14 October 2002

The Editor
British Medical Journal

Dear Editor

**Celebrex's relative GI safety is overstated?**

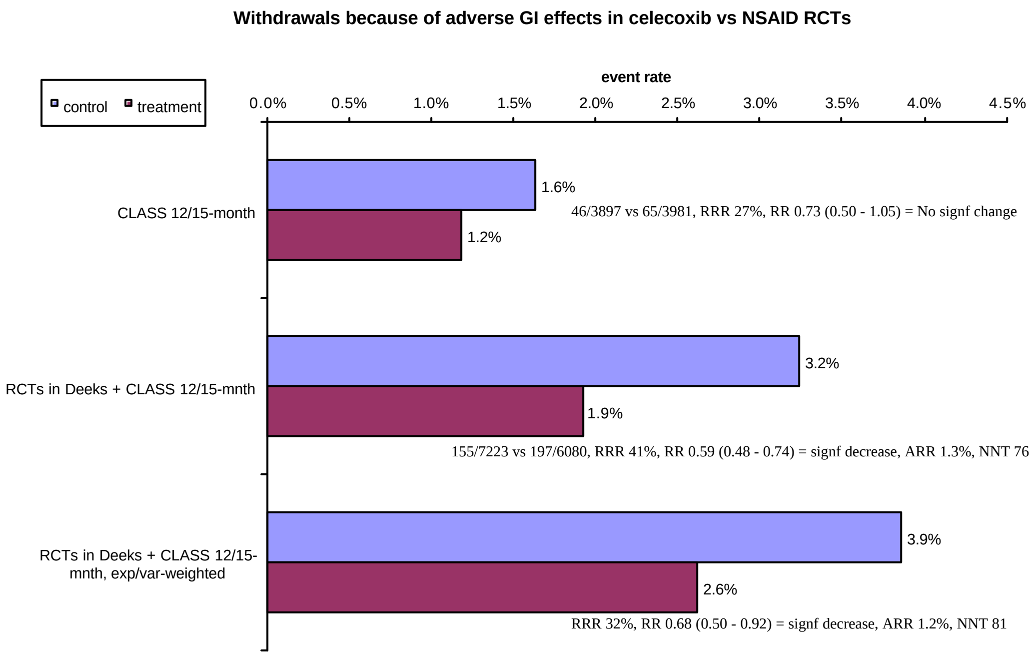
The editorial by Roger Jones[1] makes important points about the limitations of the meta-analysis by Jon Deeks and colleagues[2] for celecoxib. However, we also note that the Deeks meta-analysis does not account for the 12-15-month data for the CLASS study compiled by the FDA[3 4] and cited by Peter Juni and colleagues' critical editorial.[5] We have abstracted the 12-15 month CLASS data (www.fda.gov/ohrms/dockets/ac/01/briefing/3677b1_03_med.pdf tables 6, 13, 25, 26, 29, 30, www.fda.gov/ohrms/dockets/ac/01/briefing/3677b1_04_stats.doc tables 2, 5, 7) and applied them to the Deeks analysis, and find that these give a different picture.

The FDA's analysis of CLASS is more complete than that published in JAMA[6], which the Juni editorial criticised for not accounting for the 12-15 month data. We believe Deeks and colleagues offer an unconvincing explanation for limiting the analysis to the six-month follow up for CLASS, insufficient to justify the post hoc changes in design, outcomes and analysis. CLASS's 12-15 month data amount to 7878 person-years of follow-up, compared with 1252 person-years from the other RCTs measuring withdrawals because of gastrointestinal (GI) events. We also believe these data should have been included in figure 2 of the Deeks paper 'Celecoxib vs NSAID Any GI adverse effects', materially affecting those results, as shown below.

