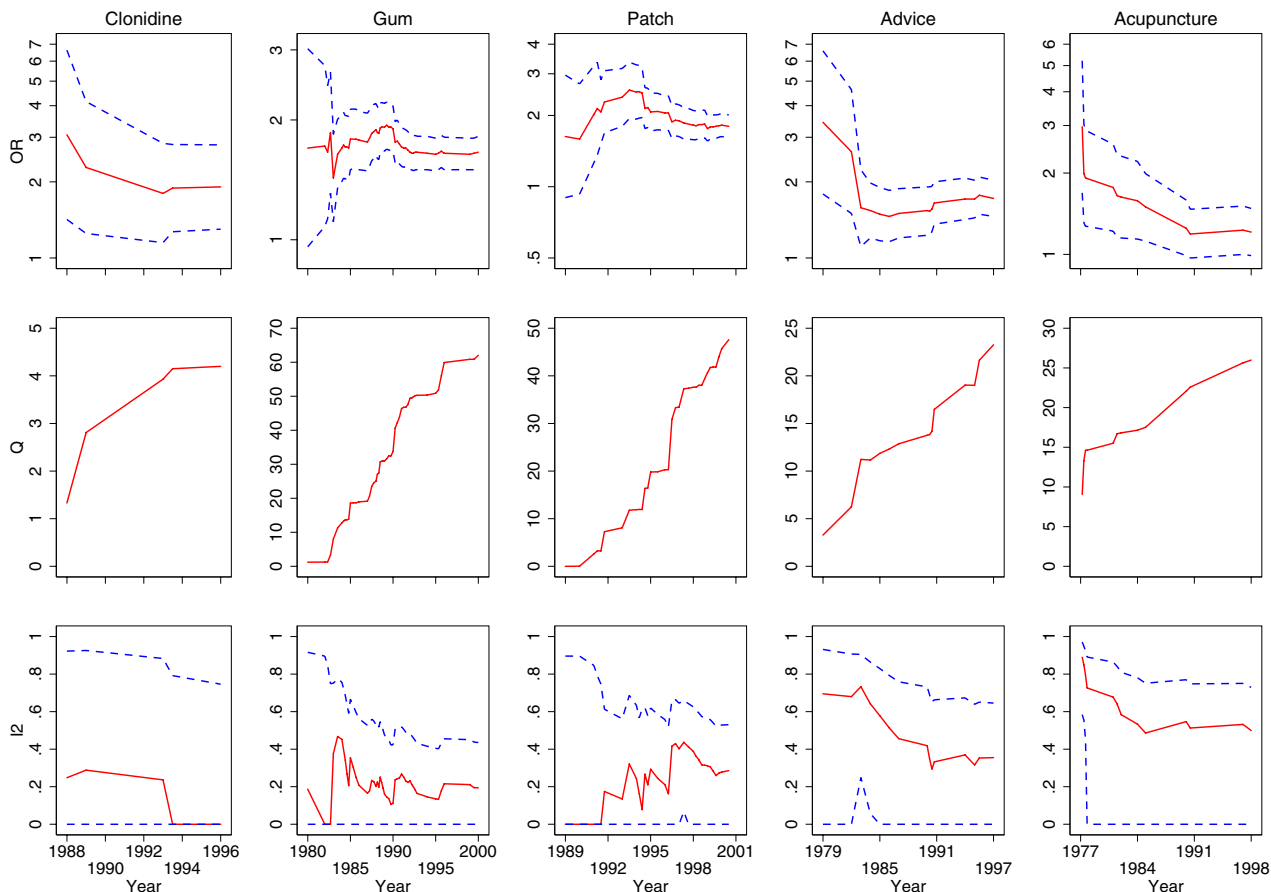
**Results**

Figure 1 describes the evolution of the OR (first row),  $Q$  (second row) and  $I^2$  (third row) for each of the five interventions. We performed cumulative meta-analyses in the order in which studies were published. As expected,  $Q$  is non-decreasing over the entire period. Moreover, the average changes in  $Q$  are equally inconsistently reflected in the changes in the OR. For instance, the OR for nicotine replacement using patches changes only by 12% from 1996 to 2000 while  $Q$  changes by more than 130%. Note the difficulty of comparing  $Q$  across studies because of the different ranges as exemplified by the varying ordinates.

