HEDS Discussion Paper 09/08

Published paper

Madan J, Stevenson MD, Cooper KL, Ades AE, Whyte S, Akehurst R. Consistency between direct and indirect trial evidence: is direct evidence always more reliable? /Value in Health /2011;14(6):953-960.


ScHARR
School of Health and
Related Research

The University Of Sheffield.

ScHARR

Health Economics and Decision Science
Discussion Paper Series

No. 09/08

Consistency between direct trial evidence and Bayesian Mixed Treatment Comparison: Is head-to-head evidence always more reliable?

Jason Madan, Katy L Cooper, Matt D Stevenson, Sophie Whyte, Ron Akehurst

Health Economics and Decision Science, University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA, UK

Correspondence to: Jason Madan, Health Economics and Decision Science,University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA, UK.
Telephone: +44114 222 0827.
Fax: +44114 272 4095.
Email:j.madan@sheffield.ac.uk

Consistency between direct trial evidence and Bayesian Mixed Treatment Comparison: Is head-to-head evidence always more reliable?

Authors:
Jason Madan, BSc, MSc
Dr Katy L Cooper, BSc, PhD
Dr Matt D Stevenson, BSc, PhD
Dr Sophie Whyte, BSc, MMath, PhD
Professor Ron Akehurst, BSc (Econ), Hon MFPH

Corresponding author:
Jason Madan
Health Economics and Decision Science
School of Health and Related Research
University of Sheffield
30 Regent Court
Sheffield S1 4DA
Tel: +44 (0)114 222 0827
email: j.madan@sheffield.ac.uk

Acknowledgements
The research underlying this paper was funded by Amgen Ltd, and a research grant from Amgen (EUROPE) GmbH was provided to support the production of the manuscript. Amgen staff reviewed and made suggested edits to the manuscript, but final content, authorship and right to publication remained with the research team.

Key words: Bayesian methods, granulocyte colony-stimulating factors, G-CSFs, febrile neutropenia, mixed treatment comparison, methodology

Running title: The benefits of MTC: A case study

Abstract

Objectives

This study aims to highlight the benefits of Bayesian mixed treatment comparison (MTC), within a case study of the efficacy of three treatments (pegfilgrastim, filgrastim and lenograstim) for the prevention of febrile neutropenia (FN) following chemotherapy.

Methods

Two published meta-analyses have assessed the relative efficacy of the three treatments based on head-to-head trials. In the present study, all the trials from these meta-analyses were synthesised within a single network in a Bayesian MTC. Following a systematic review, the evidence base was then updated to include further recently-published trials. The meta-analyses and MTC were re-analysed using the updated evidence base.

Results

Using data from the previously-published meta-analyses only, the relative risk of FN for pegfilgrastim vs. no treatment was estimated at 0.08 (95% confidence interval: 0.03, 0.18) from the head-to-head trial and 0.27 (95% credible interval: 0.12, 0.60) from the MTC, reflecting strong inconsistency between the results of the direct and indirect methodologies. When subsequently-published head-to-head trials were included, the meta-analysis estimate increased to 0.29 (95% confidence interval: 0.15, 0.55), while the MTC gave a relative risk of 0.34 (95% credible interval: 0.23, 0.54). The initial MTC results were therefore a better predictor of subsequent study results than was the direct trial. The MTC was also able to estimate the probability that there were clinically significant difference in efficacy between the treatments.

Conclusions

Bayesian MTC provides clinically relevant information, including a measure of the consistency of direct and indirect evidence. Where inconsistency exists, it should not always be assumed that the direct evidence is more appropriate.

Introduction

There is often a need to synthesise evidence from multiple studies when evaluating medical interventions from a clinical or health policy perspective. Meta-analysis is an established statistical method for this synthesis when the evidence consists of several head-to-head comparisons of two treatments. However, there are often more than two treatments that could be used in a given situation. Even when a head-to-head choice is being made, the two interventions may have been compared with further treatments as well as, or instead of, each other. The ability of standard meta-analysis to synthesise evidence on more than two treatments is limited.

