Expected health benefits of additional evidence: Technical Appendix

Technical appendix for the Patient-Centered Outcomes Research Institute
December 2012

Karl Claxton,[1,2] Susan Griffin,[2] Hendrik Koffijberg[3] and Claire McKenna[2]

1. Department of Economics and Related Studies, University of York, UK (www.york.ac.uk/economics)
2. Centre for Health Economics, University of York, UK (www.york.ac.uk/inst/che).
3. Julius Centre for Health Sciences and Primary Care, University Medical Centre Utrecht, The Netherlands (www.juliuscentrum.nl/julius).

Contents

Appendix A

Early thrombolysis using streptokinase for the treatment of acute myocardial infarction

Contents

A1. Introduction

Meta-analytical techniques combine the results from multiple studies of the same treatments in the same patients in order to provide a single estimate of effect that takes account of all of the available evidence. Cumulative meta-analysis is the process by which a meta-analysis is updated every time a new trial appears. This allows the results of successive studies to be viewed as a continuum, which can provide insight into trends over time in the estimation of the treatment effect. One time point of interest may be that at which the difference in outcomes between alternative treatments meets criteria for clinical and statistical significance. Furthermore the output from a cumulative meta-analysis can, for example, be compared to clinical practice over time in order to determine when clinical practice matches the indication of the accumulated evidence.

A classic example of such a use of cumulative meta-analysis examines the use of streptokinase as thrombolytic treatment for acute myocardial infarction.(1, 2) Lau et al. updated a meta-analysis(3) that had established that the weight of evidence supported the use of streptokinase for preventing mortality in the treatment of acute myocardial infarction. They presented the results as both traditional and cumulative meta-analyses,(2) and in an additional paper compared the temporal relationship between the accumulating data from the trials and the recommendations of clinical experts.(1)

When a series of studies provide information on the same question they may be suitable for meta-analysis. The process of planning a meta-analysis can be similar to that of an individual trial, and the resulting design should aim to minimise bias.(4) In order for the results of meta-analyses to be meaningful it is important that they be conducted in a scientifically rigorous manner with efforts made to assess and minimise the extent of bias. We do not discuss those methods here, but note that the this case study was not found to exhibit signs of bias.(5, 6)

We now re-examine this cumulative meta-analysis in order to estimate the health impacts of the accumulating evidence from a UK perspective. Background information relevant to the UK setting has been extracted from Boland et al. which examined the use of thrombolysis for the treatment of acute myocardial infarction in a UK setting.(7) The accumulation of evidence on the benefits of intravenous streptokinase in preventing death following acute myocardial infarction is examined from the year 1959 up to 1988 (see Table A3 at end of appendix for study details. Data for the full set of the original trials could not be obtained). The purpose of this case study is to demonstrate how information on health impacts can be derived from a cumulative meta-analysis. As such, we do not take the opportunity to revise the original analysis, either to take account of information subsequently available or to try and incorporate additional comparators or health outcomes.

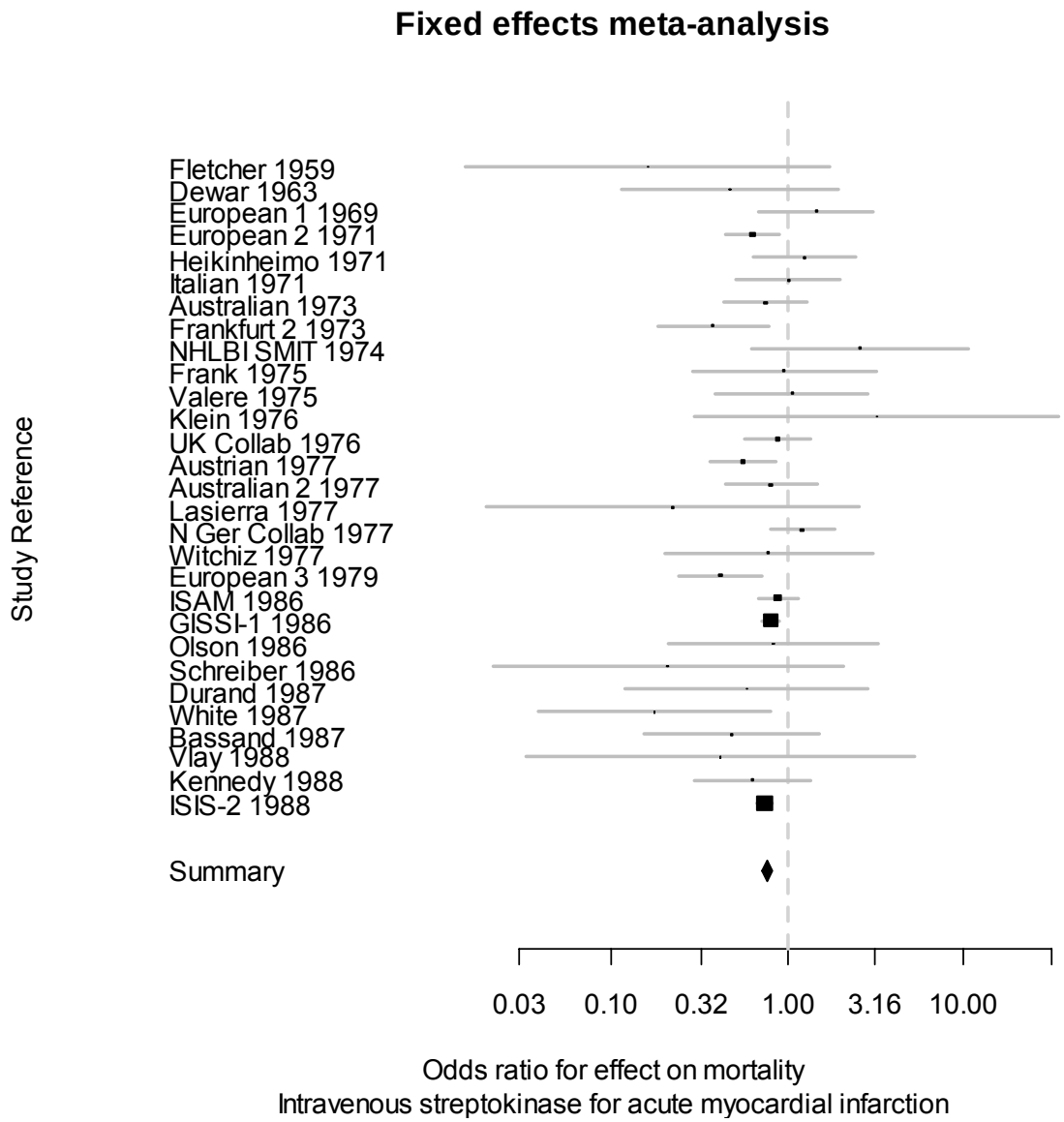
A1.1 Fixed or random effects meta-analysis

In determining which statistical model to use to combine the trial results it is necessary to make a judgement about just how similar are the group of studies. If it is believed that each of the studies should measure the same treatment effect, because for example they are undertaken in the same setting and in groups of patients with the same characteristics, then a fixed effect meta-analysis will be appropriate. The mathematical model underlying a fixed effect meta-analysis assumes that any variation between studies in the observed treatment effect is due to random chance. However, if it is believed that there are some differences between the studies that could cause some variation in the treatment effect being measured then a random effects meta-analysis would be more appropriate. The mathematical model underlying a random effects meta-analysis assumes that between study variation in the observed treatment effect is due to a combination of random chance and of differences between the studies.

While statistical tests can be performed to describe the level of between study variation in outcomes it is best to start with an examination of the study characteristics. The group of studies in this example all address the question of whether streptokinase is beneficial in the treatment of acute myocardial infarction. However, they have some differences, for example, in the dose, mode and timing of administration of streptokinase, the treatment provided to patients in the control arm, the setting for the trial and the

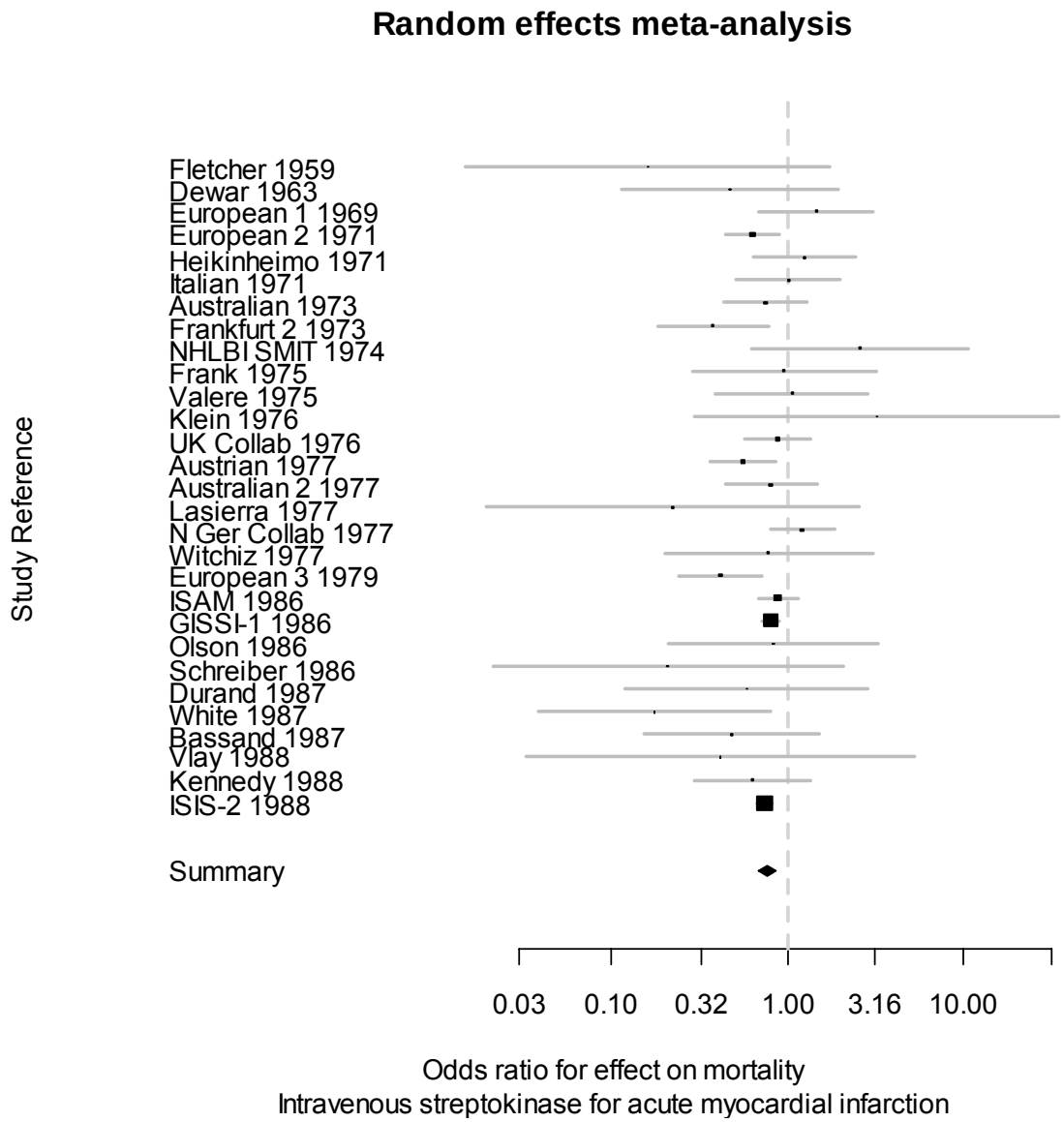
inclusion criteria that determine the patient characteristics.(3) This would suggest that a random effects meta-analysis would be appropriate. Statistical tests such as Cochran's Q and Woolf's test assess how likely is the observed between study variation if any differences are truly only due to random error. Given that assumption, the probability of observing random variation in treatment effects at least as large as that in this group of studies is 9% (the results can be expressed as a p-value of 0.09). Typically the results of such tests would be compared to a statistical significance level in order to know whether the use of a fixed effect approach should be ruled out. At a statistical significance level of 10% the use of a fixed effect approach would be rejected. An alternative approach attempts to quantify the extent of between study variation by estimating the proportion of the total variance that cannot be attributed to random chance. This provides the I^2 statistic, and in this group of studies that produces a value of 27%. The higher the value of the I^2 statistic, the more likely it is that a random effects approach should be used. As a broad rule, values lower than 25% would support the use of a fixed effect meta-analysis.(8)

Throughout this appendix we present the results from both a fixed and random effects approach. The random effects is applied only to the treatment effect and we maintain a fixed effect for the pooled baseline odds. The numbers of deaths and numbers randomised to each arm form the basis for the calculations. All of the meta-analyses are Bayesian with non-informative priors, but equally they could utilise a classical or frequentist framework.(4, 9) All subsequent calculations are based 10,000 samples drawn from the posterior distributions for the pooled baseline odds and odds ratio. Figure A1 shows a forest plot of the included trials, the year in which the results were published, and a summary pooled odds ratio from a fixed effect meta-analysis. Figure A2 shows the same information with the pooled odds ratio from a random effects meta-analysis.