The utility of  $I^2$  is immediately apparent. First, the use of standard limits from 0 to 1 allows comparisons to be made between different meta-analysis. Second, the absolute level of  $I^2$  may be interpreted in a meaningful way. For instance, we see that between-study variance ( $\tau^2$ ) played a minimal role in comparison to within-study variance ( $\sigma^2$ ) prior to 1993 for clonidine while essentially disappearing after this time. In contrast, the results for physician advice and acupuncture showed that  $\tau^2$  was proportionally large in earlier cumulations. While there was a decrease in the overall level over time, the absolute amounts remained above 35%.

Third, changes in  $I^2$  correlate well with changes in the OR, but plateauing of the OR does not necessarily mean a plateauing of  $I^2$ . The dips and rises in the OR are reflected in the rise or fall of  $I^2$ . This can be seen quite clearly in the early and middle periods in the cumulations of both gum and patch. The slight perturbations in the OR are magnified in trends and absolute levels of  $I^2$ . However, a steady level of the OR does not necessarily mean that the  $I^2$  value stabilises, too. This is seen in the different behaviour of  $I^2$  during the later periods of the cumulations for gum and patch.

Since  $I^2$  and  $H^2$  are related arithmetically by equation (5), we would expect to see similarities in their graphical representations. This is borne out in Figure 2 in which  $I^2$  and  $H$  values are plotted across time for the cumulative meta-analyses of the five interventions.



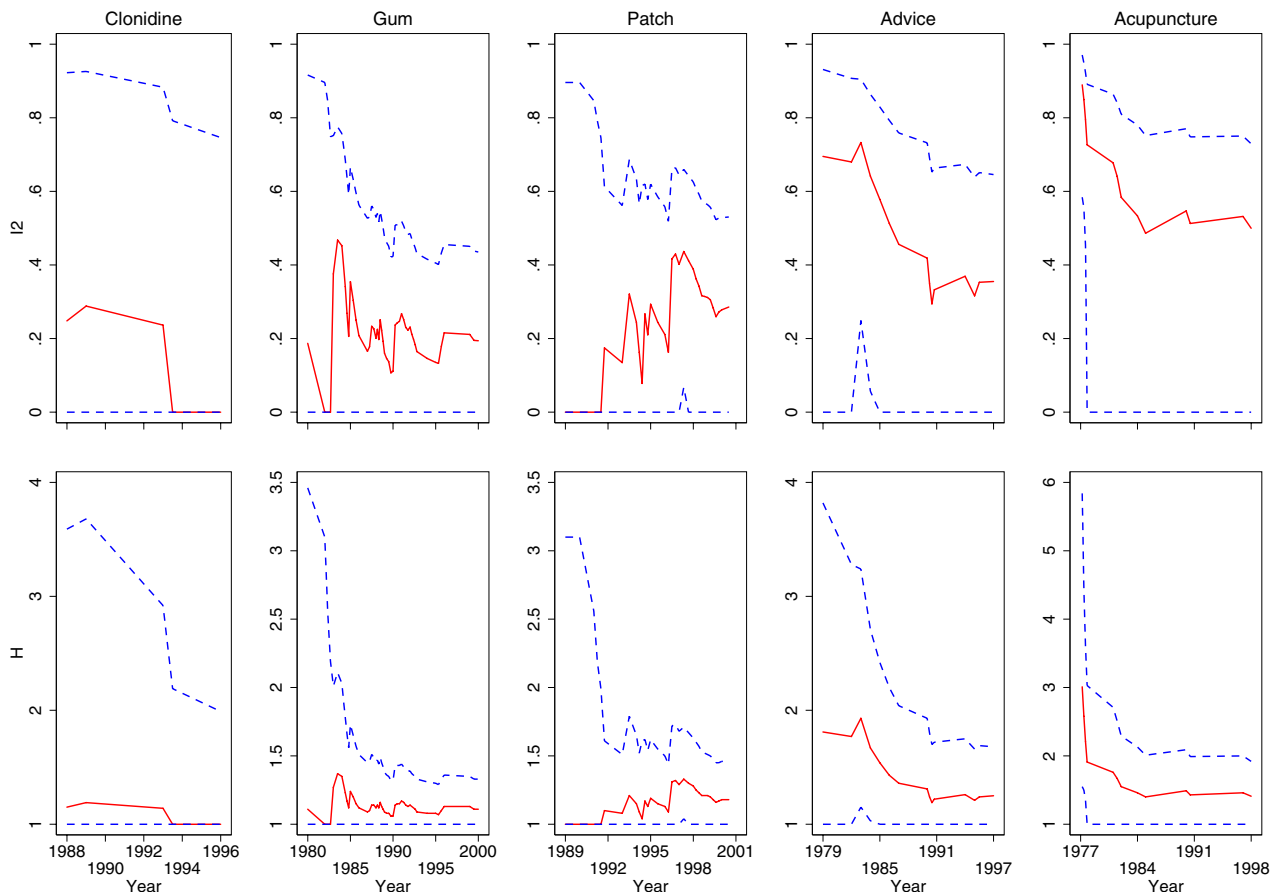
**Figure 1**  
 Cumulative meta-analytic estimates of the odds ratio (OR) for clonidine, nicotine replacement therapy (gum and patch), physician advice and acupuncture as smoking cessation interventions across time are shown in the first row of graphs. The second row outlines the development of Cochrane's Q. The last row describes the trends in  $I^2$ . All analyses were cumulated by publication date.