The estimation of the relative effectiveness of two treatments for which no head-to-head evidence is available is known as an indirect comparison. Bayesian methods provide a coherent framework for indirect comparisons, due to their explicit representation of parameter uncertainty. Where there are head-to-head comparisons in addition to the direct evidence, then the method can be extended to synthesise both direct and indirect evidence and produce an estimate of effectiveness that draws on the complete evidence base (1). This is known as a mixed treatment comparison (MTC).

The benefit of indirect comparisons is clear where no head-to-head trials exist. However, the value of going beyond the direct evidence with a mixed treatment comparison is more controversial. It is our experience that clinicians and decision-makers are often sceptical of the value of indirect evidence when direct evidence is available. This is reflected in the current methods guide for technology appraisal issued by the National Institute for Health and Clinical Excellence, which states that 'the Institute has a preference for data from head-to-head RCTs' (2).

Our view is that the synthesis of direct and indirect evidence has the potential to provide useful information alongside the head-to-head comparison. To illustrate this potential, as well as demonstrate additional benefits from the Bayesian indirect comparison methodology, we carried out a synthesis of the evidence around granulocyte colony-stimulating factors (G-CSFs) for prevention of febrile neutropenia (FN) during chemotherapy. Three G-CSFs are currently in use: filgrastim, pegfilgrastim, and lenograstim. The evidence synthesis aimed to inform economic modelling of the most cost-effective of these treatments for prevention of FN.

Methods

In order to populate the economic model with efficacy evidence, we carried out a systematic review of the literature. Two systematic reviews and meta-analyses of G-CSFs in relation to reducing FN events were identified. Kuderer et al(3) analysed studies of primary G-CSF prophylaxis (pegfilgrastim, filgrastim or lenograstim) versus no primary prophylaxis, whilst Pinto et al(4) analysed studies of prophylaxis with pegfilgrastim versus filgrastim. We assumed that these two systematic reviews comprehensively identified the relevant literature for the period covered by their search strategy. No head-to-head studies had been identified by the published systematic reviews comparing lenograstim with the other two G-CSFs, and only one head-to-head trial had been found of pegfilgrastim vs. no primary G-CSF (5). The magnitude of the effect recorded in this trial appeared inconsistent with expectations given

the results from meta-analyses of pegfilgrastim vs. filgrastim and filgrastim vs. no primary G-CSF. In order to quantify the extent of this apparent inconsistency, and to provide indirect estimates of the efficacy of lenograstim relative to the other two G-CSFs, we carried out an MTC of the evidence base as identified by Kuderer and Pinto.

In addition to the systematic reviews, our search of the literature identified three further studies of pegfilgrastim vs. no primary G-CSF(5-8) that had been published subsequently (incorporating four comparisons, since one of the studies assessed two populations). This provided an additional source of evidence to test whether the direct evidence alone, or the synthesis of evidence provided by the MTC, proved a more accurate prediction of the results of future head-to-head trials. To assess this, we carried out a conventional random-effects meta-analysis of the complete direct evidence base (including the additional studies) for the comparison of pegfilgrastim vs. no primary G-CSF. We also carried out another MTC on the entire evidence base, to assess the impact of the additional evidence.

In our view, one of the benefits of Bayesian methods in evidence synthesis is the ability to characterise uncertainty around parameters in terms of a posterior density function. This allows for statements to be made about the information present in an evidence base that go beyond whether a null hypothesis can be rejected at an arbitrary level of significance, to give instead results that are easier to interpret clinically. A commonly presented example is the credible interval, defined so that there is a 5% chance (for a 95% credible interval) that the true value lies outside the range. Conventional confidence intervals are often wrongly interpreted this way, illustrating the greater intuitive appeal of the credible interval (9).

In fact, the posterior density can be used to calculate the probability that the true value lies within any given range. If the range is defined as less than one, and the outcome measure is undesirable, this gives the probability that the treatment is superior to the control (avoiding the need to specify an arbitrary significance level). For equivalence comparisons, a range could be identified to represent clinical equivalence. For each of the pair-wise comparisons generated by the MTC, we calculated a 95% credible interval, the probability that treatment one was superior to treatment two, and the probability that the two treatments were equivalent. We defined equivalency for the sake of this case study as a relative risk of 1 +/- 0.1.