Summary odds ratio 0.77 (95% credible interval 0.72-0.82)

Figure A1. Forest plot with fixed effect meta-analysis

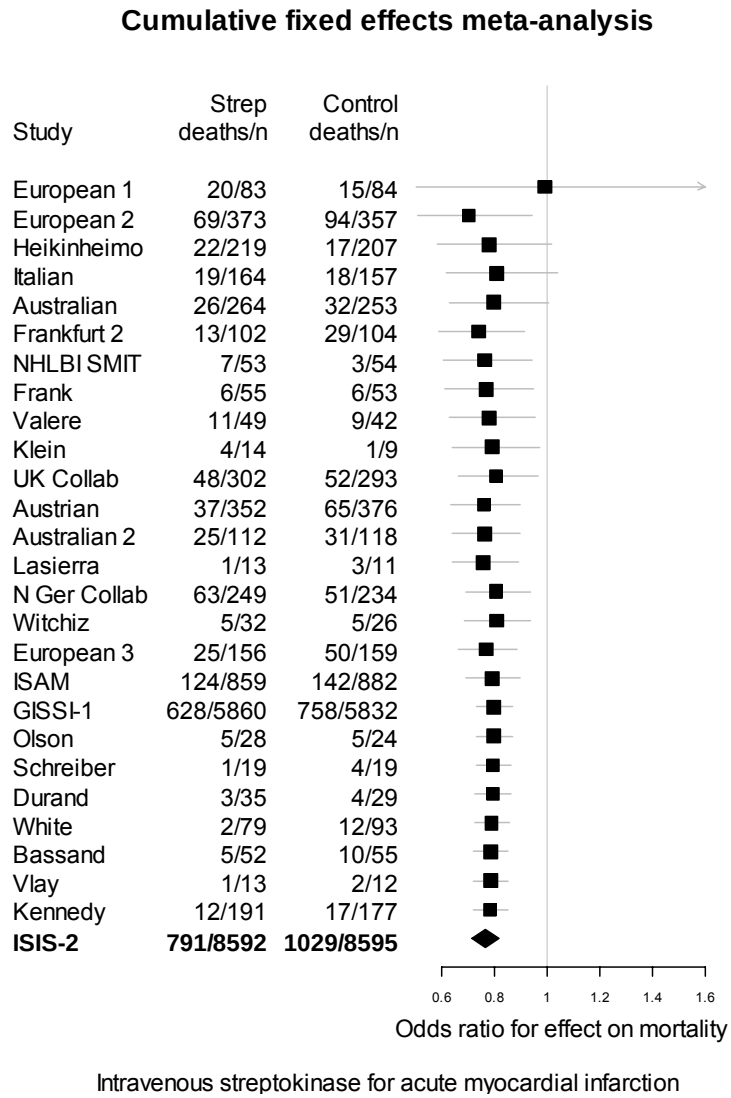


Summary odds ratio 0.76 (95% credible interval 0.67-0.85)

Figure A2. Forest plot with random effects meta-analysis

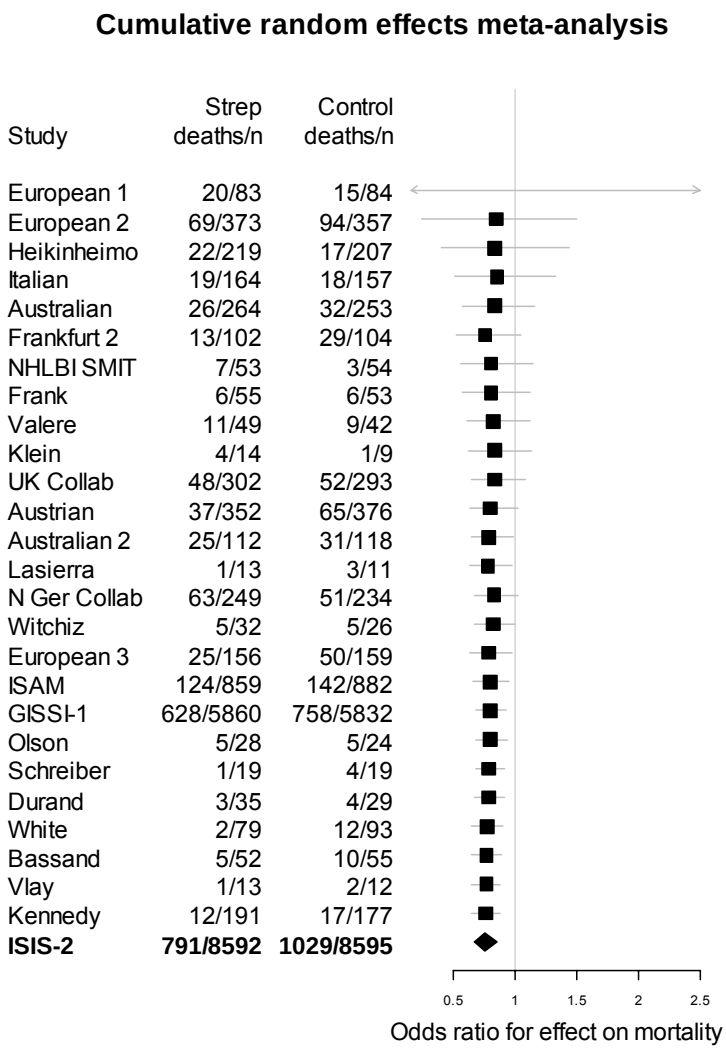
A1.2 Cumulative meta-analysis

Figures A3 and A4 present the same set of studies in the form of a cumulative fixed effect and cumulative random effects meta-analysis respectively. They also show the number of deaths and number of patients randomised (n) to either streptokinase (strep) or control in each of the studies. Each line on the forest plot represents a pooled estimate of effect that includes the results of the current trial along with all the previous trials. The first pooled estimate is provided only once the results of three studies are available. This is because in the random effects model the between study variation is estimated on the basis of observed differences between the trials, and so a minimum of three studies must be included (unless subjective informative priors are to be used).



Summary odds ratio 0.77 (95% credible interval 0.72-0.82)

Figure A3. Forest plot with cumulative fixed effect meta-analysis



Summary odds ratio 0.76 (95% credible interval 0.67-0.85)

Figure A4. Forest plot with cumulative random effects meta-analysis

A2. Information required to estimate absolute health impacts

The meta-analysis provides a summary measure in terms of a relative treatment effect. In order to describe the absolute number of events it is necessary to incorporate additional information on the baseline risk and the number of patients expected to benefit.

The baseline risk of death is taken from the pooled baseline odds from the corresponding meta-analysis. An alternative source of information for the baseline risk could come from observational data in the UK.(7) The population size is assumed to be the number of people who experience acute myocardial infarction and who reach hospital in time to receive thrombolytic treatment. Acute myocardial infarction affects 273,000 people each year in England and Wales. Of these 137,000 die within 30 days and over half (more than 68,500) of these deaths occur prior to the patient reaching hospital or other medical assistance. Therefore we set the annual population size to 204,500 (273,000-68,500).

With this information as the cumulative meta-analysis is updated with each new trial we can estimate the number of deaths per year that would be expected if (i) all patients were not treated with streptokinase (represented by the control arm in the included studies); (ii) all patients were treated with streptokinase; (iii) perfect information allowed all patients to receive the optimal treatment. These estimates are shown in Tables A1 and A2 (corresponding to fixed effect and random effects meta-analyses respectively). By comparing the number of deaths that would be expected under each of these scenarios, it is possible to make some assessments about the value of altering clinical practice (changing implementation) and the value of additional research.

A2.1 Current clinical practice

The value of altering clinical practice depends on what patients currently receive. The counterfactual to the health outcomes that patients might experience as a consequence of further research is the health outcomes they would receive if treated in accordance with clinical practice. Current clinical practice may change over time with or without research and can be assessed retrospectively. Because we perform this cumulative meta-analysis retrospectively we can contrast the results of the accumulating trials with evidence of how clinical practice changed over time. Boland et al. report that practice changed to use of streptokinase after the results of GISSI-1 (1986) and ISIS-2 (1988) were published. If the proportion of eligible patients receiving streptokinase is known at each time point the value of switching from a mixed provision to either one or the other can also be calculated.

However, when assessing the value of conducting a further trial our use of cumulative meta-analysis becomes prospective or proactive. A plan to repeatedly look at updated meta-analysis affects assessments about statistical significance in the classical or frequentist framework because repeated looks increase the possibility of satisfying some pre-specified significance level by chance. We must also look forward and consider how the results might cause clinical practice to change. For example, we may argue that a trial which produced a statistically significant result, and that when added to a cumulative meta-analysis produces a statistically significant result, may persuade clinicians to alter their practice. Tables A1 and A2 make the following the comparisons:

(i) present results against each comparator (streptokinase and control)

(ii) assume current practice switches to providing streptokinase once the odds ratio of streptokinase compared to no streptokinase reaches statistical significance at the 5% level in both the most recent trial and the cumulative meta-analysis. We refer to this as Evidence Based Practice (EBP).

Cross-referencing Figure A1 with A3, and Figure A2 with A4, we can see that based on the rule outlined above, evidence based practice would change following publication of the results of European 2 in 1971 for a fixed effect meta-analysis, and following publication of the results of the results of European 3 in 1979 for a random effects meta-analysis.

Table A1. Number of deaths over one year based on cumulative fixed effect meta-analyses*

Study	Year	Control	Strep	Perfect information	Control v strep	Strep v perfect info	Control v perfect info	Evidence based practice	Practice v perfect info
		(A)	(B)	(C)	(A-B)	(B-C)	(A-C)	u	uB+(1-u)A-C
European 1	1969	46958	44929	41451	2029	3478	5507	0	5507
European 2	1971	52285	39574	39555	12711	19	12730	1	19
Heikinheimo	1971	42596	34726	34668	7870	58	7928	1	58
Italian	1971	38879	32492	32410	6387	82	6469	1	82
Australian	1973	36763	30388	30348	6375	40	6415	1	40
Frankfurt 2	1973	38761	30074	30070	8687	4	8691	1	4
NHLBI SMIT	1974	35905	28556	28550	7349	6	7355	1	6
Frank	1975	34548	27600	27594	6948	6	6954	1	6
Valere	1975	35489	28767	28758	6722	9	6731	1	9
Klein	1976	36127	29646	29632	6481	14	6495	1	14
UK Collab	1976	35946	29998	29987	5948	11	5959	1	11
Austrian	1977	36443	28944	28943	7499	1	7500	1	1
Australian 2	1977	37449	29839	29838	7610	1	7611	1	1
Lasierra	1977	37429	29648	29647	7781	1	7782	1	1
N Ger Collab	1977	37283	31185	31183	6098	2	6100	1	2
Witchiz	1977	37321	31254	31252	6067	2	6069	1	2
European 3	1979	38722	31114	31114	7608	0	7608	1	0
ISAM	1986	38106	31399	31399	6707	0	6707	1	0
GISSI-1	1986	37352	30974	30974	6378	0	6378	1	0
Olson	1986	37588	31150	31150	6438	0	6438	1	0
Schreiber	1986	37101	30699	30699	6402	0	6402	1	0
Durand	1987	36419	30099	30099	6320	0	6320	1	0
White	1987	35542	29154	29154	6388	0	6388	1	0
Bassand	1987	35388	28948	28948	6440	0	6440	1	0
Vlay	1988	34872	28506	28506	6366	0	6366	1	0
Kennedy	1988	34096	27786	27786	6310	0	6310	1	0
ISIS-2	1988	34010	27181	27181	6829	0	6829	1	0

*minimum 3 studies included in meta-analysis so Fletcher and Dewar incorporated in row 1 with European 1

At each row in the table the odds of death with streptokinase and control are calculated from the cumulative meta-analysis of the current trial with all previous trials. The Bayesian meta-analysis undertaken in WinBUGS provides 10,000 estimates from the posterior distributions for the baseline odds and the odds ratio.(10) The baseline odds is the odds of death for a patient receiving control and the baseline odds multiplied by the odds ratio is the odds of death for a patient receiving streptokinase. These odds are converted to probabilities and multiplied by the size of the annual eligible population in order to calculate the number of deaths expected over one year. To calculate the number of deaths expected with perfect information we take the minimum of the odds of death among control and streptokinase from each of the 10,000 paired estimates and multiply by the size of the eligible population.

As streptokinase is always expected to produce fewer deaths than control, the comparison of control to streptokinase (A-B) shows the number of deaths that could be avoided by implementing streptokinase. The comparison of streptokinase to perfect info (B-C) shows the number of deaths that could be avoided by gathering further information on the effectiveness of streptokinase. The comparison of control to perfect info (A-C) shows the number of deaths that could be avoided by implementing streptokinase and by gathering further information on the effectiveness of streptokinase.

Table A2. Number of deaths over one year based on cumulative random effects meta-analyses*

Study	Year	Control	Strep	Perfect information	Control v strep	Strep v perfect info	Control v perfect info	Evidence based practice	Utilisation v perfect info
		(A)	(B)	(C)	(A-B)	(B-C)	(A-C)	u	uB+(1-u)A-C
European 1	1969	51311	43702	37438	7609	6264	13873	0	13873
European 2	1971	52163	40122	38634	12041	1488	13529	0	13529
Heikinheimo	1971	42119	35371	34177	6748	1194	7942	0	7942
Italian	1971	38178	33079	32273	5099	806	5905	0	5905
Australian	1973	36282	30793	30428	5489	365	5854	0	5854
Frankfurt 2	1973	38463	30237	30102	8226	135	8361	0	8361
NHLBI SMIT	1974	35197	29059	28753	6138	306	6444	0	6444
Frank	1975	33909	27947	27719	5962	228	6190	0	6190
Valere	1975	34803	29266	29046	5537	220	5757	0	5757
Klein	1976	35369	30082	29815	5287	267	5554	0	5554
UK Collab	1976	35527	30303	30141	5224	162	5386	0	5386
Austrian	1977	35770	29349	29269	6421	80	6501	0	6501
Australian 2	1977	36905	30172	30133	6733	39	6772	0	6772
Lasierra	1977	36939	29928	29902	7011	26	7037	0	7037
N Ger Collab	1977	36892	31395	31327	5497	68	5565	0	5565
Witchiz	1977	37015	31319	31280	5696	39	5735	0	5735
European 3	1979	38337	31213	31186	7124	27	7151	1	0
ISAM	1986	37953	31328	31314	6625	14	6639	1	0
GISSI-1	1986	37331	30790	30782	6541	8	6549	1	0
Olson	1986	37521	30997	30991	6524	6	6530	1	0
Schreiber	1986	37166	30465	30462	6701	3	6704	1	0
Durand	1987	36512	29893	29889	6619	4	6623	1	0
White	1987	35816	28789	28785	7027	4	7031	1	0
Bassand	1987	35606	28507	28506	7099	1	7100	1	0
Vlay	1988	35203	28132	28131	7071	1	7072	1	0
Kennedy	1988	34496	27351	27351	7145	0	7145	1	0
ISIS-2	1988	34099	27021	27021	7078	0	7078	1	0

*minimum 3 studies included in meta-analysis so Fletcher and Dewar incorporated in row 1 with European 1

Figures A5 and A6 show the relationship between the comparison of value of implementation (changing practice to the treatment that is expected to produce least deaths based on current evidence, which is always streptokinase) and the value of information (reducing the consequences of uncertainty around which treatment will produce least deaths). The numbers are taken from Tables A1 and A2.