Note that  $I^2$  is limited by [0, 1] and trends that approach these limits will be compressed. On the other hand, while  $H$  has a lower limit of 1, it has no theoretical upper limit. However, since very little heterogeneity is expected in such high values, the scale is wasted, producing a compression effect in the lower portion of the ordinate range. This effect is much more pronounced than the trends seen in the  $I^2$  curves, resulting in poorer resolution.

The evolution of measurable heterogeneity in relation to time was considered previously. When plotted in this manner, periods of increased or decreased heterogeneity may be identified for further study. For instance,  $I^2$  was about twice its final value in the mid-1980's for nicotine replacement therapy using gum (Figure 3, column 1). Fol-

lowing this period,  $I^2$  values hovered around 20% until 2000. The increase in heterogeneity may be due to the differences in the components of the intervention [17-20] or quality of the research [20-22] in those studies that were conducted at the time. For instance, one of the studies in 1983 [17] assigned participants into one of four groups, only one of which included the intervention of interest (nicotine gum). In the meta-analysis, the results for the other three groups were collapsed to provide a "control" group against which the effect of nicotine gum could be compared.

By cumulatively performing a meta-analysis by ascending (or descending) control event rates, additional information may be gained in the assessment of heterogeneity



**Figure 2**  
A comparison between trends in  $I^2$  (first row) and  $H$  (second row) across time for smoking cessation interventions.

especially with regard to the severity of the condition in the study population. Note that small control event rates were associated with smaller values of  $I^2$  in nicotine replacement therapy using gum (Figure 3, column 2). In studies in which the control event rate of smoking cessation was at least 10%, the proportional contribution of between-study variance to the total variance was about 20%.