Looking at withdrawals because of both serious upper GI events and endoscopic ulcers, the 12-15 month FDA data for CLASS showed no statistically significant reduction in risk (relative risk (RR) 0.73 (95% CI 0.50 to 1.05)), distinct from the 39% RRR for CLASS's 6-month data suggested in Deeks figure 4. Likewise, for withdrawals due to serious GI events only (not endoscopic ulcers), the 12-15 month data meant the incidence of serious events (n= 20 / 3897) was nearly that of the other NSAIDS (n=24 / 3981),[7] not the 11 vs. 20 effect described by Deeks for the 6-month data.[2,5]

Combining the 12-month CLASS GI withdrawal data with the seven RCTs in Deeks fig 2 'Celecoxib vs NSAID Any GI adverse effects' gave an overall RRR of 41% (variance-weighted RR 0.59 (0.48 - 0.74), fixed effects method). Adjusting for the longer exposure experienced in CLASS (12-15 months rather than 12 weeks in the seven other RCTs) decreased the overall RRR to 32% (exposure/variance-weighted RR 0.68 (0.50 – 0.92)) (see Figure 1 below) - somewhat less than the 46% reported in Deeks fig 2.

**Figure 1. Withdrawals because of adverse GI effects in celecoxib vs. NSAID RCTs**




These analyses suggest celecoxib still causes statistically significant reductions in GI adverse events overall, but appreciably less than that suggested for the seven other RCTs by Deeks fig 2.

Furthermore, Deeks et al reported no statistically significant difference between low-dose aspirin and non-aspirin use for both endoscopic ulcers (four RCTs, 51% vs. 73% RRR, p 0.18) and for CLASS (specific outcome not stated, 19% vs 50% RRR, p 0.44). However, using the 12-15-month data for CLASS suggests a different picture. Non-aspirin users had a statistically significant 42% RRR (22 / 3154 vs 39 / 3169, RR 0.58 (0.35 - 0.95)), whereas aspirin users showed no reduction in risk (24 / 743 vs 26 / 812, RR 1.02 (0.59 - 1.74)). The difference between the subgroups' RRRs over the 12 months was statistically significant (p 0.03).

Taken in entirety (combining both endoscopic ulcers with CLASS's GI withdrawals+ulcers)[1], significant differences between subgroups persist. When the 12-15-month CLASS data are included, the meta-analysis gives a non-significant 28% RRR for aspirin use (RR 0.72 (0.48 – 1.06)) compared with a 72% RRR for non-aspirin use (RR 0.28 (0.22 – 0.35)), the difference between RRRs being statistically significant (p <0.01). Extending the analysis to adjust for the greater exposure conferred by CLASS gave a 4% RRR for aspirin users (exposure/variance-weighted RR 0.96 (0.63 - 1.46), versus 52% for non-aspirin use (e/v-w RR 0.48 (0.33 – 0.70)), p <0.01 (see Figures 2 and 2A below):

[1] Numbers of CLASS withdrawals were comparatively low when compared with more sensitive GDUs found on 12-week mandatory endoscopy in the four other RCTs. Numbers of ulcers detected by routine endoscopy at 12 weeks reported in Deeks et al figure 5 (25% control incidence) were considerably higher than numbers of withdrawals because of adverse GI effects for corresponding RCTs reported in Deeks figure 2 (6%) and in CLASS (1.6%). Hence combining the two sets of data understates ulcer burden occurring in CLASS.

**Figure 2. Aspirin use in celecoxib vs. NSAID RCTs, GDU + withdrawals from GI events: RCTs in Deeks + CLASS 12/15-month, exposure/variance-weighted**