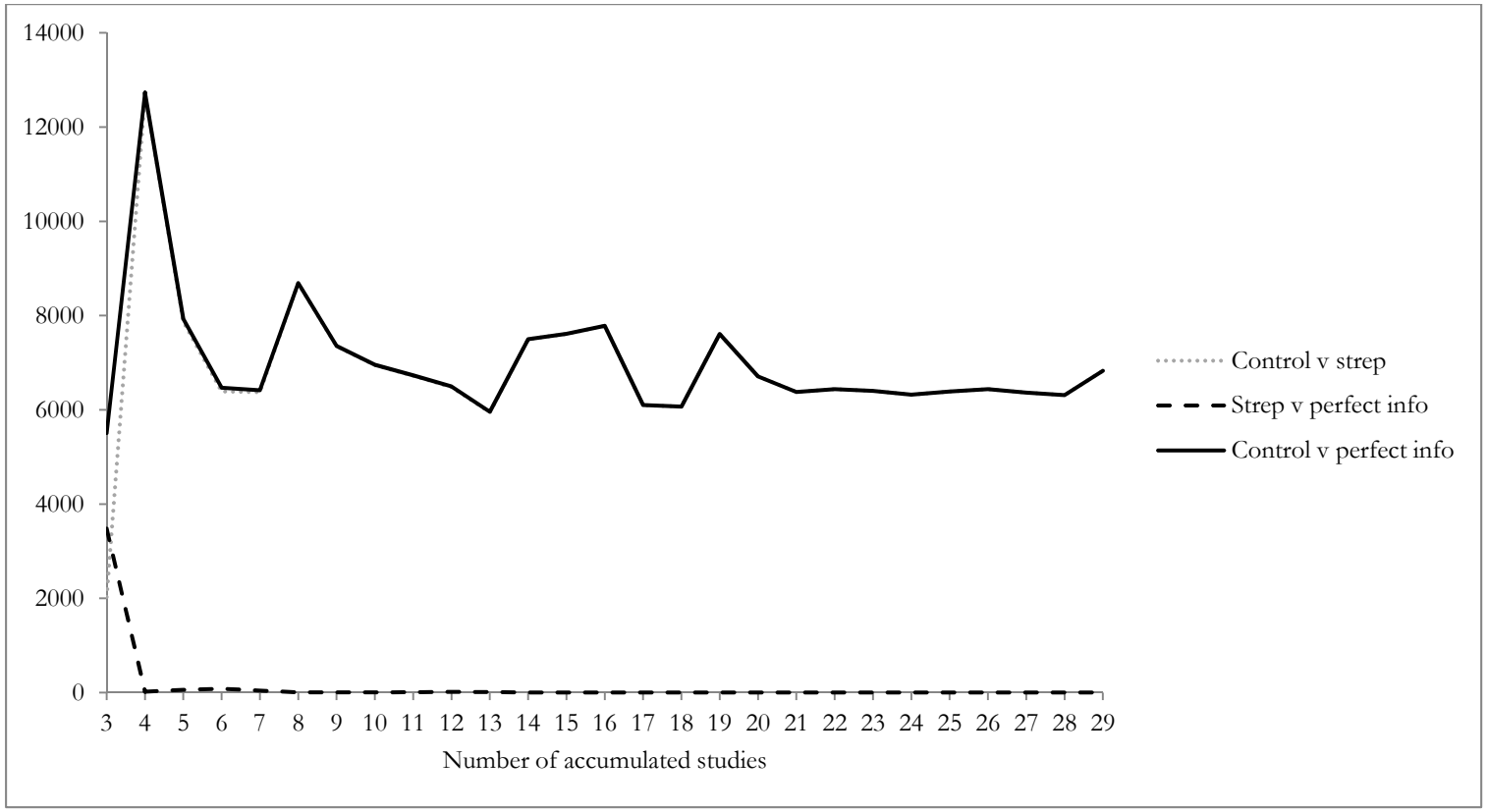


Figure A5. Value of implementation and value of perfect information fixed effect meta-analysis

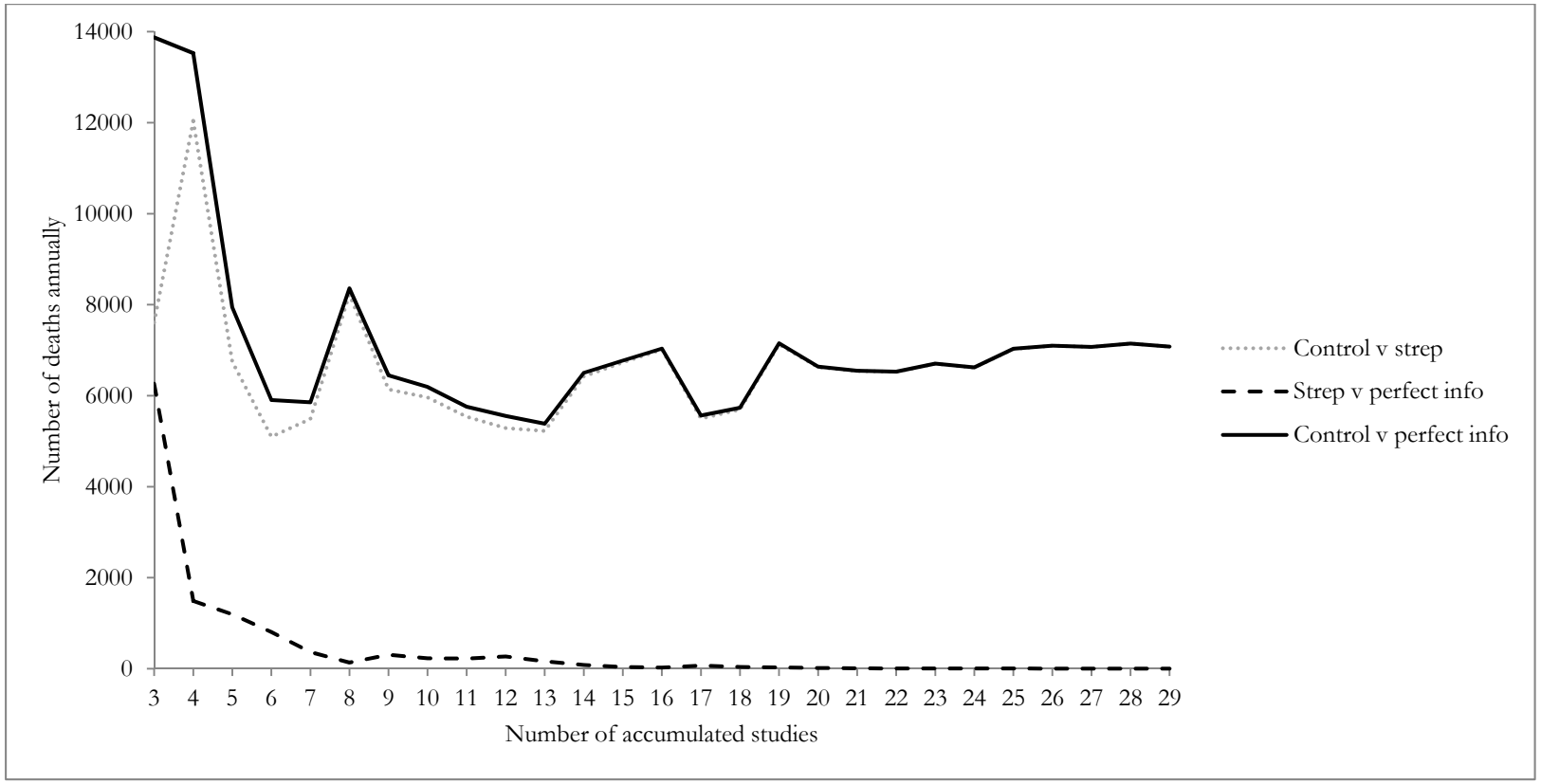


Figure A6. Value of implementation and value of perfect information random effects meta-analysis

Providing streptokinase is always expected to result in fewer deaths than providing control (no streptokinase) across the whole range of accumulated studies. This figure shows the value of switching from no streptokinase to streptokinase (Control v strep) and the value of perfect information (Strep v perfect info). The value of perfect information compared to streptokinase indicates the value of reducing uncertainty in whether streptokinase would result in the fewest deaths. The value of perfect information compared to control indicates the value of switching implementation from control to streptokinase and the value of reducing uncertainty in whether streptokinase is the most effective treatment option. Visually it can be seen that the line 'Control v perfect info' is the summation of the line 'Control v strep' and 'Strep v perfect info'. There is greater uncertainty for longer in the random effects meta-analysis compared to the fixed effect meta-analysis. The value of implementation is broadly similar between the fixed effect and random effects meta-analyses.

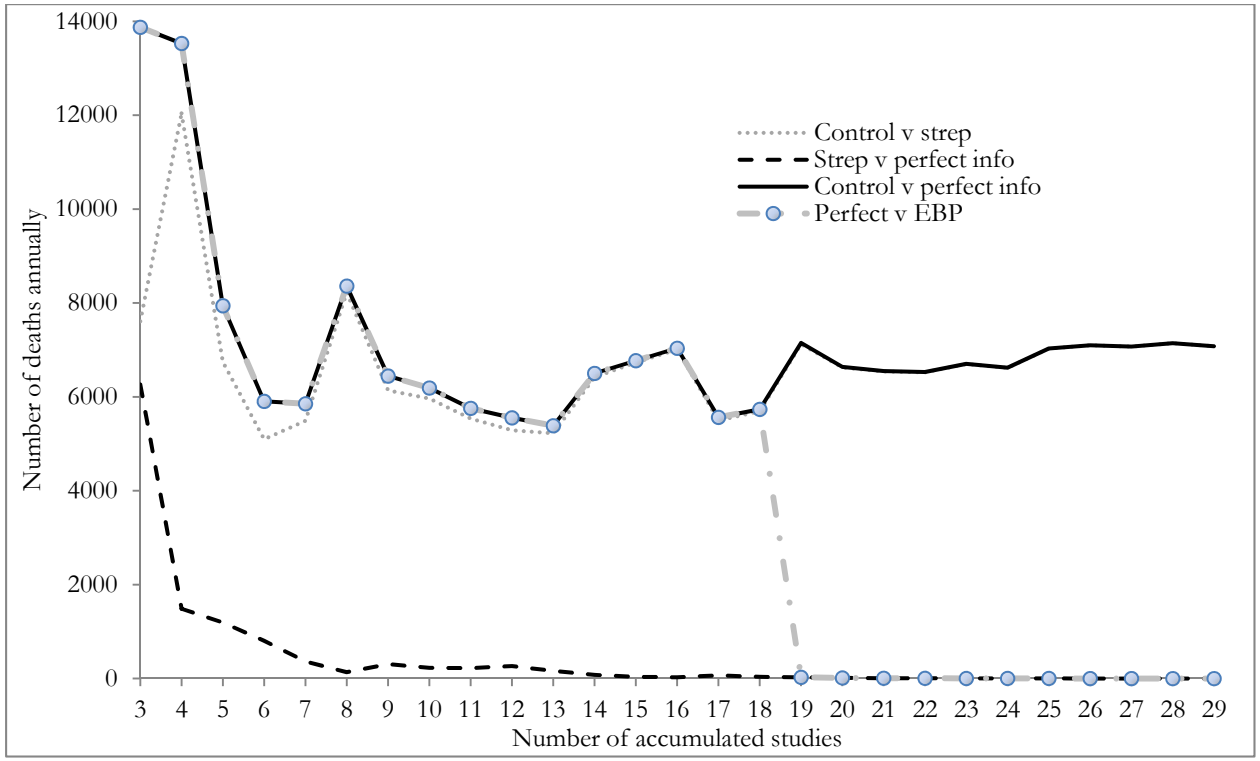


Figure A7. Value of implementation and value of perfect information random effects meta-analysis

Figure A7 adds an additional line to Figure A6 that compares the value of providing the treatment indicated by perfect information to the value of providing the treatment suggested by the evidence based practice (EBP) rule of switching to streptokinase once the odds ratio from the current trial and the pooled odds ratio from the concurrent cumulative meta-analysis are statistically significant at the 5% level. As discussed earlier, EBP would change from control to streptokinase with the publication of the results of European 3, which is trial number 19.

Tables A1 and A2 show the total expected consequences of uncertainty throughout the cumulative meta-analysis (the number of deaths expected with perfect information compared to the number expected with current information). In Figures A8 and A9 the total expected consequences are broken down at three time points in order to show the underlying distribution. The total probability that current treatment is not the most effective is estimated by the proportion of the 10,000 samples in which the alternative treatment results in fewer deaths. For that proportion of samples where the current treatment is the most effective there are zero excess deaths. The histograms show the distribution of excess deaths for perfect information compared to current treatment as control (Figures A8 and A9), and current treatment as streptokinase (Figures A10 and A11).

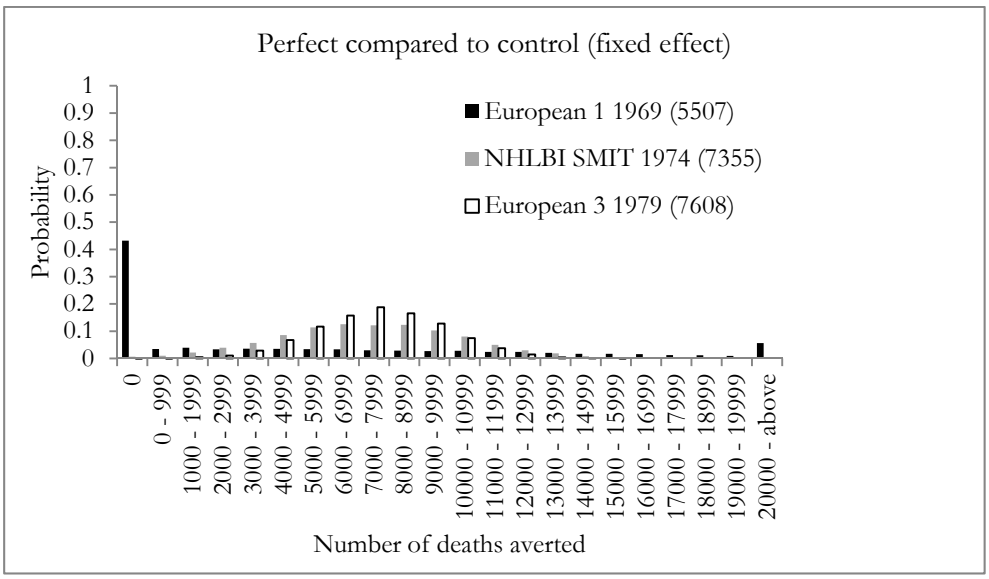


Figure A8. Histogram of consequences of uncertainty for three time points.

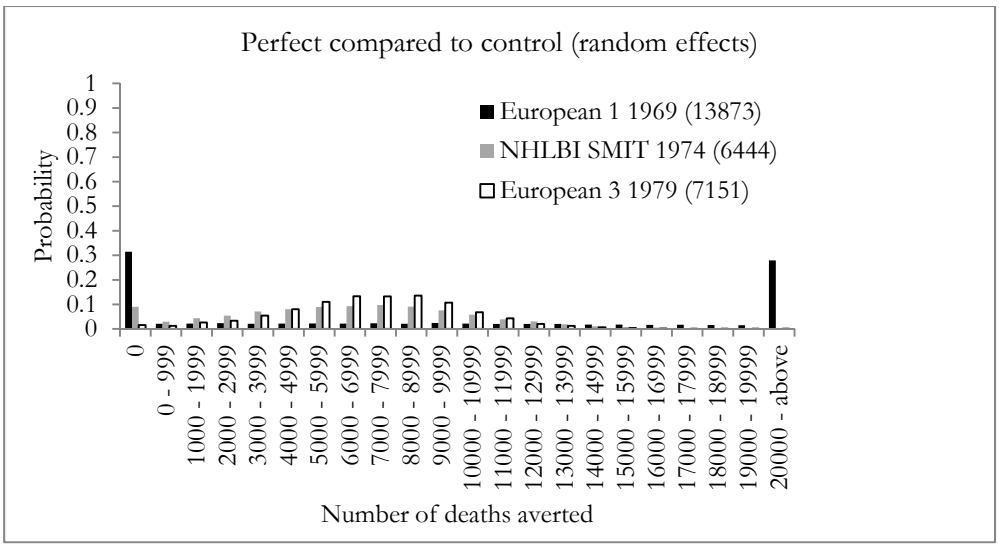


Figure A9. Histogram of consequences of uncertainty for three time points.

Figures A8 and A9 show the distribution of the number of deaths averted by providing the treatment indicated by perfect information instead of utilising control. The final category of '20000-above' is the summation of small probabilities (< 0.001) of observing consequences of uncertainty in terms of numbers of deaths averted in brackets of width 1000 up to a maximum of 51,392 (fixed effect) and 90,401 (random

effects) following the publication of European 1 in 1969. The numbers in brackets in the legend next to study name show the total expected number of deaths averted with perfect information relative to control. Figures A8 and A9 show the value of switching implementation from control to streptokinase and reducing uncertainty in the effectiveness of streptokinase.