MTCs were carried out using the Bayesian modelling software WinBugs.(10). All posterior distributions were estimated using weakly informative prior distributions; gamma (0.001,0.001) distributions for precision parameters and normal (0,0.001) distributions for the logit of the probability of FN.

Results

The studies included in the analysis (those identified in the existing systematic reviews and the additional trials identified in our literature review) are described in Table 1, and the network of evidence is illustrated by Figure 1.

Table 2 compares results from the MTC performed on the evidence base reported in the existing systematic reviews with the results from the meta-analyses they report. The results are reasonably consistent, with one exception – in the initial evidence base, there is a marked

difference between the direct and indirect estimates for the relative risk of FN with pegfilgrastim vs. placebo (0.08 vs. 0.27), such that each point estimate lies outside the confidence / credible interval of the other. Table 3 presents the updated MTC including the additional studies not identified in the existing systematic reviews, alongside a meta-analysis of the updated evidence base for pegfilgrastim versus no G-CSF. This demonstrates that, once this additional evidence is included, the direct and indirect estimates are reasonably consistent (0.29 vs. 0.34), and that the additional evidence has brought the head-to-head estimate in line with the MTC estimate, rather than the other way around. As shown in Table 1, the MTC estimate from the original dataset was a better predictor of subsequent trial results than the head-to-head estimate. For the complete evidence base, the divergence between the direct and the synthesised evidence is greatest for the comparison of lenograstim with no G-CSF, although each estimate lies within the confidence / credible interval generated by its opposing method.

The MTC uses Markov Chain Monte Carlo methods (11) to estimate the posterior distribution for each relative risk. Figure 2 illustrates this distribution for the FN relative risk for pegfilgrastim vs. lenograstim. From these distributions, estimates can be made of the comparative efficacy of lenograstim against the other two G-CSFs, despite the lack of direct evidence. The median relative risk of FN for pegfilgrastim vs. lenograstim is 0.71 and for lenograstim vs. filgrastim is 0.88. The posterior distributions can also be used to calculate the probability that one G-CSF is superior to another or that two G-CSFs are clinically equivalent (as defined in the methods) can be calculated. Table 4 gives these results for the three pairwise comparisons of G-CSFs. The MTC showed that there is strong evidence for the superiority of pegfilgrastim over filgrastim; the probability that pegfilgrastim is superior is 98%, and that the two drugs are clinically equivalent is 4%. The evidence that pegfilgrastim is superior to lenograstim is convincing, with an 86% probability of superiority and a probability of equivalence of 13%. The evidence that lenograstim is superior to filgrastim was less conclusive, with the comparative figures being 71% and 30%.

Discussion

The case study we present illustrates the argument that MTC provides a useful check of the consistency between the direct and indirect evidence. If inconsistency is found, we would caution against the automatic acceptance of the direct result. In our example, the MTC estimate proved the more accurate predictor of future head-to-head studies. If our economic modelling called for an estimate of the relative efficacy of each treatment against no G-CSF, and we had only the evidence from the two existing systematic reviews to draw on, then relying only on the head-to-head evidence would have provided us with a less accurate estimate of the efficacy of pegfilgrastim.

It is true that, in this case study, there was only one head-to-head study in the initial evidence base for the relative efficacy of pegfilgrastim vs. placebo, and it is not unheard of for a single trial to prove to be an outlier, particularly if small. Whilst pointing out that the study was the largest single trial in the entire evidence base, we would agree that an MTC is most likely to be of value to decision-makers when head-to-head evidence is limited. However, we would argue that there may be value in carrying out an MTC even if several head-to-head studies exist, particularly in support of health economic evaluations. In most cases, we would expect the direct and indirect evidence to provide similar results; a reassuring additional finding. However, it is possible for such inconsistencies to arise with a larger head-to-head evidence base, and our view is that it is useful to identify such cases also.