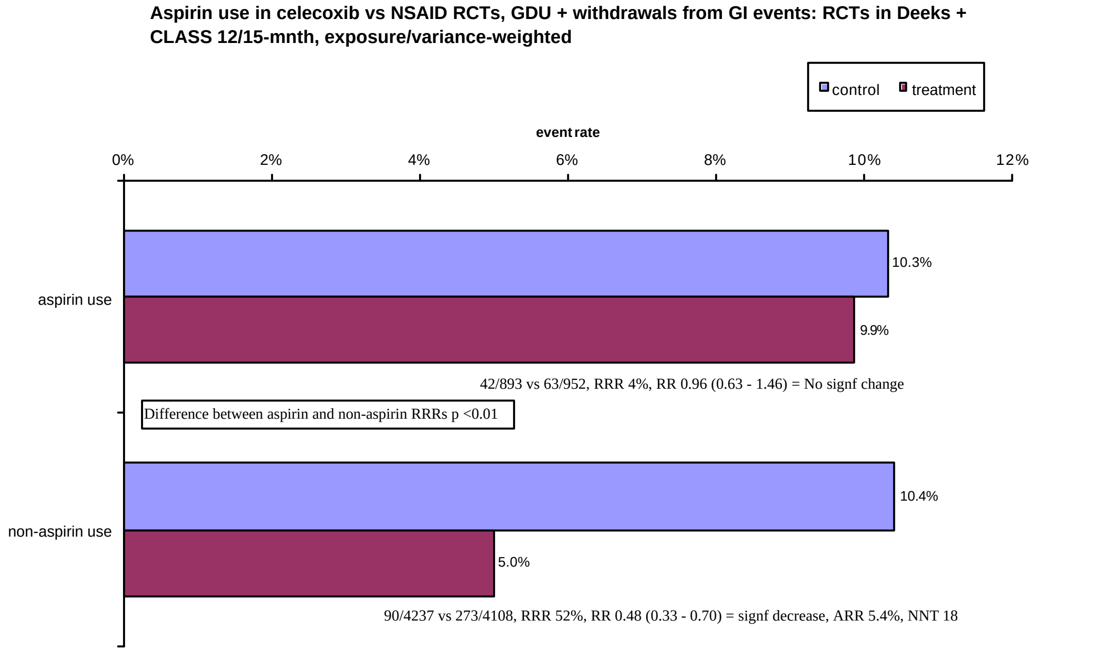
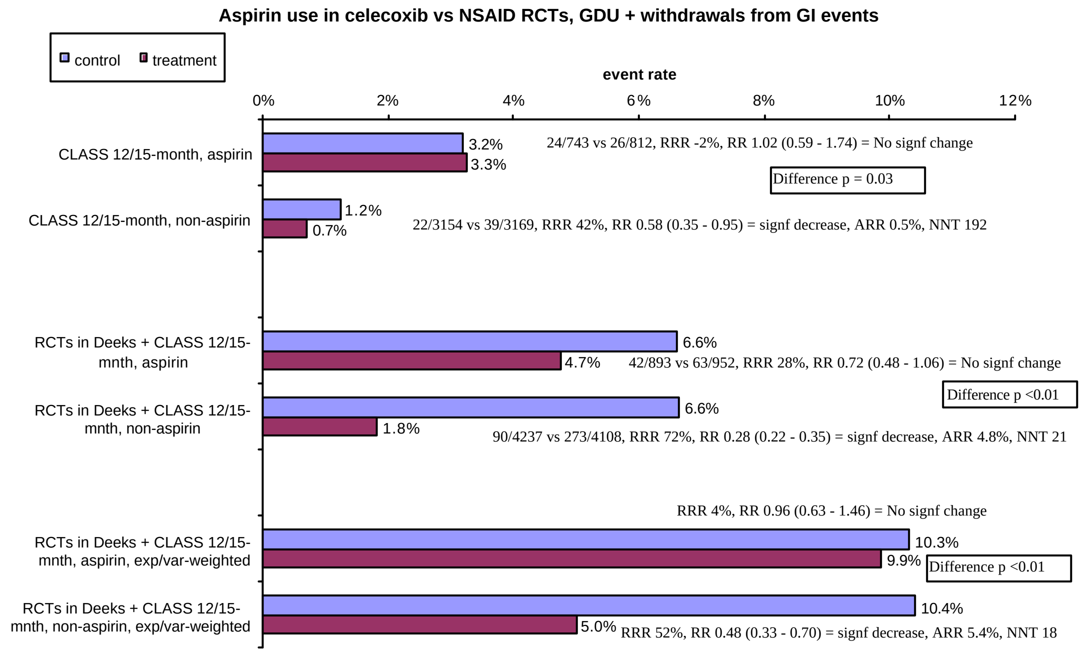



**Figure 2A. Aspirin use in celecoxib vs NSAID RCTs, GDU + withdrawals from GI events**



Hence we disagree with the implication that the benefits of celecoxib extend equally to aspirin users, and agree with NICE's current precautionary recommendation to withhold celecoxib from aspirin users.