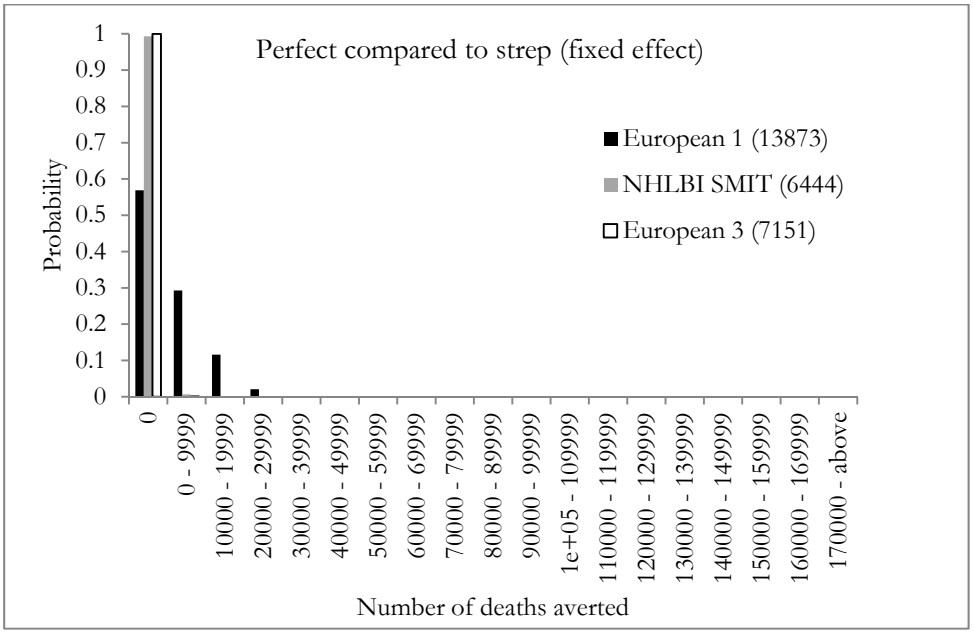


Figure A10. Histogram of consequences of uncertainty for three time points.

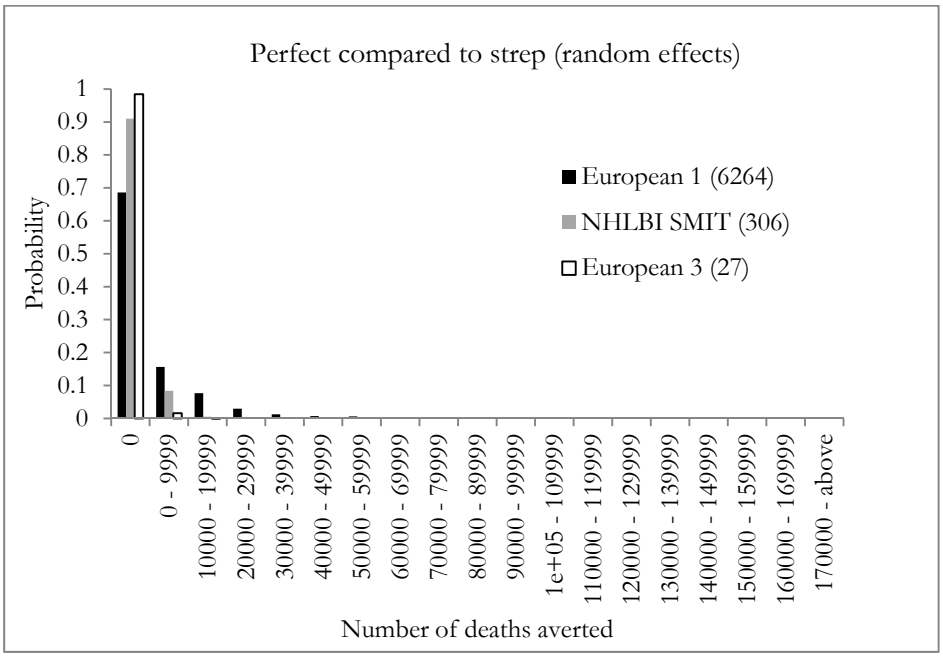


Figure A11. Histogram of consequences of uncertainty for three time points.

Figures A10 and A11 show the distribution of the number of deaths averted by providing the treatment indicated by perfect information instead of utilising streptokinase. Numbers in brackets in the legend next to study name show the expected number of deaths averted with perfect information relative to using streptokinase. Because streptokinase is the treatment expected to produce fewest deaths at each of

the timepoints considered, Figures A10 and A11 show the distribution of the value of reducing uncertainty in the effectiveness of streptokinase at each of those timepoints. In the random effects meta-analysis of the first three studies including European 1 there is very large variance in the pooled odds ratio and hence a minute possibility of consequences up to 170,000 deaths per year.

A2.2 Minimum clinical difference

The calculations so far have estimated the number of deaths with perfect information by taking the minimum odds of death in each of the 10,000 samples. Incorporating a minimum clinical difference in the calculation means that the minimum odds are selected only if they are lower than the odds associated with current practice by some specified amount. Figures A12 and A13 consider a range of minimum clinical differences in the absolute odds of death. They show how the reduction in number of deaths with perfect information relative to current information falls if we only switch practice from control to streptokinase once the improvement in odds of death exceeds ever larger minimum clinical differences.

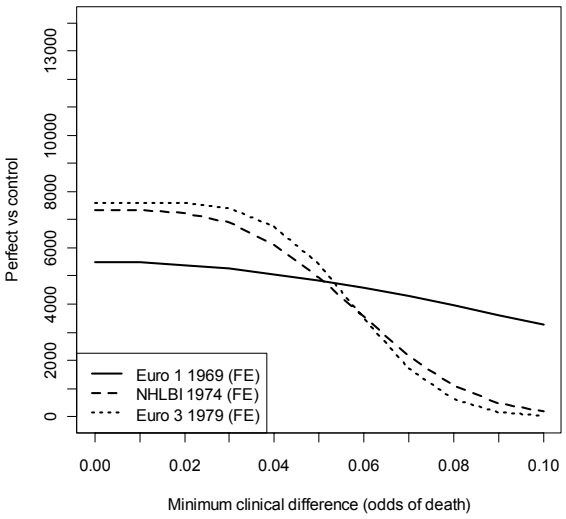


Figure A12. Deaths averted by perfect information compared to control for range of minimum clinical differences (fixed effect meta-analysis)

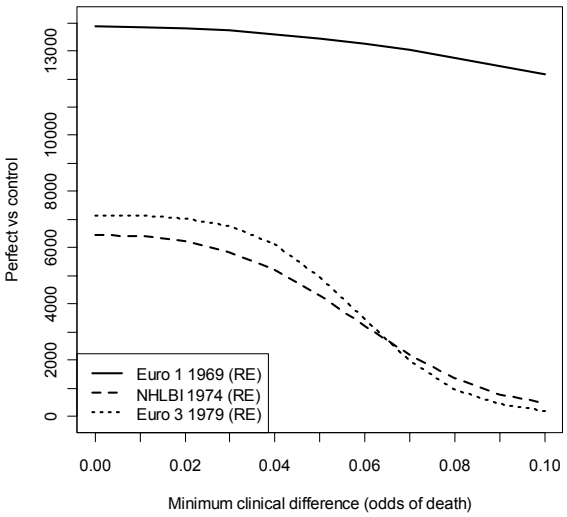


Figure A13. Deaths averted by perfect information compared to control for range of minimal clinical differences (random effects meta-analysis)

A minimum clinical difference means that practice changes to streptokinase only if perfect information shows that the odds of death are reduced by more than the minimum difference when compared to the odds of death on control. This minimum clinical difference in the odds of death can also be expressed as a minimum reduction in the expected number of deaths per year. The baseline odds of death from which we calculate the minimum reduction is updated with each successive study in the cumulative meta-analysis. Consequently the minimum number of deaths implied by any given reduction in the odds of death varies as evidence accumulates. Figures A14 and A15 show how the minimum clinical difference in the odds of death translates into minimum reductions in the number of deaths at three time points in the fixed and random effects cumulative meta-analyses.

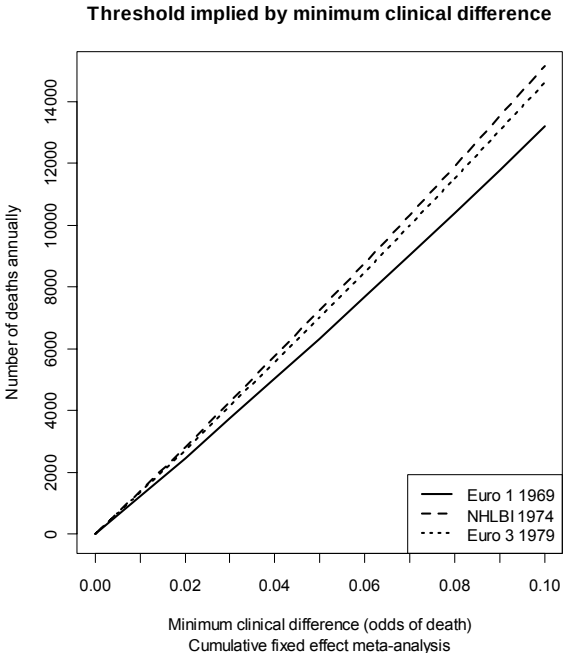
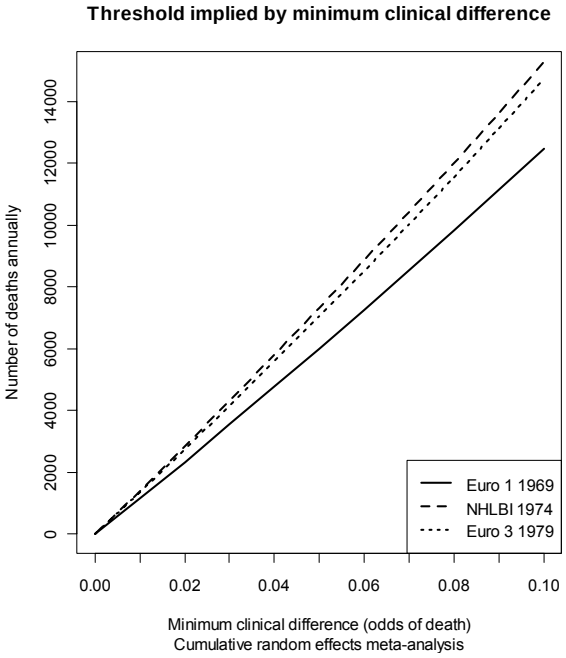


Figure A14. Minimum reduction in number of deaths implied by minimum clinical difference (fixed effect meta-analysis)



A15. Minimum reduction in number of deaths implied by minimum clinical difference (random effects meta-analysis)

A2.3 Health impacts for patients enrolled in clinical trials

Figure A16 repeats the cumulative fixed effect meta-analysis alongside additional information in each row. The first two columns after the study name are calculated using the evidence available prior to the current study and describe: (i) whether streptokinase would be supported by the evidence based practice (EBP) rule; (ii) the number of deaths expected in patients randomised within the trial compared the number of deaths expected had they received EBP. The third and fourth columns are calculated once the current study is incorporated into the cumulative meta-analysis and they describe: (iii) the value of changing implementation from streptokinase to control; and (iv) the value of reducing uncertainty in the effectiveness of streptokinase.

Study	Strep in use before study	Deaths enrolled v EBP	Value imp after study strep v control	Value perfect after study perfect v strep
European 1				
European 2	0	-3.7	0	19
Heikinheimo	1	12.9	0	59
Italian	1	6	0	82
Australian	1	7.9	0	40
Frankfurt 2	1	3.2	0	4
NHLBI SMIT	1	2.3	0	6
Frank	1	1.9	0	6
Valere	1	1.4	0	9
Klein	1	0.3	0	14
UK Collab	1	9.3	0	11
Austrian	1	10.9	0	1
Australian 2	1	4.3	0	0
Lasierra	1	0.4	0	0
N Ger Collab	1	8.9	0	2
Witchiz	1	0.8	0	2
European 3	1	4.7	0	0
ISAM	1	32.8	0	0
GISSI-1	1	191.3	0	0
Olson	1	0.7	0	0
Schreiber	1	0.6	0	0
Durand	1	0.9	0	0
White	1	2.9	0	0
Bassand	1	1.7	0	0
Vlay	1	0.4	0	0
Kennedy	1	5.5	0	0
ISIS-2	**1**	**265.2**	**0**	**0**

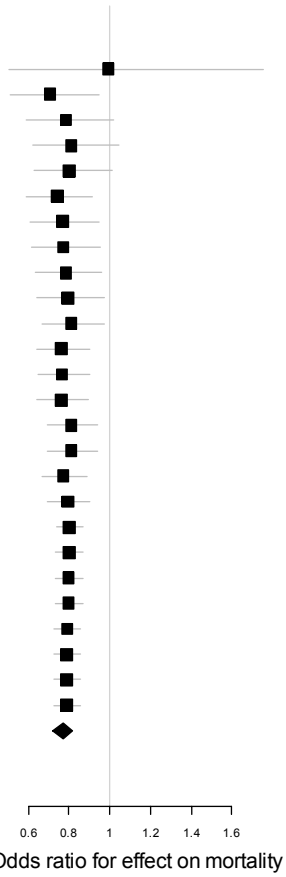


Figure A16. Health effects of next study based on fixed effect cumulative meta-analysis

Streptokinase is always expected to produce fewer deaths than control (no streptokinase). When streptokinase is not used in all patients, those who are randomised to the streptokinase arm of the next study have a lower expected number of deaths compared to current practice: each study is expected to result in a health gain. If streptokinase is used in current practice, those randomised to the control arm in the next study have a higher expected number of deaths compared to current practice: each study is expected to result in a health loss.

The EBP use of streptokinase assumes that practice switches to streptokinase once a statistically significant (at 5%) reduction is observed in the odds ratio from the most recent study and in the pooled odds ratio when it is incorporated into a cumulative meta-analysis of all preceding studies.

The value of implementation (streptokinase versus control) after the study shows the value of future efforts to switch practice to use streptokinase instead of control. The value of perfect information compared to streptokinase shows the maximum potential value from a new study in reducing uncertainty as to whether streptokinase is the best treatment. Figure A17 repeats Figure A16 for a random effects meta-analysis.