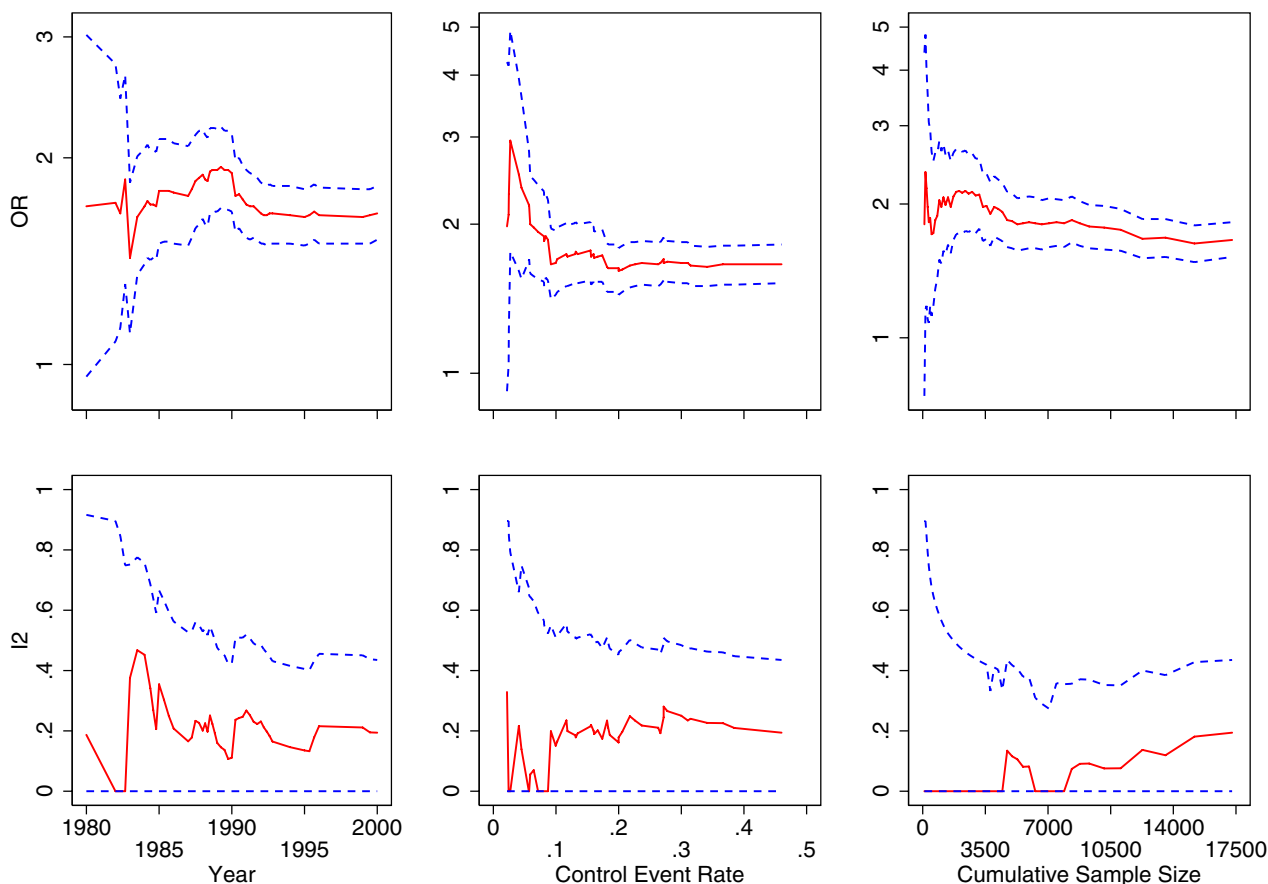
Meta-analyses may be performed cumulatively over the studies arranged in order of increasing sample size (Figure 3, column 3). The smaller number of endpoints associated with smaller studies impacts on their variability. In the case of nicotine replacement therapy using gum, the variability seemed to arise from within-study sources rather than between-study sources. It was not until more

than 3,500 patients were studied cumulatively in 28 studies that appreciable heterogeneity as measured by  $I^2$  was detected.

Results for the meta-analyses performed according to year, control event rate and sample size for patches, physician advice and acupuncture are available in Additional File 1.

**Discussion**

We have provided a description of the use of cumulative meta-analytic techniques to trace the evolution of measures of heterogeneity. By evaluating the patterns in heterogeneity, temporal or other relationships may be examined with the view of evaluating the impact of specific levels of heterogeneity in association with the overall estimate of effect. For instance, the increase in



**Figure 3**  
Trends in the summary odds ratio (OR) and  $I^2$  for nicotine replacement therapy using gum. The first column shows trends by year of publication, the second by control event rate, and the last by cumulative sample size.

heterogeneity measured using  $I^2$  showed an approximate doubling over the early 1980's that may be due to differences in study quality or components of the intervention [14].

The techniques described above will facilitate the examination of sources or relationships that contribute to heterogeneity. The measured examination of such sources has been advocated widely [8,23,24] and, prior to the introduction of  $I^2$  and related measures, was limited to general advice – differentiating between clinical and statistical heterogeneity [23], covariate-specific evaluations [25], etc. The use of the newly developed measures will allow the more systematic examination of meta-analytic variation.