Summary results can be seen in the Table below. Excel spreadsheet calculations and tables detailing the above findings are available on PHARMAC's website at www.pharmac.govt.nz publications page.

Methods for calculating person-year weighted incidence rates, weighted rate ratios and relative risk reductions are described in the Appendix below.

We note there appears to be significant funnel plot asymmetry for the seven RCTs reported for GI withdrawals (slope 0.90, intercept 5.4, R2 0.45), with minor improvement when the CLASS results are included (see figures 3 and 4 below):

**Figure 3. Funnel plot asymmetry, celecoxib RCTs (withdrawals because of adverse GI effects)**

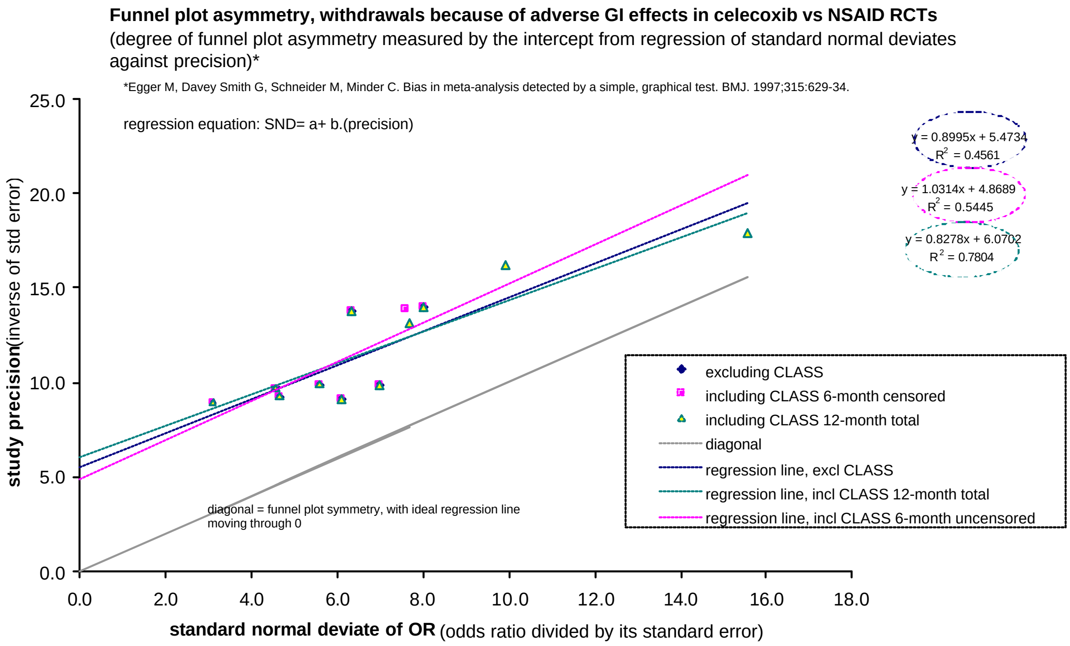
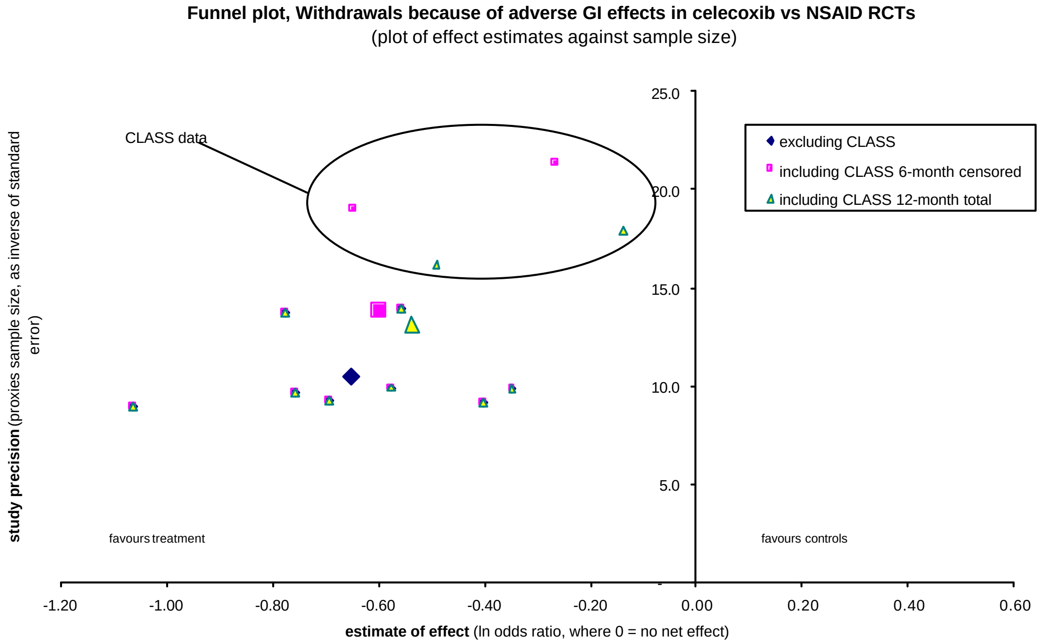