Study	Strep in use before study	Dead enrolled cf practice	Value imp after study strep v control	Value perfect after study perfect v strep
European 1				
European 2	0	-13.9	12041	1488
Heikinheimo	0	-12.9	6748	1194
Italian	0	-5.4	5098	806
Australian	0	-6.6	5489	365
Frankfurt 2	0	-2.7	8226	136
NHLBI SMIT	0	-2.1	6138	307
Frank	0	-1.7	5963	228
Valere	0	-1.4	5537	220
Klein	0	-0.4	5287	268
UK Collab	0	-7.8	5223	162
Austrian	0	-9	6422	79
Australian 2	0	-3.5	6733	38
Lasierra	0	-0.4	7011	25
N Ger Collab	0	-8.5	5498	68
Witchiz	0	-0.9	5697	38
European 3	0	-4.3	0	27
ISAM	1	30.7	0	14
GISSI-1	1	188.9	0	8
Olson	1	0.8	0	6
Schreiber	1	0.6	0	3
Durand	1	1	0	4
White	1	3	0	3
Bassand	1	1.9	0	1
Vlay	1	0.4	0	1
Kennedy	1	6.1	0	0
ISIS-2	**1**	**300.3**	**0**	**0**

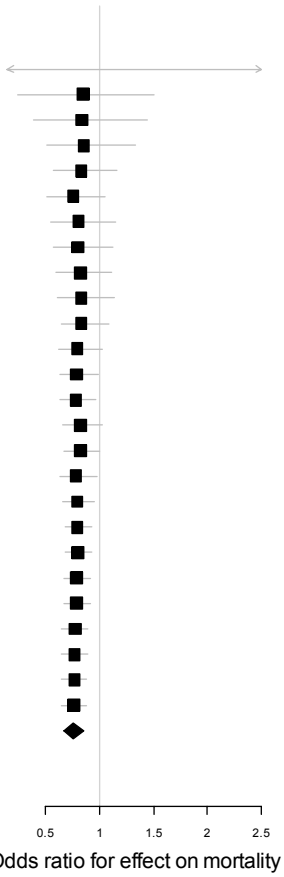


Figure A17. Health costs for patients enrolled in next study based on cumulative random effects meta-analysis of prior studies

The health costs can for patients enrolled in the next study can also be calculated by comparing the number of deaths expected within the trial to the number of deaths expected if current practice was to utilise the most effective treatment indicated by the preceding cumulative meta-analysis. The most effective treatment is always expected to be streptokinase. Figure A18 shows the expected health costs for patients enrolled in the next study if it is assumed they would otherwise have received streptokinase. This is shown alongside the value of perfect information relative to providing streptokinase that was predicted before the study is accumulated in the meta-analysis. Larger trials conducted at a time at which the value of perfect information had declined such as GISSI-1 (number 21) and ISIS-2 (number 29) were expected to produce greater harm by randomising patients to an ineffective treatment than any health gain from reducing uncertainty in the effectiveness of streptokinase.

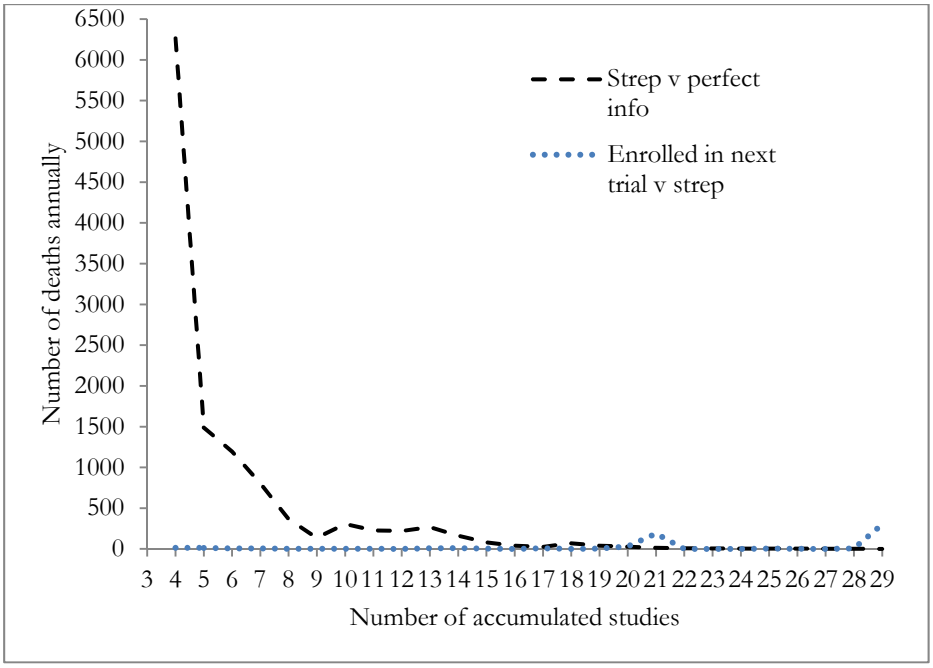


Figure A18. Health costs and perfect information based on cumulative random effects meta-analysis

A3. Cumulative meta-analysis by year based on number of deaths

Figure A19 shows how the accumulating evidence affects the uncertainty in the reduction in number of deaths per year from switching practice to streptokinase, but the trials are grouped by year instead of accumulating individually. Rather than plotting the pooled odds ratio and its associated confidence interval it shows the distribution of the expected reduction in the number of deaths (based on the fixed effect meta-analysis). The mean expected reduction in deaths is plotted along with percentiles from the distribution. The more extreme percentiles are drawn as fainter lines to demonstrate that they are less probable.

Figure A18 illustrates that the expected (mean) reduction in number of deaths by switching practice to streptokinase very soon stabilises to around 6-7000 deaths annually, as shown in Table A1. A decision taken in 1969 to use streptokinase would be expected to benefit patients (saving 2,029 lives), but there is a 10% chance that it would actually harm patients to the extent of a loss of more than 12,183 lives. Grouping all the studies published in 1971 together (European 1, Heikinheimo, Italian) means that the 97.5th percentile does not exclude the possibility of harm. When grouping by year the benefits of streptokinase achieve statistical significance at the 5% level at the end of 1973 (note no studies were published in 1972). By 1986 the publication of the very large GISSI-1 and two very small studies results in an estimate of zero probability of harm from switching to provide streptokinase. So while decisions taken in earlier years are expected to produce on average the same reduction in the number of deaths per year as decisions taken in later years, we can see that the potential consequences of error are reduced for later decisions as a result of the additional information accumulated from trials.

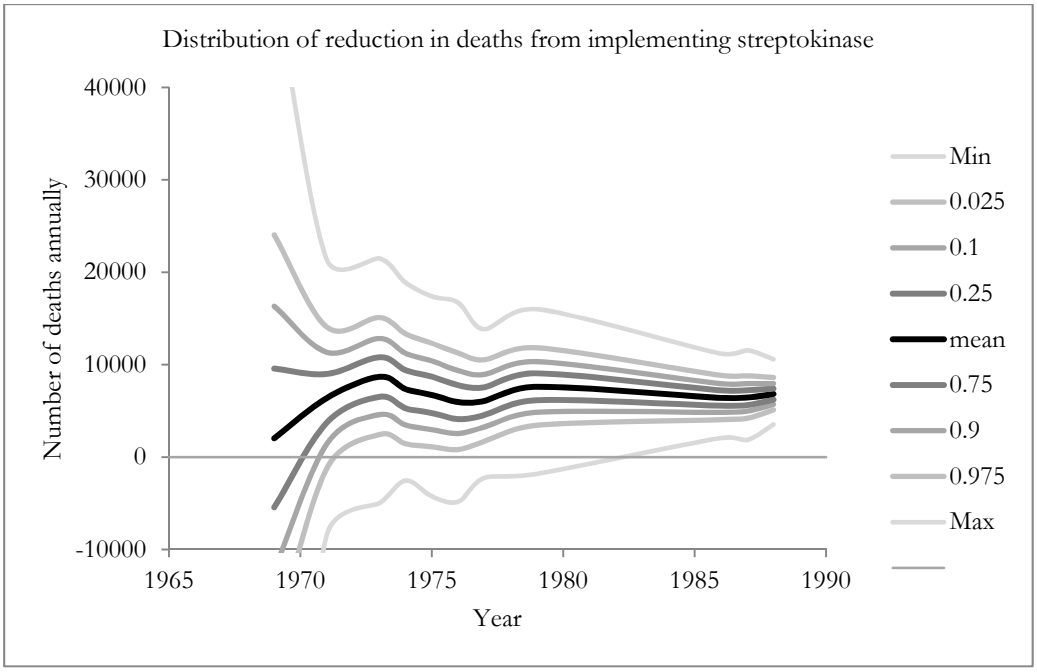


Figure A19. How accumulating evidence affects uncertainty in the treatment decision (fixed effect)

Figure A20 repeats Figure A19 for the cumulative random effects meta-analysis. The uncertainty is greater in the random effects meta-analysis than the fixed effect meta-analysis as it allows for between study variation. A decision taken in 1969 to use streptokinase would be expected to benefit patients (saving 7,609 lives), but there is a 10% chance that it would actually harm patients to the extent of a loss of more than 16,559 lives.

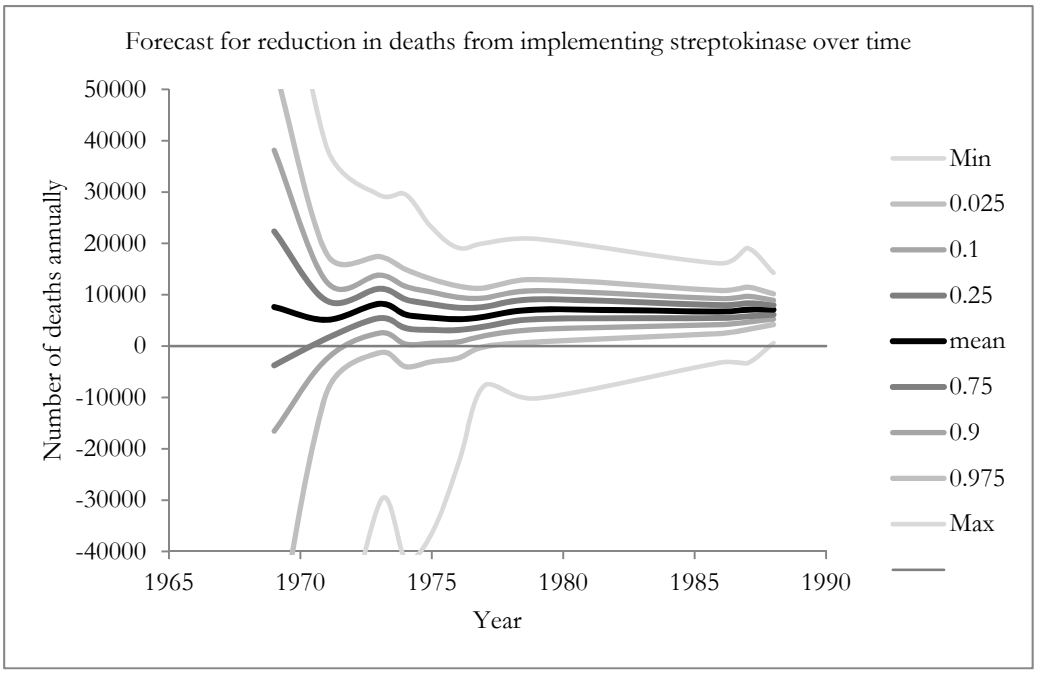


Figure A20. How accumulating evidence affects uncertainty in the treatment decision (random effects)

Table A3. Data used in analysis

		Strep		Control				
Trial	Date	Deaths	Treated	Deaths	Treated	N	Include?	Source
Fletcher	1959	1	12	4	11	23	Y	Yusuf et al 1985
Dewar	1963	4	21	7	21	42	Y	Yusuf et al 1985
European 1	1969	20	83	15	84	167	Y	Yusuf et al 1985
European 2	1971	69	373	94	357	730	Y	Yusuf et al 1985
Heikinheimo	1971	22	219	17	207	426	Y	Yusuf et al 1985
Italian	1971	19	164	18	157	321	Y	Yusuf et al 1985
Australian	1973	26	264	32	253	517	Y	Abstract: Bett et al. Lancet 1973;301(7794)57-60
Frankfurt 2	1973	13	102	29	104	206	Y	Yusuf et al 1985
NHLBI SMIT	1974	7	53	3	54	107	Y	Yusuf et al 1985
Frank	1975	6	55	6	53	108	Y	Yusuf et al 1985
Valere	1975	11	49	9	42	91	Y	Yusuf et al 1985
Klein	1976	4	14	1	9	23	Y	Yusuf et al 1985
UK Collab	1976	48	302	52	293	595	Y	Yusuf et al 1985
Austrian	1977	37	352	65	376	728	Y	Yusuf et al 1985
Australian 2	1977	25	112	31	118	230	Y	Calculated from Yusuf (combined Australian) and Bett
Lasierra	1977	1	13	3	11	24	Y	Yusuf et al 1985
N Ger Collab	1977	63	249	51	234	483	Y	Yusuf et al 1985
Witchiz	1977	5	32	5	26	58	Y	Yusuf et al 1985
European 3	1979	25	156	50	159	315	Y	Yusuf et al 1985
ISAM	1986	124	859	142	882	1741	Y	Abstract: Schroder et al. Journal of the American College of Cardiology 1987;9(1)197-203
GISSI-1	1986	628	5860	758	5832	11712	Y	Abstract: Lancet 1986;1(8478):397-402 and web results
Olson	1986	5	28	5	24	52	Y	Yusuf et al 1985
Baroffio	1986						N	In Italian - results/abstract not found yet
Schreiber	1986	1	19	4	19	38	Y	Yusuf et al 1985
Cribier	1986						N	Abstract n/a: Haemostasis 1986;16 (Suppl 3): 122-129
Sainsous	1986						N	Abstract n/a: Haemostasis 1986;16 (Suppl 3): 140-147
Durand	1987	3	35	4	29	64	Y	Article: Clin Cardiol 1987; 10:383-392
White	1987	2	79	12	93	219*	Y	Abstract: NEJM 1987;317:850-855
Bassand	1987	5	52	10	55	107	Y	Abstract: American Journal of Cardiology 1987;60(7):435-439
Vlay	1988	1	13	2	12	25	Y	Article: Chest 1988;93(4):716-721
Kennedy	1988	12	191	17	177	368	Y	Article: Circulation 1988;77:345-352
ISIS-2	1988	791	8592	1029	8595	17187	Y	Abstract: Lancet 1988;2(8607):349-360
Wisenberg	1988		41		25		N	Info not in abstract: American Journal of Cardiology 1988; 62(16)1011-1016

*Deaths not reported for full N

References

1. Antman EM, Lau J, Kupelnick B, Mosteller F, Chalmers TC. A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts: Treatments for myocardial infarction. JAMA: The Journal of the American Medical Association. 1992;268(2):240-8.
2. Lau J, Antman EM, Jimenez-Silva J, Kupelnick B, Mosteller F, Chalmers TC. Cumulative meta-analysis of therapeutic trials for myocardial infarction. The New England journal of medicine. 1992;327(4):248-54. Epub 1992/07/23.
3. Yusuf S, Teo K, Woods K. Intravenous magnesium in acute myocardial infarction. An effective, safe, simple, and inexpensive intervention. Circulation. 1993;87(6):2043-6.
4. Whitehead A, Wiley J. Meta-analysis of controlled clinical trials: John Wiley & Sons West Sussex,, UK; 2002.
5. Egger M, Smith GD. Misleading meta-analysis. BMJ. 1995;310(6982):752-4.
6. Egger M, Smith GD, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ. 1997;315(7109):629-34.
7. Boland A, Dundar Y, Bagust A, Haycox A, Hill R, Mujica Mota R, et al. Early thrombolysis for the treatment of acute myocardial infarction: a systematic review and economic evaluation. Health technology assessment (Winchester, England). 2003;7(15):1-136. Epub 2003/05/30.
8. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ. 2003;327(7414):557-60. Epub 2003/09/06.
9. O'Hagan A, Forster J, Kendall MG. Bayesian inference: Edward Arnold London; 1994.
10. Spiegelhalter D, Thomas A, Best N, Gilks W. BUGS: Bayesian inference using Gibbs sampling, Version 0.50. MRC Biostatistics Unit, Cambridge. 1995.