The discussion regarding the cumulative meta-analysis of effect estimates (eg, OR, relative risks) has previously focused on demonstrating that statistical significance could have been reached at an earlier period or using a smaller number of patients [3-5]. In contrast, the use of cumulative meta-analysis in the examination of heterogeneity does not have as its primary objective the anticipation of any attainment of statistical significance in the level of variability.

Our paper had several limitations. First, we relied on available Cochrane reviews in calculating summary estimates. We assumed that the authors of the systematic reviews conducted a reasonable effort to identify,

appraise, extract, and analyse the individual trial reports relating to the particular intervention. In 1998, Cochrane reviews were found to be more methodologically rigorous in comparison to paper-based meta-analysis [26] and we have had no cause to have reservations about the generalisation of this finding to the present time.

Second, we limited our analysis to interventions designed to promote smoking cessation. The choice of topic may very well have resulted in peculiar or distinctive results. However, we were primarily interested in describing an approach to assessing heterogeneity rather than concentrating on the effectiveness of the interventions.

Third, we limited our description of the relationship between heterogeneity and specific covariates to three: year, control event rate and sample size. We submit that additional covariates may be used in similar ways to trace the development of heterogeneity. Some useful covariates may include study-related variables such as study quality, patient-related variables such as age or severity of disease, treatment-related variables such as duration or delay in receipt of treatment, etc. As with meta-regression techniques, the examination of these relationships depends on the availability of the variables in the individual study reports.

The evaluation of heterogeneity should be seen as crucial to the overall approach used in systematic reviews. We believe that the present focus on the meta-analytic summaries of primary effect estimates without regard for the sources of variability will lead to the production of an incomplete picture of the intervention in question. The development of measures of heterogeneity will certainly improve the subsequent description and examination of the phenomenon;  $I^2$  is already included in the standard software used in the preparation of Cochrane reviews. In the same way that we demand the reporting of measures of variability with conventional measures of central tendency (ie., standard deviation with the mean), should we not also expect quantitative summaries to include  $I^2$  (or similar measures)?

## Conclusions

When combined with cumulative meta-analytic techniques, measures of heterogeneity such as  $I^2$  and  $H$  allow for the easy description of the impact of variability. This adds an important perspective in describing the evolution of pooled effect estimates.

## Competing interests

None declared.

## Authors' contributions

EV conceptualised the study and defined the protocol. EV and SZ shared data extraction and analysis responsibilities. The manuscript was prepared by both authors. EV and SZ approved the final version of the paper.

## Additional material

### Additional File 1

*Trends in the summary odds ratio (OR) and  $I^2$  by year of publication, control event rate and cumulative sample size.* ∞ Figure A2-1. *Nicotine replacement therapy: gum.* ∞ Figure A2-2. *Nicotine replacement therapy: patch.* ∞ Figure A2-3. *Physician advice.* ∞ Figure A2-4. *Acupuncture.* Click here for file  
[<http://www.biomedcentral.com/content/supplementary/1471-2288-4-18-S1.pdf>]

## Acknowledgements

None declared.