**Figure 4. Funnel plot, celecoxib RCTs (withdrawals because of adverse GI effects)**




Such asymmetry raises the question of possible selection bias (e.g. publication bias), although might be equally explained by alternatives such as poor methodological quality of smaller studies, true heterogeneity, size of effect differs according to study size, artefact, and chance.[8]

This all said, results from the Success-I trial might again influence overall results.[9] But future analyses must take account of the full CLASS data. In the meantime the results presented for celecoxib suggesting favourable GI safety need careful scrutiny.

Finally, we note too the relatively high NNTs to prevent GI adverse events seen with celecoxib when compared with older NSAIDs (see figures 1 and 2 above), let alone negligible improvement in musculo-skeletal symptoms. Also, Cox-2 inhibitors are expensive relative to older NSAIDS in the New Zealand setting. A preliminary pharmacoeconomic analysis gives a figure of over NZ$500,000/QALY, even when using the 6-month CLASS data.


Scott Metcalfe
Public Health Physician

Sean Dougherty
Analyst

Wayne McNee
Chief Executive

Pharmaceutical Management Agency (PHARMAC)
Level 1 Old Bank Chambers
98 Customhouse Quay
PO Box 10 254
Wellington 6001
NEW ZEALAND

## Table. Summary results, recalculation of Deeks et al (BMJ 2002) meta-analysis of celcoxib RCTs, adverse GI effects

| | duration | total pts | | overall variance-weighted effects (fixed effects method) | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | (no adjustment for study duration/quality) | | | ARR info | | aspirin vs non-aspirin use | | |
| | (years) | patients | person-years measured | RR (95% CLs) | RRR | signf | (baseline event rates) | NNT | RRR ASA/nonASA | X2 heterogeneity | p-value |
| **Withdrawals because of adverse GI effects in Celecoxib vs NSAID RCTs** | | | | | | | | | | | |
| **RCTs in Deeks et al 2002 fig 2 'Celecoxib vs NSAID Any GI adverse effects'** | | | | | | | | | | | |
| 7 RCTs in Deeks, which excluded CLASS* | 0.23 | 5,425 | 1,252 | 0.54 (0.41,0.70) | 46% | -ve | 6.3% | 34 | | | |
| * same as Deeks et al 2002 fig 2 'Celecoxib vs NSAID Any GI adverse effects' | | | | | | | | | | | |
| **CLASS 6- and 12-month results for clinically-significant upper gastrointestinal events + gastroduodenal ulcer (CSUGIE+GDU)**** | | | | | | | | | | | |
| 6-mnth, uncensored*** | 0.50 | 7,878 | 3,939 | 0.63 (0.41,0.98) | 37% | -ve | 1.2% | 220 | | | |
| 12-mnth, uncensored | 1.00 | | 7,878 | 0.71 (0.49,1.04) | 29% | | 1.6% | 222 | | | |