Of course, if it is felt that the head-to-head estimate is always 'more valid' than the estimate from the complete evidence base, there is little to be gained by identifying discrepancies between the two. In many situations, there will be clinical arguments for preferring the direct evidence. In this case study, for example, it could be argued that the baseline risk in the head-to-head trials was lower than in the trials forming the indirect evidence base, and that this casts doubt on the indirect estimate. One benefit of identifying inconsistencies is that it encourages discussion of these issues. In response to the point raised above, it can be noted that the baseline risk and treatment effect relationship does not appear to apply within the head-to-head trials (as can be seen in table one). Furthermore, Kuderer et al point out that the evidence for such a relationship across the entire evidence base appears weak. Therefore, it does not seem that such a relationship can explain the particularly strong treatment effect recorded in the Vogel trial. In general, comparing the indirect and direct evidence can highlight when further investigation of factors driving heterogeneity is worthwhile, and (in a few cases), there may be an argument for placing some weight on the indirect evidence. For example, the patient populations in those trials may reflect the population of interest more closely than the head-to-head trial

There are further benefits to Bayesian MTC. Bayesian analysis differs from the classical approach in that it models uncertainty around parameters. This technical difference allows it to provide clinically relevant information more easily. It is the reason why indirect comparisons can be made within a coherent framework. Results can be adapted without difficulty to the situation. The example we gave involving the probability of clinical equivalence shows how results can be presented that are clinically significant rather than statistically significant. It is particularly relevant to situations where treatments are thought a priori to be similar e.g. where drugs are of the same class, and have similar modes of action.

Conclusions

Bayesian evidence synthesis, carried out in addition to conventional meta-analysis, provides useful information for clinical and policy decision-making. This includes a measure of the consistency between direct and indirect evidence. The presence of inconsistency warrants further investigation. Whilst there are strong reasons for preferring direct evidence, it should not be assumed that the head-to-head estimate is always more appropriate.

Reference List

(1) Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. Stat Med 2004 Oct 30;23(20):3105-24.

(2) National Institute for Health and Clinical Excellence (NICE). Guide to the methods of technology appraisal. http://www.nice.org.uk/media/B52/A7/TAMethodsGuideUpdatedJune2008.pdf; 2008.

(3) Kuderer N, Dale D, Crawford J, Lyman G. Impact of primary prophylaxis with granulocyte colony-stimulating factor on febrile neutropenia and mortality in adult cancer patients receiving chemotherapy: a systematic review. JOURNAL OF CLINICAL ONCOLOGY 2007;25(21):3158-6731.

(4) Pinto L, Liu Z, Doan Q. Comparison of pegfilgrastim with filgrastim on febrile neutropenia, grade IV neutropenia and bone pain: a meta-analysis of randomized controlled trials. Curr Med Res Opin 2007;23(9):2283-95.

(5) Vogel C, Wojtukiewicz M, Carroll R, et al. First and Subsequent Cycle Use of Pegfilgrastim Prevents Febrile Neutropenia in Patients With Breast Cancer: A Multicenter, Double-Blind, Placebo-Controlled Phase III Study. J Clin Oncol 2005;23(6):1178-84.

(6) Balducci L, Al-Halawani H, Charu V, et al. Elderly cancer patients receiving chemotherapy benefit from first-cycle pegfilgrastim. Oncologist 2007 Dec;12(12):1416-24.

(7) Romieu G, Clemens M, Mahlberg R. Pegfilgrastim supports delivery of FEC-100 chemotherapy in elderly patients with high risk breast cancer: A randomized phase 2 trial. Critical Reviews in Oncology/Hematology 2007;64:64-72.

(8) von Minckwitz G, Kummel S, du Bois A, et al. Pegfilgrastim +/- ciprofloxacin for primary prophylaxis with TAC (docetaxel/doxorubicin/cyclophosphamide) chemotherapy for breast cancer. Results from the GEPARTRIO study. Annals of Oncology 2008;19:292-8 (Epub Sept 2007).

(9) Cowan G. Statistical Data Analysis Oxford Press, 1998.

(10) Fryback DG, Stout NK, Rosenberg MA. An elementary introduction to Bayesian computing using WinBUGS. Int J Technol Assess Health Care 2001;17(1):98-113.

(11) Gelman A, Rubin DB. Markov chain Monte Carlo methods in biostatistics. Stat Methods Med Res 1996 Dec;5(4):339-55.

(12) Balducci L, Al-Halawani H, Charu V, et al. Elderly cancer patients receiving chemotherapy benefit from first-cycle pegfilgrastim. Oncologist 2007 Dec;12(12):1416-24.