## References

- Mulrow CD: **The medical review article: state of the science.** *Ann Intern Med* 1987, **106(3)**:485-488.
- Mulrow CD, Thacker SB, Pugh JA: **A proposal for more informative abstracts of review articles.** *Ann Intern Med* 1988, **108(4)**:613-615.
- Lau J, Antman EM, Jimenez-Silva J, Kupelnick B, Mosteller F, Chalmers TC: **Cumulative meta-analysis of therapeutic trials for myocardial infarction.** *N Engl J Med* 1992, **327(4)**:248-254.
- Lau J, Schmid CH, Chalmers TC: **Cumulative meta-analysis of clinical trials builds evidence for exemplary medical care.** *J Clin Epidemiol* 1995, **48(1)**:45-57. discussion 59-60
- Berkey CS, Mosteller F, Lau J, Antman EM: **Uncertainty of the time of first significance in random effects cumulative meta-analysis.** *Control Clin Trials* 1996, **17(5)**:357-371.
- Deeks JJ, Altman DG, Bradburn MJ: **Statistical methods for examining heterogeneity and combining results from several studies in meta-analysis.** In: *Systematic Reviews in Health Care: Meta-analysis in Context* Edited by: Egger M, Davey Smith G, Altman DG. London: BMJ Publishing Group; 2001.
- Cochran WG: **The combination of estimates from different experiments.** *Biometrics* 1954, **10**:101-129.
- Hardy RJ, Thompson SG: **Detecting and describing heterogeneity in meta-analysis.** *Stat Med* 1998, **17(8)**:841-856.
- Higgins JP, Thompson SG: **Quantifying heterogeneity in a meta-analysis.** *Stat Med* 2002, **21(11)**:1539-1558.
- Higgins JP, Thompson SG, Deeks JJ, Altman DG: **Measuring inconsistency in meta-analyses.** *BMJ* 2003, **327(7414)**:557-560.
- Takkouche B, Cadarso-Suarez C, Spiegelman D: **Evaluation of old and new tests of heterogeneity in epidemiologic meta-analysis.** *Am J Epidemiol* 1999, **150(2)**:206-215.
- White AR, Rampes H, Ernst E: **Acupuncture for smoking cessation.** In: *The Cochrane Library Issue 3 Oxford: Update Software; 2003.*
- Gourlay SG, Stead LF, Benowitz NL: **Clonidine for smoking cessation.** In: *The Cochrane Library Issue 3 Oxford: Update Software; 2003.*
- Silagy C, Lancaster T, Stead L, Mant D, Fowler G: **Nicotine replacement therapy for smoking cessation.** In: *The Cochrane Library Issue 3 Oxford: Update Software; 2003.*
- Silagy C, Stead LF: **Physician advice for smoking cessation.** In: *The Cochrane Library Issue 3 Oxford: Update Software; 2003.*
- Mantel N, Haenszel W: **Statistical aspects of the analysis of data from retrospective studies of disease.** *J Natl Cancer Inst* 1959, **22(4)**:719-748.
- British Thoracic Society: **Comparison of four methods of smoking withdrawal in patients with smoking related diseases.**

**Report by a subcommittee of the Research Committee of the British Thoracic Society.** *Br Med J (Clin Res Ed)* 1983, **286(6365)**:595-597.

18. Russell MA, Merriman R, Stapleton J, Taylor W: **Effect of nicotine chewing gum as an adjunct to general practitioner's advice against smoking.** *Br Med J (Clin Res Ed)* 1983, **287(6407)**:1782-1785.
19. Fagerstrom KO: **Effects of nicotine chewing gum and follow-up appointments in physician-based smoking cessation.** *Prev Med* 1984, **13(5)**:517-527.
20. Killen JD, Maccoby N, Taylor CB: **Nicotine gum and self-regulation training in smoking relapse prevention.** *Behav Ther* 1984, **15**:234-248.
21. Hjalmarson AI: **Effect of nicotine chewing gum in smoking cessation. A randomized, placebo-controlled, double-blind study.** *JAMA* 1984, **252(20)**:2835-2838.
22. Schneider NG, Jarvik ME, Forsythe AB: **Nicotine vs. placebo gum in the alleviation of withdrawal during smoking cessation.** *Addict Behav* 1984, **9(2)**:149-156.
23. Thompson SG: **Why sources of heterogeneity in meta-analysis should be investigated.** *BMJ* 1994, **309(6965)**:1351-1355.
24. Thompson SG, Smith TC, Sharp SJ: **Investigating underlying risk as a source of heterogeneity in meta-analysis.** *Stat Med* 1997, **16(23)**:2741-2758.
25. Thompson SG, Sharp SJ: **Explaining heterogeneity in meta-analysis: a comparison of methods.** *Stat Med* 1999, **18(20)**:2693-2708.
26. Jadad AR, Cook DJ, Jones A, Klassen TP, Tugwell P, Moher M, Moher D: **Methodology and reports of systematic reviews and meta-analyses: a comparison of Cochrane reviews with articles published in paper-based journals.** *JAMA* 1998, **280(3)**:278-280.

### Pre-publication history

The pre-publication history for this paper can be accessed here:

<http://www.biomedcentral.com/1471-2288/4/18/prepub>

Publish with **BioMed Central** and every scientist can read your work free of charge

*"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."*

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)