| 6-mnth, total (uncensored + censored) | | | | 0.65 (0.42,0.99) | 35% | -ve | 1.3% | 220 | | | |
| 12-mnth, total (uncensored + censored) | | | | 0.73 (0.50,1.05) | 27% | | 1.6% | 223 | | | |
| **source: FDA NDA20-998 Witter J. Medical officer review. http://www.fda.gov/ohrms/dockets/ac/01/briefing/3677b1_03_med.pdf. Tables 25,26,29 | | | | | | | | | | | |
| *** same as Deeks et al 2002 fig 4 'Serious upper gastrointestinal events + ulcers' | | | | | | | | | | | |
| **combined RCTs in Deeks et al plus CLASS for [withdrawals due to any adverse GI effects]** | | | | | | | | | | | |
| RCTs in Deeks, + CLASS 6-mnth, uncensored*** | 0.39 | 13,303 | 5,191 | 0.56 (0.44,0.70) | 44% | -ve | 3.7% | 61 | | | |
| RCTs in Deeks, + CLASS 12-mnth, total | | | | 0.59 (0.48,0.74) | 41% | -ve | 3.5% | 70 | | | |
| **Gastrointestinal impact of aspirin use in Celecoxib vs NSAID RCTs** | | | | | | | | | | | |
| **GDU detected by routine endoscopy at 12 weeks, RCTs in Deeks et al 2002 fig 5 (excludes CLASS)** | | | | | | | | | | | |
| non-aspirin use | 0.23 | 2,022 | 467 | 0.27 (0.21,0.34) | 73% | -ve | 25.7% | 5 | | | |
| aspirin use | 0.23 | 290 | 67 | 0.52 (0.30,0.87) | 48% | -ve | 25.0% | 8 | 0.66 | 3.69 | 0.00 |
| **CLASS 6- and 12-month withdrawals because of clinically-significant upper gastrointestinal events + gastroduodenal ulcer (CSUGIE+GDU)*** | | | | | | | | | | | |
| non-aspirin, 6-mnth, uncensored | 0.50 | 6,323 | 3,162 | 0.52 (0.29,0.91) | 48% | -ve | 1.0% | 206 | | | |
| aspirin, 6-mnth, uncensored | 0.50 | 1,555 | 778 | 0.91 (0.45,1.81) | 9% | | 2.1% | 517 | 0.19 | 1.03 | 0.09 |
| non-aspirin, 12-mnth, total | 1.00 | 6,323 | 6,323 | 0.58 (0.35,0.95) | 42% | -ve | 1.2% | 193 | | | |
| aspirin, 12-mnth, total | 1.00 | 1,555 | 1,555 | 1.02 (0.59,1.74) | -2% | | 3.2% | -1930 | -0.04 | 1.55 | 0.03 |
| *source: FDA NDA20-998 Witter J. Medical officer review. http://www.fda.gov/ohrms/dockets/ac/01/briefing/3677b1_03_med.pdf. Tables 26,30 | | | | | | | | | | | |
| **combined [RCTs in Deeks et al 12-week endoscopic GDUs] plus [CLASS CSUGIE+GDU]** | | | | | | | | | | | |
| non-aspirin use, [CLASS 6-mnth, uncensored] | 0.43 | 8,345 | 3,628 | 0.27 (0.21,0.34) | 73% | -ve | 10.0% | 14 | | | |
| aspirin use, [CLASS 6-mnth, uncensored] | 0.46 | 1,845 | 844 | 0.63 (0.40,0.97) | 37% | -ve | 9.0% | 30 | 0.51 | 13.90 | 0.00 |
| non-aspirin use, [CLASS 12-mnth, total] | 0.81 | 8,345 | 6,790 | 0.28 (0.22,0.35) | 72% | -ve | 8.3% | 17 | | | |
| aspirin use, [CLASS 12-mnth, total] | 0.88 | 1,845 | 1,622 | 0.72 (0.48,1.06) | 28% | | 8.3% | 43 | 0.39 | 16.85 | 0.00 |