Appendix B

Corticosteroids following traumatic head injury

Contents

B1. Introduction

The case study of CRASH demonstrates the health consequences of not considering the value of additional research on the use of steroids following traumatic head injury. It relates to an example of where additional evidence, through the funding of a large randomised control trial (CRASH), was required to prevent thousands of iatrogenic deaths.(1-2) The existing evidence base before CRASH was insufficient to reliably determine whether there was a clinical benefit or harm from the use of steroids following head injury. This lack of reliable evidence led to wide variation in the clinical use of steroids worldwide.(3-6) Formal analytic methods for establishing the value of additional research were not used at the time that CRASH was funded. A retrospective analysis of the evidence available before CRASH is conducted in order to show that formal methods of value of information analysis would have been useful for quantifying uncertainty in the effects of steroid use in traumatic head injury and could have indicated the best course of action to prevent thousands of unnecessary deaths.

B2. Background to the case study

In the late 1990s the benefit or harm of using steroids to treat patients following traumatic head injury (THI) was unclear. A systematic review of randomised controlled trials (RCTs) of corticosteroids in acute THI concluded that despite 25 years of RCTs examining the effects of corticosteroids on death and disability their effects remain unclear(7). In the review, the risk of death in those given corticosteroids was 1.8% less than in the control group but the 95% confidence interval (CI) was 5.7% less to 2.5% more. The lack of reliable evidence led to wide variation in the clinical use of steroids worldwide. A 1995 US survey of the intensive care management of head injury patients found that corticosteroids were used in 64% of trauma centres(3), while a 1998 UK survey found that corticosteroids were used in 12% of 263 intensive care units(6).

The annual incidence of severe head injury is estimated to be 15 per 100,000 people(8). This results in an annual incidence of approximately 8,800 in the UK (corresponding to a population estimate of 58.9 million in 2000, the year that the grant application for CRASH was submitted). Serious head injury often leads to death or disability with profound effects on the subsequent quality of life of affected individuals(9). The unproven effectiveness of steroids prompted the need for a large RCT. The CRASH trial (Corticosteroid Randomisation After Significant Head injury) was successfully funded by the UK Medical Research Council as the largest RCT ever to be conducted in head injury to examine the effects of a short term infusion of corticosteroids on death and disability(1-2). The application for funding of CRASH was successful on the basis that there was: (i) inadequate evidence of the effects of steroids in THI(7), and (ii) recent evidence of benefit from corticosteroids in acute spinal cord injury led to renewed interest in their role in head injury(10-11). However, the results of CRASH were alarming. After enrolling 10,008 randomised adults with head injury, the risk of death at 6 months was higher in the corticosteroid group than in the placebo group (1248 [25.7%] versus 1075 [22.3%] deaths; relative risk 1.15, 95% CI 1.07-1.24; p=0.0001)(2). Similarly, the risk of death or severe disability was higher in the corticosteroid group (1828 [38.1%] versus 1728 [36.3%] dead or severely disabled; relative risk 1.05, 95% CI 0.99-1.10; p=0·079)(2). The results reliably refute any reduction in mortality or severe disability with corticosteroids in the 6 months after head injury.

The prevention of thousands of unnecessary iatrogenic deaths following head injury hinges on the fact that the funding application for CRASH was successful. Had the application failed the collection of valuable information on best clinical practice would not have been obtained and people would continue to be harmed by steroids. This case study examines the evidence available before CRASH to quantify the value of obtaining further evidence on the use of steroids in THI and the expected health consequences of not obtaining the evidence.

B3. Evidence before CRASH

Before CRASH there were 19 RCTs comparing the use of corticosteroids with a control group (placebo or no treatment) in acute THI. These trials dating from 1972 to 1995 were of varying study quality, length of follow-up, steroids administered (predominantly prednisolone, betamethasone, cortisone, dexamethasone, hydrocortisone, methylprednisolone, prednisone, and triamcinolone), doses and time to

administration. The primary outcome reported in 16 of the trials was number of deaths at the end of the study period(12-27), while the Glasgow outcome scale (GOS)(28), which categorises people into one of five health states: (i) dead; (ii) persistent vegetative; (iii) severe disability; (iv) moderate disability; and (v) recovery, was used to assess neurological outcomes in 7 of the trials(13, 15-20). A further 2 trials reported the combined number of people dead, vegetative and severely disabled at end of study(21-22).

A systematic review of corticosteroids for acute THI published in the British Medical Journal (BMJ) in 1997(7) used a fixed effect meta-analysis to synthesise the evidence on death or disability. The pooled odds ratio for death from the fixed effect analysis was 0.91 with 95% confidence interval of 0.74 to 1.12, while the pooled odds ratio for death or disability was 0.90 with 95% confidence interval of 0.72 to 1.11.(7) The assumption underlying the fixed effect model is that each trial result is estimating a common unknown pooled effect, while any variation in the estimated effect size across studies is due to sampling error. If this assumption is considered to be too restrictive, a random effects model can be used, which allows the unknown pooled effect to vary between studies. In the case of THI, a random effects model that specifically allows for the existence of between study heterogeneity as well as within study variability may be more appropriate given the different sources of variation between the studies. Furthermore, Bayesian approaches have been advocated for random effects meta-analysis(29-30). Bayesian methods allow for greater uncertainty than the classical approach to statistical inference by allowing for the fact that both the overall population effect and the between study effect in random effects meta-analyses are estimated by the data(29).

A Bayesian fixed effect and random effects meta-analysis is used to re-synthesise the evidence available before CRASH and to include three of the 19 RCTs that were not included in the original synthesis(14, 25, 27). The model is based on a binomial likelihood that uses the event data for all GOS outcomes and number of individuals in each study arm directly, without the need to assume that the log odds ratio from individual studies are normally distributed.

B3.1 The effect of steroids on the primary endpoint of mortality

Figure B1 shows the forest plot of the evidence available before CRASH for the primary outcome of death. The summary odds ratio (OR) from the Bayesian fixed effect analysis is 1.07 with 95% credible interval (probability of containing the true effect) of 0.89 to 1.28, while the OR from the Bayesian random effects meta-analysis is 0.93 (95% CrI 0.71 to 1.18). The evidence from the random effects meta-analysis suggests that the use of steroids following THI could reduce the risk of death by 1.8% (i.e. a reduction of nearly 2 deaths for every 100 people treated) when using the average death rate in the control arms of the meta-analysis of 35.3%. However, the credible interval spans the no difference line (odds ratio = 1.0) indicating that the change in the risk of death could be 12.5% lower to 9.9% higher. Interestingly, the summary OR from the fixed effect analysis is against the use of steroids in THI but again the result is not statistically significant.

B3.1.1 Consequences of uncertainty in number of deaths averted per annum

The annual incidence of THI in the UK was approximately 8,800 in 2000 (the year that the grant application for CRASH was submitted), while a 1998 UK survey indicated that corticosteroids were used to treat THI in approximately 12% of intensive care units. An assessment of the likely consequences of uncertainty in the OR of death can be used to judge whether the scale of the consequences might justify further research. If the decision on whether or not to use steroids was judged to be 100% certain then there are no consequences and so there would be nothing to be gained by more research. However, as the probability that the decision is correct becomes less certain, the expected consequences (and hence potential value of more research) increases. Table B1 shows the probability of no consequences and the expected consequences in number of deaths averted per annum for different levels of utilisation of steroids in current practice: (i) no steroids; (ii) steroids; and (iii) 12% steroids (88% no steroids) for both fixed and random effects analysis. For the fixed effect analysis, the summary OR suggests that steroids should not be used in clinical practice. If this decision is taken under the current level of uncertainty, the probability of no consequences is 0.75 and the corresponding expected consequences are 27 deaths per annum. In contrast, the summary OR from the random effects analysis suggests that steroids should be

used in clinical practice. If steroids are used, the probability of no consequences is 0.74 with corresponding expected consequences of 40 deaths per annum.

The same analysis can be used to record the frequency of errors in the decision on whether or not to use steroids in clinical practice. Figures B2 and B3 show the distribution of consequences of uncertainty for different levels of utilisation under a fixed and random effects analysis, respectively. Most commonly there are no consequences when the decision is in line with that suggested by the summary OR. However, if this decision turns out to be the wrong the consequences of error may be relatively small, e.g., 1-50 deaths per annum, but it may be very large, although less likely, e.g. 500 deaths or more per annum. The average across the distribution of consequences gives the expected consequences of uncertainty in Table B1.

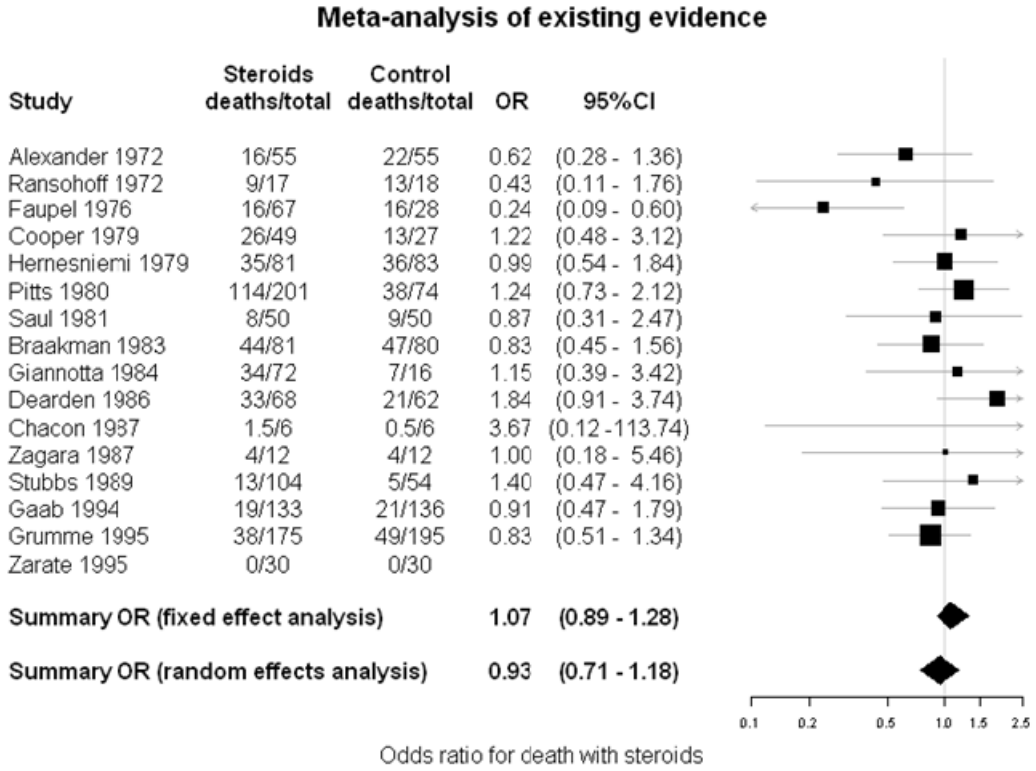


Figure B1: Fixed and random effects meta-analysis of the effect of steroids on mortality

Table B1: Expected consequences of uncertainty in number of deaths averted per annum

Level of utilisation in current practice	Fixed effect analysis		Random effects analysis	
	Probability of no consequences	Expected consequences, number of deaths averted per annum	Probability of no consequences	Expected consequences, number of deaths averted per annum
100% no steroids	0.75	27	0.26	199
100% steroids	0.25	152	0.74	40
12% steroids, 88% no steroids	-	42	-	180

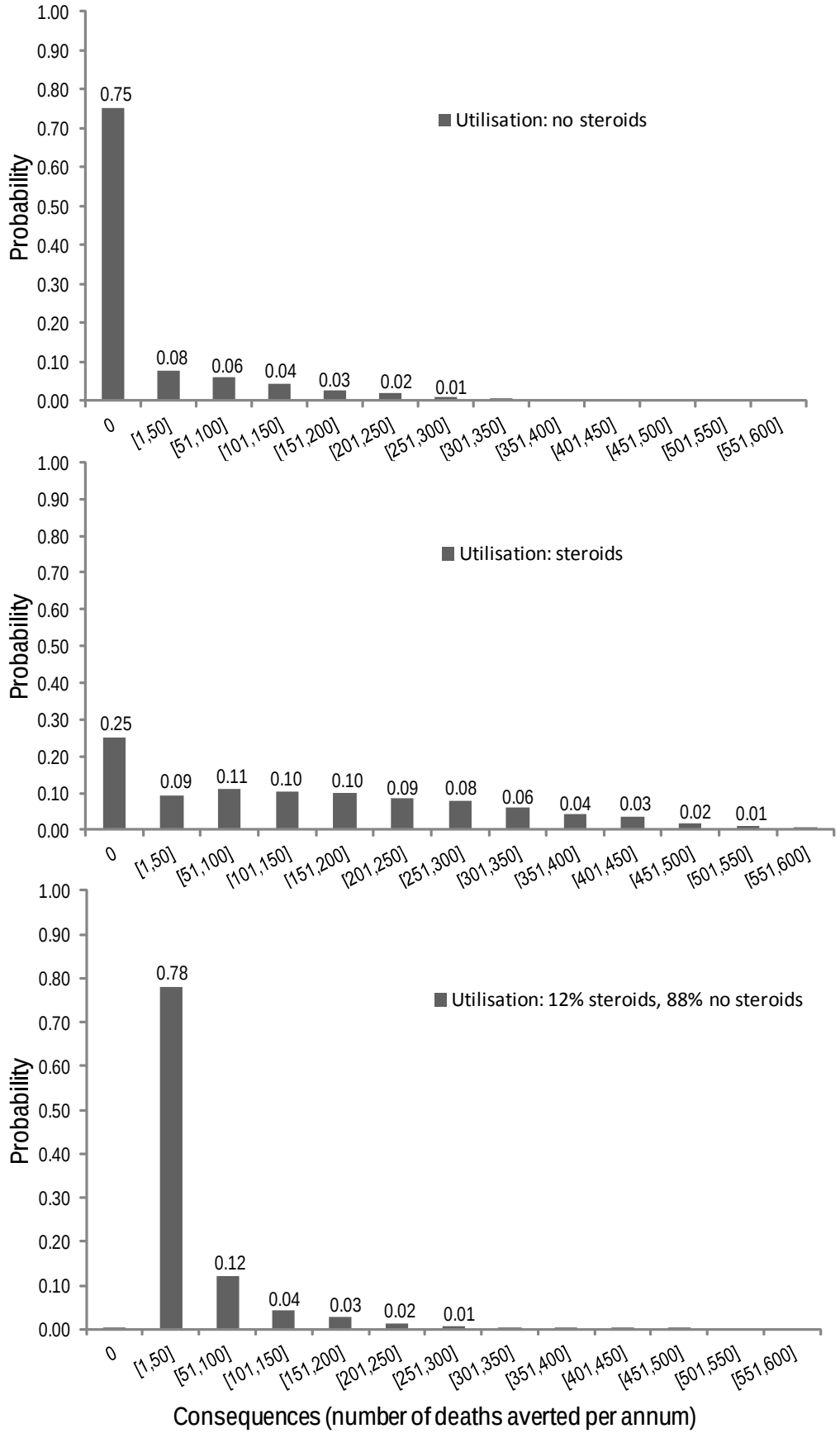


Figure B2: Distribution of the consequences of uncertainty for fixed effect analysis