Figure 1: Network of trials for mixed treatment comparison

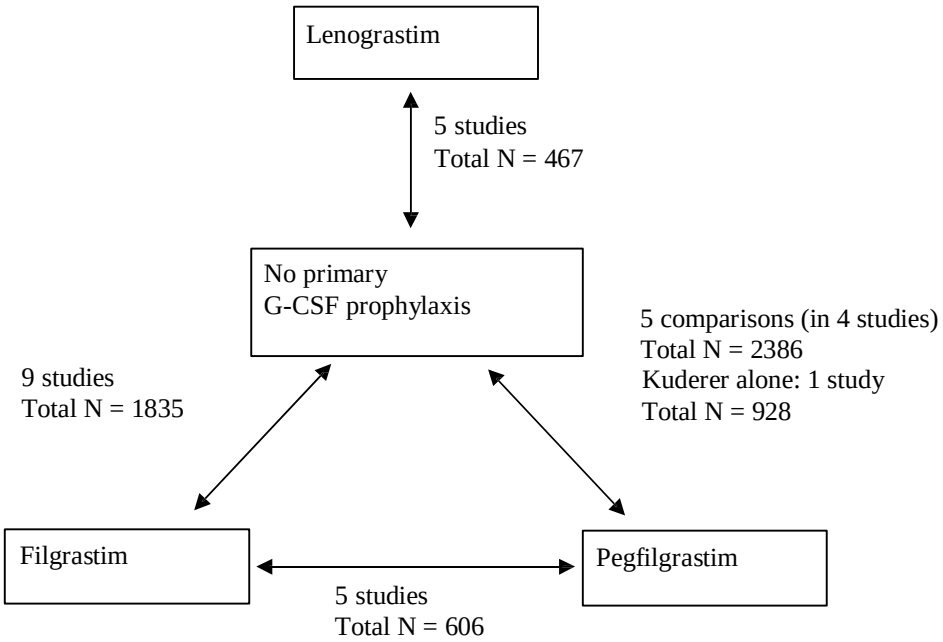

Figure 2: Probability density fwtction for posterior distribution of relative risk of febrile neutropenia wtder treatment with pegfilgrastim vs.lenograstim