## Appendix: Calculating person-year weighted incidence rates, weighted rate ratios and relative risk reductions

Person-year weighted incidence rates, weighted rate ratios (relative risks) (aRR) and weighted relative risk reductions (aRRR) can be calculated as follows:

| | |
|---|---|
| nt | = no. of patients in treated group responding, |
| Nt | = no. patients in treated group, |
| nc | = no. of patients in control group responding, |
| Nc | = no. patients in control group |
| crude response rate for treated patients | = $\sum nt/\sum Nt$ |
| crude response rate for control patients | = $\sum nc/\sum Nc$ |
| crude rate ratio (relative risk, RR) | = $(\sum nt/\sum Nt) / (\sum nc/\sum Nc)$ |
| crude odds ratio (OR) | = $(\sum nt/(\sum Nt-\sum nt)) / (\sum nc/(\sum Nc-\sum nc))$ |

Exposure-adjusted baseline event rates (aEc) can be derived by weighting control arms according to risk exposure (t.Nc),

where
- t = study duration,
- Nc = no. patients in control group, and
- t.N = risk exposure, in person-year equivalents

Pooled (adjusted) odds ratios for all studies (aOR) are weighted according to the variance of individual RCTs' odds ratios, with associated confidence limits (fixed effects, Peto one-step method)

Adjusted rate ratios (aRR) are derived from adjusted baseline event rates and pooled odds ratios, with associated confidence limits, according to the formula[10]:

$$aRR = 1 - \frac{(1-aEc).(1-aOR)}{1 - [aEc.(1-aOR)]}$$

where:
- aRR = adjusted rate ratio (ie adjusted relative risk)
- aEc = adjusted baseline event rate (ie control incidence rate, weighted according to (t.N)
- aOR = pooled (adjusted) odds ratio (weighted according to variance)

Adjusted relative risk reductions (aRRR) are derived from adjusted rate ratios, where aRRR = 1 – aRR, according to the formula:

$$aRRR = \frac{(1-aEc).(1-aOR)}{1 - [aEc.(1-aOR)]}$$

where:
- aRRR = adjusted relative risk reduction
- aRR = adjusted rate ratio
- aEc = adjusted baseline event rate
- aOR = pooled (adjusted) odds ratio

If adjusted baseline event rates are considered relevant to the New Zealand population, adjusted absolute risk reductions (aARR) are derived from adjusted baseline event rates and adjusted rate ratios, according to the formula:

$$aARR = aEc.aRRR$$

where:
- aARR = adjusted absolute risk reduction
- aEc = adjusted baseline event rate
- aRRR = adjusted relative risk reduction

Similarly, if adjusted baseline event rates are considered relevant to the New Zealand population, adjusted treatment event rates (aEt) are derived from adjusted baseline event rates and adjusted rate ratios, according to the formula:

$$aEt = aEc.\ aRR$$

where:
- aEt = adjusted treatment event rate (treated incidence rate)
- aEc = adjusted baseline event rate
- aRR = adjusted rate ratio (relative risk, derived from pooled odds ratio)

[NB Odds ratios derive from relative risks according to the formula:

$$OR = \frac{RR.(1-Ec)}{1-(RR.\ Ec)}$$

$$= \frac{Et\ .((1/Ec)-1)}{1-Et}$$

where:
OR = odds ratio
RR = rate ratio (ie relative risk)
Ec = baseline (control) event rate
Et = treatment event rate ]