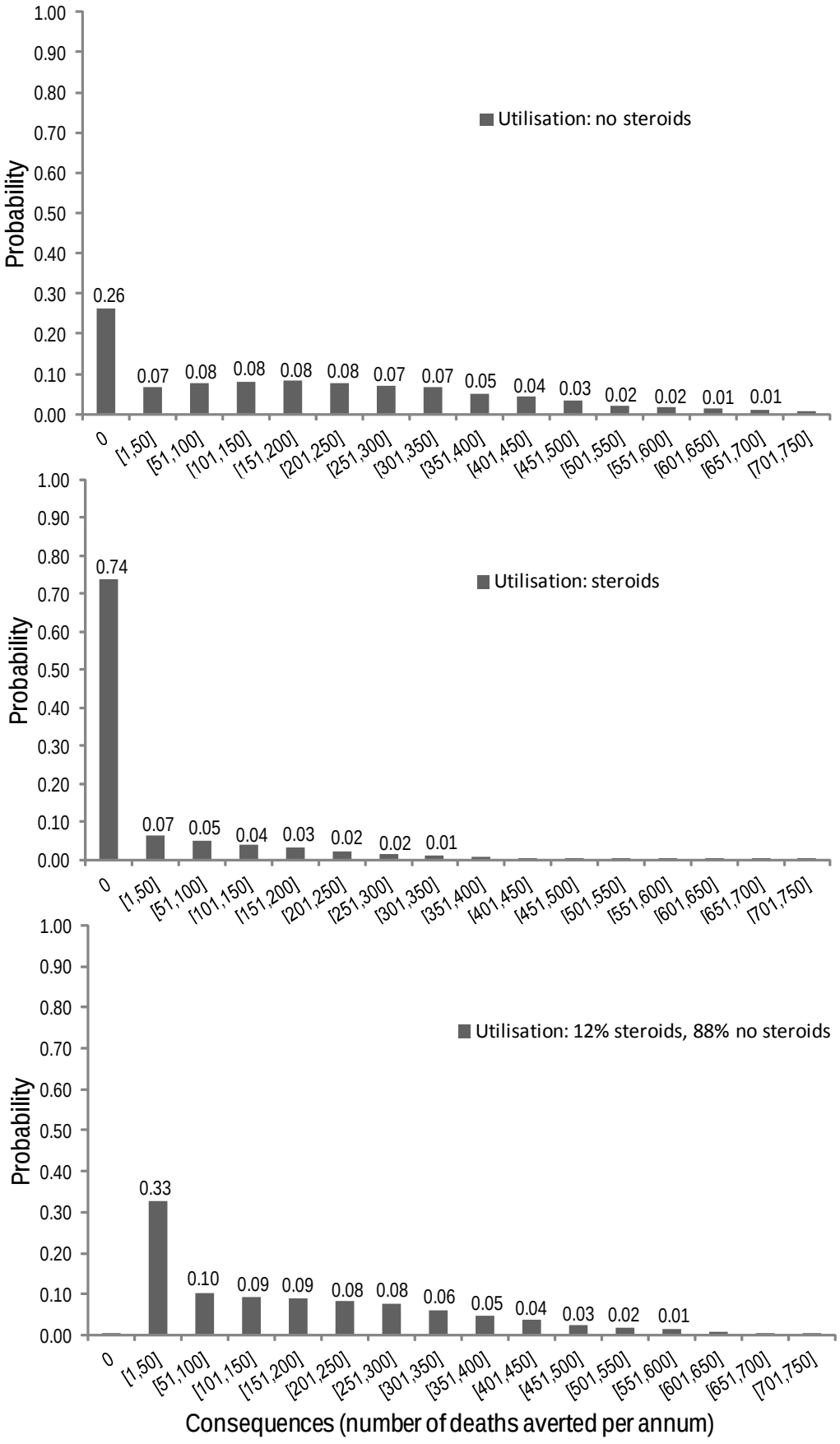


Figure B3: Distribution of the consequences of uncertainty for random effects analysis

B3.1.2 Time until research reports

The research required to provide the additional evidence will take time to complete and report. Therefore, an assessment of the potential benefits of evidence should take account of the fact that the patient population will not benefit from the results of the research until it becomes available. If treatment decisions are irreversible (e.g. an acute indication) then it is only those patients' incident after the research reports that will realise any of the potential benefits. In order to estimate the scale of the total patient population who can benefit from the research, a judgement about the time horizon over which the technology will be used is also required (e.g. a time horizon of 10 years implies a total patient population of approximately 76,000 based on the expected incidence of THI over 10 years). Table B2 shows the expected consequences of uncertainty in number of deaths averted for time taken for research to report over different technology time horizons, assuming current utilisation of 12% on steroids. The potential value of further research declines with the time to research reporting. The length of time that it takes for research to report will depend in part on the design (length of follow-up, sample size and endpoints), recruitment rates and size of the eligible patient population, as well as how efficient the organisation and data collection might be.

Table B2: Expected consequences of uncertainty in number of deaths averted for time taken for research to report over different time horizons for current utilisation of 12% on steroids (random effects analysis)

Time until research reports, years	Expected consequences (number of deaths averted) by technology time horizon (years)									
	10	9	8	7	6	5	4	3	2	1
Immediately	1,546	1,415	1,278	1,137	991	840	683	521	353	180
1	1,367	1,235	1,098	957	811	660	503	341	174	0
2	1,193	1,061	925	784	638	486	330	168	0	0
3	1,025	894	757	616	470	319	162	0	0	0
4	863	732	595	454	308	157	0	0	0	0
5	707	575	439	297	151	0	0	0	0	0
6	556	424	287	146	0	0	0	0	0	0
7	409	278	141	0	0	0	0	0	0	0
8	268	136	0	0	0	0	0	0	0	0
9	132	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0

Although there may be considerable value to additional evidence there is no guarantee that the research recommended will be conducted. Therefore, the probability that research will report at a particular time also needs to be considered. Figure B4 shows the expected consequences in number of deaths averted over a time horizon of 10 years by the time it takes for research to report and the likelihood of the research being completed. If research is certain to report, the maximum value of the evidence is 1,546 deaths averted over a 10-year period. However, if there is only a 50% chance of the research reporting within 5 years, then the value of the evidence is reduced to 353 deaths averted.

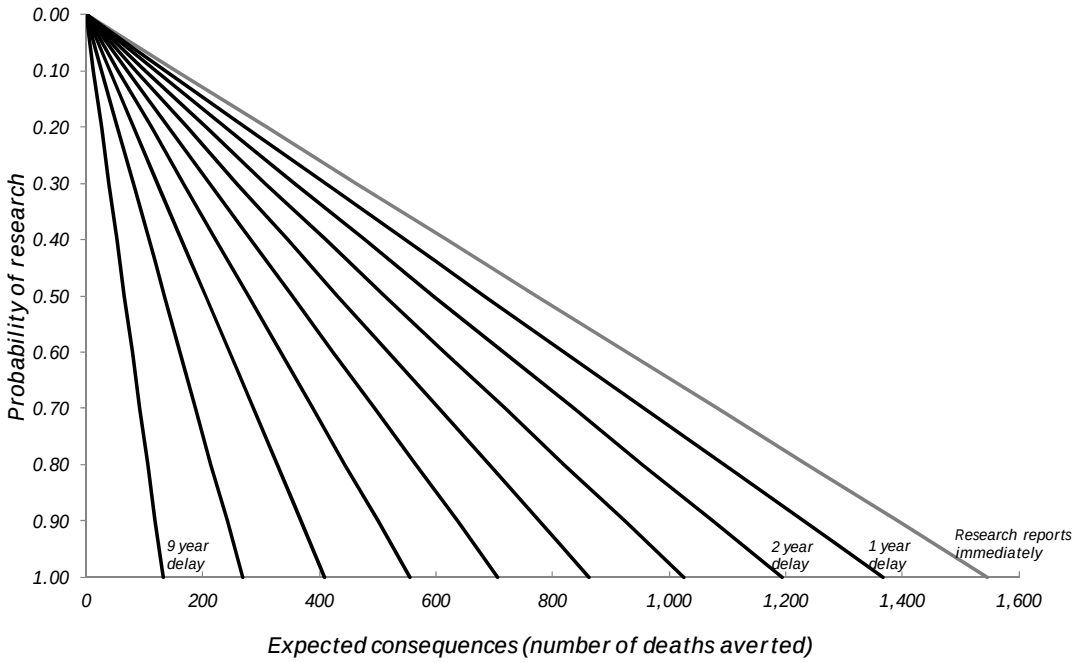


Figure B4: Expected consequences of uncertainty in number of deaths averted over a time horizon of 10 years by time for research to report and likelihood of research being completed (current utilisation of 12% on steroids and random effects analysis)

B3.1.3 Minimum clinical difference in number of deaths

A large improvement in the number of deaths averted might be required if clinical practice is unlikely to change without it. The 1998 UK survey indicated that steroids were used in clinical practice in 12% of cases. Figure B5 shows the expected consequences for a range of minimum clinical differences (effect sizes) in the absolute risk of death that might be required to change this level of utilisation. When the minimum clinical difference is 0%, the expected consequences represent the maximum value of evidence (resolving all uncertainty). As the effect size, or difference in the risk of death between steroids and no steroids, increases the decision on whether or not to use steroids in clinical practice becomes less uncertain and the value of evidence is reduced. The underlying assumption is that practice will change if perfect information (indicating the best treatment choice) shows that the difference in the risk of death between steroids and no steroids is reduced by more than the minimum clinical difference. In the fixed effect analysis, a minimum clinical difference of 4% is required to change the level of implementation such that there are no consequences in deaths per annum, whereas in the random effects analysis a 10% difference is required.

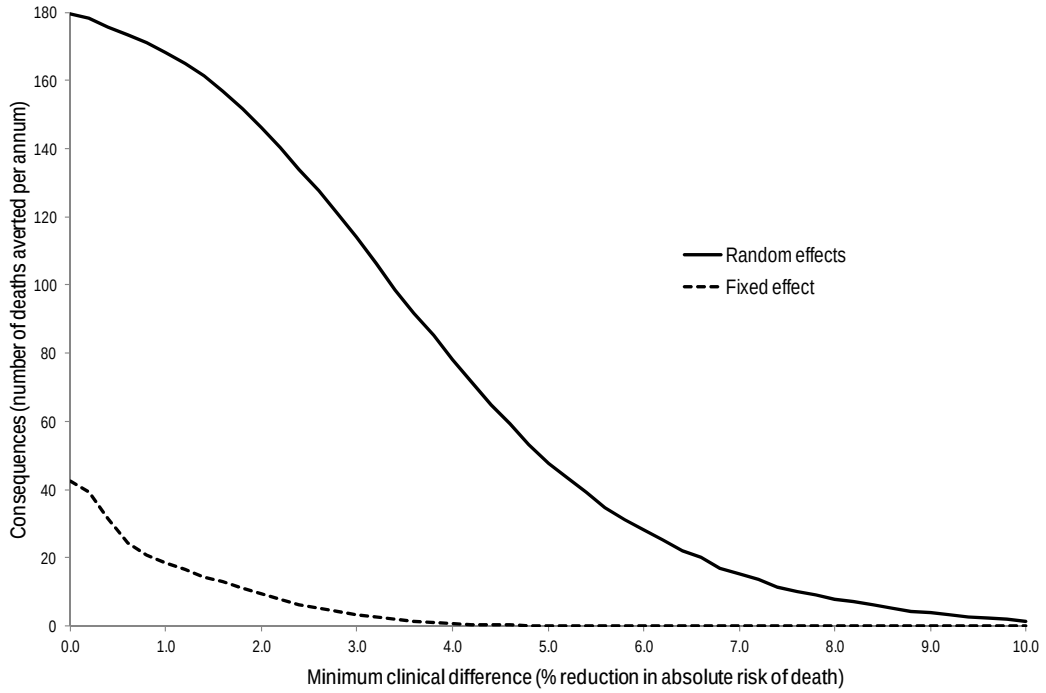


Figure B5: Consequences of uncertainty in number of deaths averted per annum for a minimum clinical difference in the absolute risk of death needed to change clinical practice

B3.2 The effect of steroids on other aspects of outcome

The health impact of THI extends beyond the number of deaths per annum. Individuals who survive are likely to experience health consequences in terms of disability and reduction in life expectancy and quality of life. The magnitude of these health impacts will also be affected by the use of steroids. Therefore, it's important to consider all outcomes of health when quantifying the value of additional evidence. The Glasgow Outcome Scale (GOS) is used to assess neurological outcomes in THI. Table B3 shows the proportion of individuals expected to be in each of the GOS outcomes by treatment based on the results of the evidence synthesis on all outcomes (random effects analysis). The risk of being vegetative or severely disabled is higher on average with steroids compared with the control arms of the RCTs, while the risk of being moderately disabled or good recovery is lower with steroids.

Table B3: Distribution of GOS outcomes by treatment and number of years lived in full health

	Percentage of individuals (95% CrI) by treatment		Health-related quality of life weights (SE)	Years lived in full health‡
GOS outcome	Steroids	No steroids		
Dead	33.5 (22.8, 45.2)	35.3 (24.8, 46.9)	0.00	0.00
Vegetative	4.8 (2.8, 7.5)	3.8 (2.4, 5.9)	0.08 (0.16)	0.56
Severe disability	13.5 (8.3, 20.1)	10.7 (7.1, 15.8)	0.26 (0.25)	3.24
Moderate disability	11.6 (8.6, 14.8)	12.1 (9.2, 15.1)	0.63 (0.27)	10.51
Good recovery	36.5 (28.1, 44.8)	38.0 (30.1, 45.6)	0.85 (0.19)	15.39

‡For an average age of 50 years and discounted at a rate of 3.5% per annum
CrI, credible interval; SE, standard error

When the evidence for the worse health outcomes of dead, vegetative and severely disabled are combined the odds ratio is 1.10, with 95% CrI of 0.81 to 1.53[1]. The use of steroids increases the risk of negative outcomes (total number of individuals dead, vegetative and severely disabled) by 2% when using the average rate of 49.8% in the control arms of the RCTs, but the 95% CrI indicates that the change in the risk of negative outcomes could be 15.9% lower to 23.0% higher. Although steroids are expected to reduce the number of deaths on average (when considering the outcome of death alone in a random effects analysis), individuals surviving are more likely to be in one of the worse health outcomes.

This increase in negative outcomes (proportion of individuals vegetative or severely disabled) should be compared with any improvement in number of deaths in order to come to an assessment about the effects of steroids on all aspects of health outcome. The outcome of survivors can be quantified in terms of their subsequent quality of life. Shavelle et al (2007) has estimated the expected life expectancy of an individual following THI by age and severity of disability (taking account of any change in health status over the individual's lifetime)(31). In order to quantify an individual's remaining life expectancy in terms of years lived in full health, health-related quality of life weights are used to weight survival in worse health states lower than survival in full health[2]. Table B3 shows the number of years lived in full health and corresponding health-related quality of life weights for the GOS outcomes. The number of years expected to be lived in full health in a vegetative or severely disabled outcome (0.56 and 3.24 years, respectively) is considerably lower than a moderately disabled or good recovery outcome (10.51 and 15.39 years, respectively) for an average age of 50 years.

B3.2.1 Consequences of uncertainty in number of years lived in full health per annum

It is now possible to combine the evidence on the risk of being in a particular GOS outcome with the quality of life associated with that outcome in order to quantify the effects of steroids on the number of years lived in full health. Again if the decision on whether or not to use steroids was judged to be 100% certain then there would be no consequences and no value to additional evidence. However, as the probability that the decision is correct becomes less certain, the expected consequences increases. Figure B6 shows the distribution of consequences in number of years lived in full health per annum for the different levels of utilisation of steroids in current practice. The likelihood of no consequences is 63% if steroids are not implemented in practice and 37% if they are. The maximum value of evidence (expectation across the distribution of consequences) is 1,067 years of full health per annum, 2,711 years, and 1,264 years for current implementation of no steroids, steroids, and 12% utilisation of steroids, respectively.