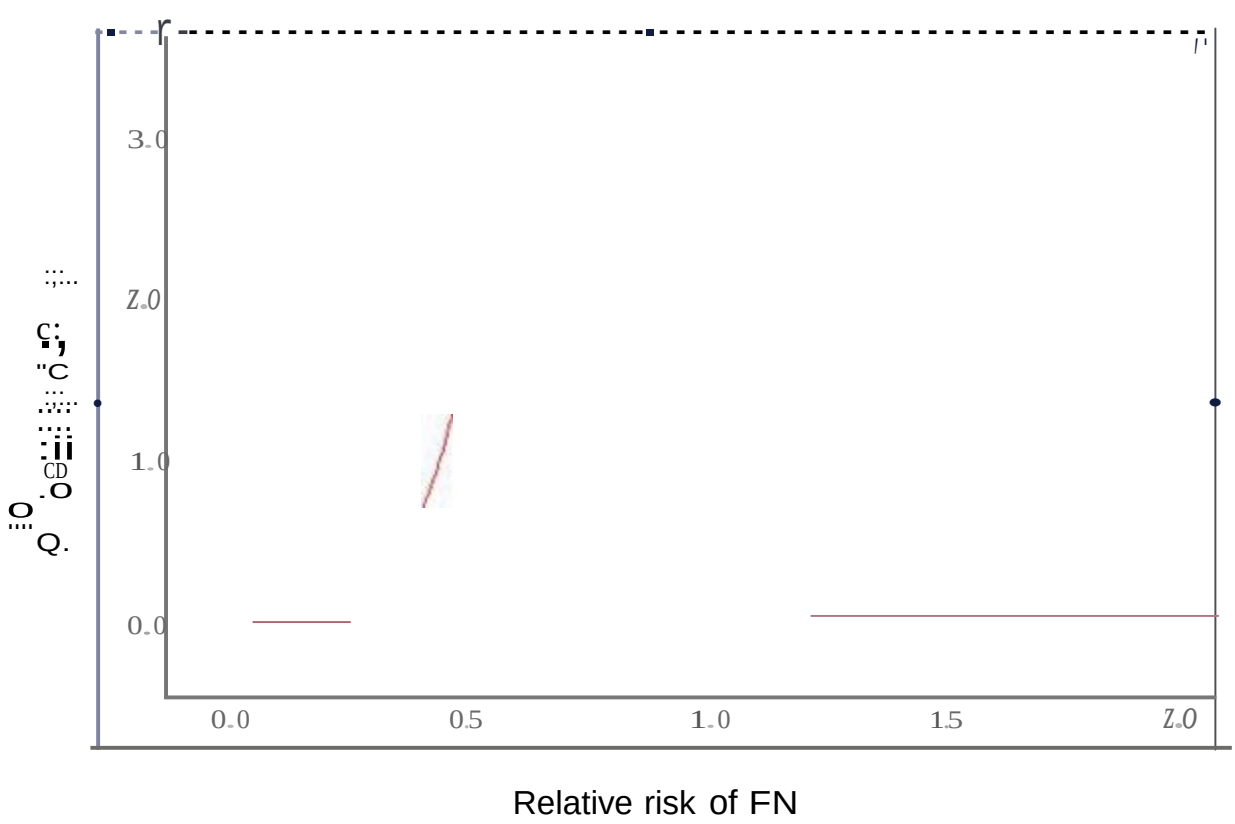

Table 1: Incidence of febrile neutropenia (FN) in all studies of prophylactic G-CSFs

Study	Total N (arm 1)	FN rate (%)		Total N (arm 2)	FN rate (%)		p-value	Relative risk of FN (95% confidence interval)
	Primary pegfilgrastim			**No primary G-CSF**				
Vogel 2005	463	6/463	(1.3%)	465	78/465	(17%)	<0.001	0.08 (0.04-0.18)
*Balducci 2007: solid tumour	343		(4%)	343		(10%)	0.001	0.40 (0.22-0.73)
*Balducci 2007: NHL	73		(15%)	73		(37%)	0.004	0.41 (0.22-0.76)
*Romieu 2007 (cycle 1 only)	30	4/30	(13%)	29	5/29	(17%)		0.77 (0.23-2.62)
*von Minckwitz 2008	314	17/314	(5%)	253	55/253	(22%)	<0.001	0.25 (0.15-0.43)
	Primary filgrastim			**No primary G-CSF**				
Doorduijn 2003	197	72/197	(37%)	192	86/192	(45%)	NR	0.82 (0.64-1.04)
Osby 2003 (C**H**OP)	101		(34%)	104		(50%)	<0.001	0.68 (0.49-0.95)
Osby 2003 (C**N**OP)	125		(32%)	125		(50%)	<0.001	0.64 (0.47-0.87)
Zinzani 1997	77	4/77	(5%)	72	15/72	(21%)	0.004	0.25 (0.09-0.73)
Pettengell 1992	41	9/41	(22%)	39	17/39	(44%)	0.04	0.5 (0.25-0.98)
Timmer-Bonte 2005	90	16/90	(18%)	85	27/85	(32%)	0.01	0.56 (0.33-0.96)
Trillet-Lenoir 1993	65	17/65	(26%)	64	34/64	(53%)	0.002	0.49 (0.31-0.78)
Crawford 1991	95		(40%)	104		(77%)	<0.001	0.52 (0.4-0.68)
Fossa 1998	129	25/129	(19%)	130	38/130	(29%)	NR	0.66 (0.42-1.03)
	Primary lenograstim			**No primary G-CSF**				
Chevallier 1995	61	36/61	(59%)	59	42/59	(71%)	NS	0.83 (0.64-1.08)
Gisselbrecht 1997	82	52/82	(63%)	80	62/80	(78%)	NS	0.82 (0.67-1.0)
Bui 1995 (cycle 1)	22	5/22	(23%)	26	15/26	(58%)	0.02	0.39 (0.17-0.90)
Gebbia 1994	23	5/23	(22%)	28	18/28	(64%)	<0.001	0.34 (0.15-0.77)
Gebbia 1993	43		(12%)	43		(33%)	<0.05	0.36 (0.14-0.90)
	Primary pegfilgrastim			**Primary filgrastim**				
Green 2003	77	10/77	(13%)	75	15/75	(20%)	NS	0.65 (0.31-1.35)
Holmes 2002: phase III	149	14/149	(9%)	147	27/147	(18%)	0.029	0.51 (0.28-0.94)
Holmes 2002: phase II	46	5/46	(11%)	25	2/25	(8%)[#]	NS	1.36 (0.28-6.49)
Grigg 2003	14	0/14	(0%)	13	1/13	(8%)	NR	N/A
Vose 2003 (cycles 1 & 2)	29	6/29	(21%)	31	6/31	(19%)	NS	1.07 (0.39-2.95)