⇒ To account for the quality of contributing RCTs, each RCT can be weighted according to a quality score, for example PHARMAC's following modification of the Jadad criteria [11] (score 0-5):

| **Criterion (modified)** | **Source of bias (Cochrane Handbook taxonomy)** | **Scoring system** |
|---|---|---|
| Randomisation | Selection bias / confounding, i.e. systematic differences in comparison groups | Adequate =1, Inadequate/nil = 0 |
| Concealed allocation | Selection bias / confounding | Adequate =1, Unclear/not described = 0 Inadequate/nil = 0 |
| Blinding of receipt | Performance bias, i.e. systematic differences in care provided apart from the intervention being evaluated; recipients | Adequate, described =0.5, Unclear/not described = 0.25 Inadequate/nil = 0 |
| Blinding of provision | Performance bias; providers | Adequate, described =0.5, Unclear/not described = 0.25 Inadequate/nil = 0 |
| Follow-up | Attrition bias, i.e. systematic differences in withdrawals from the trial, affecting outcome measurement | Participants adequately accounted for =1, Unclear/not described = 0 Inadequate/nil = 0 |
| Blinding of assessment | Detection bias, i.e. systematic differences in outcome assessment; assessors | 0 (presumably incorporated into Blinding of provision) |

Combining these quality-based weights with the above [variance-based weights for odds ratios] and the [exposure-based (t.N) weights for adjusted baseline incidence rates] gives quality/variance weights and quality/exposure/variance weights. These quality-containing weights can be used to then calculate quality-weighted pooled odds ratios and quality-weighted adjusted baseline incidence rates, using the above formulae, and thus quality-weighted adjusted rate ratios, etc.

## References


[1] Jones R. Efficacy and safety of COX 2 inhibitors. BMJ 2002;325:607-608. http://bmj.com/cgi/content/full/325/7365/607

[2] Deeks JJ, Smith LA, Bradley MD. Efficacy, tolerability, and upper gastrointestinal safety of celecoxib for treatment of osteoarthritis and rheumatoid arthritis: systematic review of randomised controlled trials. BMJ 2002; 325: 619-623. http://bmj.com/cgi/content/full/325/7365/619

[3]Witter J. Medical officer review. www.fda.gov/ohrms/dockets/ac/01/briefing/3677b1_03_med.pdf

[4] Lu HL. Statistical reviewer briefing document for the advisory committee. www.fda.gov/ohrms/dockets/ac/01/briefing/3677b1_04_stats.doc

[5] Juni P, Rutjes AWS, Dieppe PA. Are selective COX-2 inhibitors superior to traditional non-steroidal anti-inflammatory drugs? BMJ 2002; 324: 1287-1288. http://bmj.com/cgi/content/full/324/7349/1287

[6] Silverstein FE, Faich G, Goldstein JL, Simon LS, Pincus T, Whelton A, et al. Gastrointestinal toxicity with celecoxib vs nonsteroidal anti-inflammatory drugs for osteoarthritis and rheumatoid arthritis: the CLASS study: A randomised controlled trial. JAMA 2000; 284: 1247-1255.

[7] Bassett K, Wright JM, Puil L, Perry TL Jr, Heran B, Cole C. Cyclooxygenase-2 inhibitor update: Journal articles fail to tell the full story. Can Fam Physician 2002;48:1455-1460.

[8] Clarke M, Oxman AD, editors. Formulating the problem. Cochrane Reviewers' Handbook 4.1.5 [updated April 2002]; Section 8.11. In: The Cochrane Library, Issue 2, 2002. Oxford: Update Software. Updated quarterly. http://www.cochrane.dk/cochrane/handbook/hbook811_Special_topics.htm

[9] Singh G, Goldstein J, Fort J, Bello A, Boots S. Success-I in osteoarthritis (OA) trial: celecoxib has similar efficacy to the conventional NSAIDS [abstract]. J Rheumatol 2001; 28(suppl 6.3): 6.

[10] algebraic transformation by PHARMAC of formulae in Sackett D, Straus S, Richardson WS, Rosenberg W, Haynes B. Evidence-based medicine: how to practice and teach EBM, 2nd edition. Oxford: Churchill Livingstone, 2000. p136 Table 5.1 Formulae to convert odds ratios (ORs) and relative risks (RRs) to NNTs.

[11] Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, McQuay HJ. Assessing the quality of reports of randomized clinical trials: is blinding necessary? Control Clin Trials. 1996;17:1-12.