*Studies added as a result of updated search.

Table 2: Estimates of the comparative efficacy of FN treatments using head-to-head meta-analysis and MTC for the evidence base identified in the existing systematic reviews.

		Pre-updated evidence base from existing systematic reviews			
Arm 1	**Arm 2**	**No of studies**	**No of patients**	**a) Head-to-head meta-analysis: relative risk of FN (95% confidence interval), p-value**	**b) Bayesian mixed treatment comparison: relative risk of FN (median, 95% credible interval)**
Pegfilgrastim	No primary G-CSF	1	928	0.08 (0.04-0.18) p<0.001	0.27 (0.12 – 0.60)
Filgrastim	No primary G-CSF	9	1835	0.61 (0.53 – 0.72), p<0.001(3)	0.56 (0.41 - 0.68)
Lenograstim	No primary G-CSF	5	467	0.62 (0.44 – 0.88), p=0.007(3)	0.51 (0.30 – 0.74)
Pegfilgrastim	Filgrastim	5	606	0.64 (0.43 – 0.97), p=0.033(4)	0.49 (0.23 – 1.04)
Pegfilgrastim	Lenograstim	0	0	No direct trials	0.53 (0.22 – 1.37)
Lenograstim	Filgrastim	0	0	No direct trials	0.93 (0.53 – 1.50)

Table 3 Estimates of the comparative efficacy of FN treatments using head-to-head meta-analysis and MTC for the entire evidence base.

Treatment 1	Treatment 2	No of studies	No of patients	a) Head-to-head meta-analysis: relative risk of FN (95% confidence interval), p-value	b) Bayesian MTC**: relative risk of FN (median, 95% credible interval)
Pegfilgrastim	No primary G-CSF	5*	2386	0.290 (0.150 – 0.550), p=0.002	0.341 (0.226 – 0.535)
Filgrastim	No primary G-CSF	9	1835	0.614 (0.525 – 0.718), p<0.001(3)	0.553 (0.433 – 0.671)
Lenograstim	No primary G-CSF	5	467	0.623 (0.442 – 0.879), p=0.007(3)	0.486 (0.278 – 0.714)
Pegfilgrastim	Filgrastim	5	606	0.644 (0.430 – 0.965), p=0.033(4)	0.619 (0.406 – 0.994)
Pegfilgrastim	Lenograstim	0	0	No direct trials	0.710 (0.390 – 1.420)
Lenograstim	Filgrastim	0	0	No direct trials	0.879 (0.491 – 1.407)

*There were 4 publications, but one trial(12) included two populations (lymphoma and solid tumour patients) which are analysed here separately, giving 5 comparisons in total. **MTC=mixed treatment comparison

Table 4: Estimates of the clinical and statistical comparability of G-CSFs

Treatment 1	**Treatment 2**	**Median Relative Risk of FN**	**Probability that Treatment 1 is more efficacious than Treatment 2 in preventing FN (RR > 1.0)**	**Probability that Treatment 1 is clinically superior to Treatment 2 (RR > 1.1)**	**Probability that Treatment 1 and Treatment 2 are clinically equivalent (0.9 < RR < 1.1)**	**Probability that Treatment 1 is clinically inferior to Treatment 2 (RR<0.9)**
Pegfilgrastim	Filgrastim	0.619	98%	95%	4%	1%
Pegfilgrastim	Lenograstim	0.710	86%	79%	13%	8%
Lenograstim	Filgrastim	0.879	71%	54%	30%	16%