[1] The uncertainty in the effects of steroids on the risk of death, vegetative and severely disabled is greater than on the risk of death alone (as evidenced by the wider credible interval) because fewer trials report the outcome of survivors.

[2] The conventional scale for health-related quality of life weights is between 0 (death) and 1 (perfect health). One year in a health state with a weight of 0.5 is equivalent to half a year in full health.

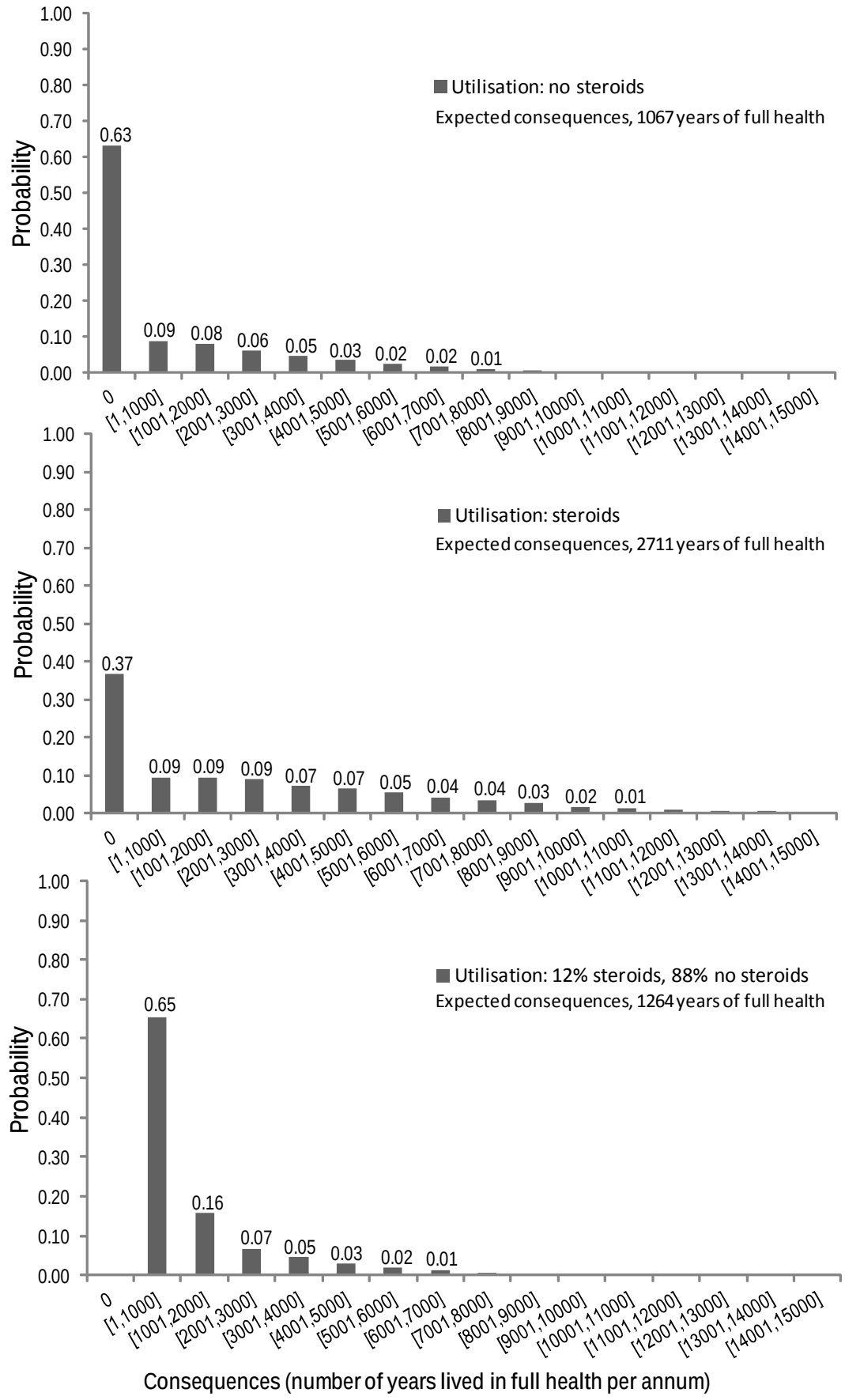


Figure B6: Distribution of consequences of uncertainty in number of years lived in full health per annum

B3.2.2 Time until research reports

As discussed above, consideration should be given to the time taken for research to complete and report, the likelihood of completion, and how much of the uncertainty is resolved. Table B4 shows the expected consequences of uncertainty in number of years lived in full health for time until research reports over different technology time horizons, assuming current implementation of 12% on steroids. If research is certain to complete and report immediately, the maximum value of evidence is 10,884 years of full health over a 10-year period. Figure B7 shows the corresponding expected consequences over 10 years taking account of the likelihood that the research will be completed. A 50% chance of the research reporting within 5 years reduces the value of evidence to 2,488 years lived in full health.

Table B4: Expected consequences of uncertainty in number of years lived in full health for time taken for research to report over different time horizons for current utilisation of 12% on steroids

Time until research reports, years	Expected consequences (number of years lived in full health) by technology time horizon (years)									
	10	**9**	**8**	**7**	**6**	**5**	**4**	**3**	**2**	**1**
Immediately	10,884	9,956	8,996	8,002	6,974	5,909	4,807	3,667	2,486	1,264
1	9,620	8,692	7,732	6,738	5,709	4,645	3,543	2,402	1,222	0
2	8,398	7,470	6,510	5,516	4,487	3,423	2,321	1,180	0	0
3	7,218	6,290	5,330	4,336	3,307	2,242	1,140	0	0	0
4	6,077	5,149	4,189	3,195	2,167	1,102	0	0	0	0
5	4,975	4,047	3,087	2,093	1,065	0	0	0	0	0
6	3,911	2,983	2,023	1,029	0	0	0	0	0	0
7	2,882	1,954	994	0	0	0	0	0	0	0
8	1,888	960	0	0	0	0	0	0	0	0
9	928	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0

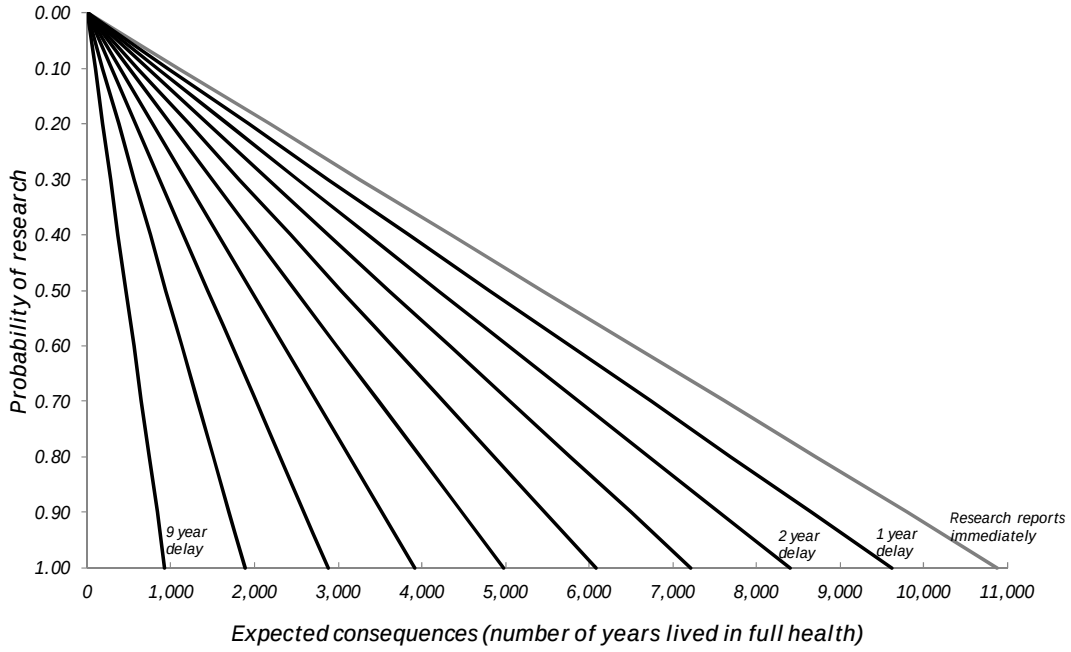


Figure B7: Expected consequences of uncertainty in number of years lived in full health over a time horizon of 10 years by time for research to report and likelihood of research being completed (current utilisation of 12% on steroids and random effects analysis)

B3.2.3 Minimum clinical difference in number of years lived in full health

Figure B8 shows the expected consequences for a range of minimum clinical differences in the absolute number of years lived in full health that might be required to change the current level of utilisation of 12% on steroids. Again the underlying assumption is that practice will change if perfect information (indicating the best treatment choice) shows that the difference in the number of years lived in full health with treatment are reduced by more than the minimum clinical difference. A minimum clinical difference of 0% represents the maximum value of evidence. As the difference increases by 3% the value of evidence required to change the current level of implementation falls rapidly.

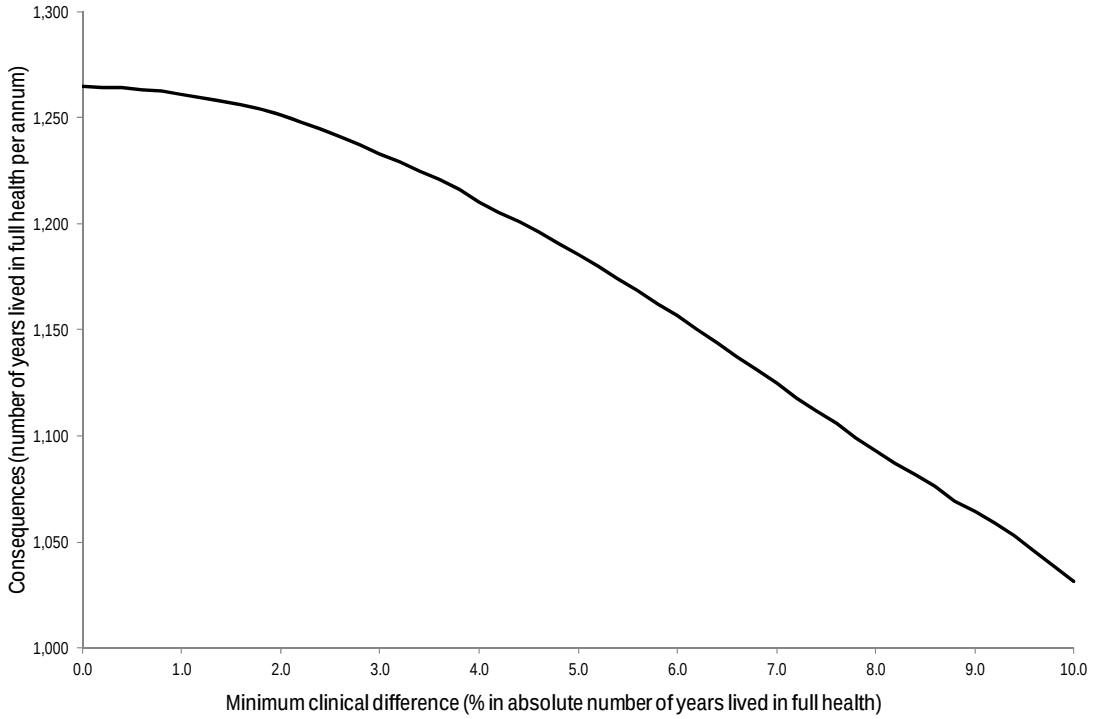


Figure B8: Consequences of uncertainty for a minimum clinical difference in the absolute number of years lived in full health needed to change clinical practice

B3.3 Type of evidence required

An assessment of the type of evidence required to reduce the uncertainty in the decision on whether or not to use steroids can be directly informed by the analysis. The different sources of uncertainty include: (i) the effect of steroids on the risk of death; (ii) the effect of steroids on the risk of disability and GOS outcomes; and (ii) uncertainty in the health-related quality of life weights applied to the life expectancy of survivors. Figure B9 separates apart the sources of uncertainty to determine which elements contribute most to the decision uncertainty in order to indicate where more research is most valuable. Uncertainty in the effects of steroids on the risk of death, vegetative or severe disability is almost exclusively responsible for the overall decision uncertainty. Future research that simply reports on the number of deaths after steroid use only would not be sufficient to reduce the decision uncertainty since no further information on the effect of steroids on disability and reduction in life expectancy and quality of life of survivors would become known. Interestingly, there is little additional value from knowing all the GOS outcomes relative to knowing the most negative ones (proportion of individuals dead, vegetative or severely disabled).

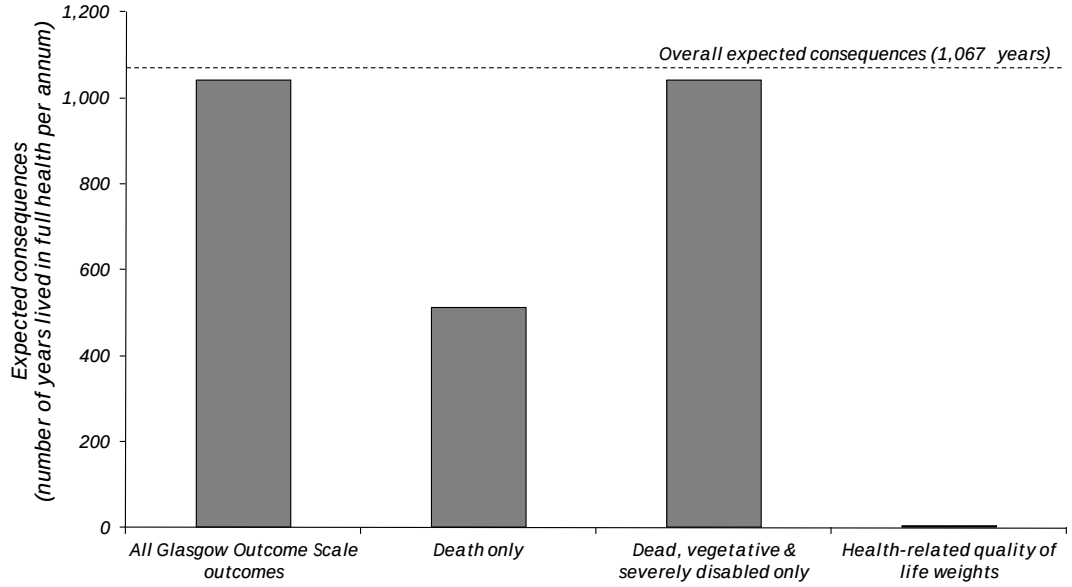


Figure B9: Consequences of uncertainty associated with different sources of uncertainty for current utilisation of no steroids

B4. Evidence after CRASH

The first results of CRASH were reported in 2004 showing the effect of corticosteroids on death within 14 days in 10,008 randomised adults with significant head injury(1). Mortality data during the first 2 weeks were obtained for 9,964 patients. Of 4,985 patients allocated corticosteroids whose outcomes were known, 1052 (21%) died within 2 weeks of randomisation, compared with 893 (18%) of 4979 allocated placebo. The corresponding relative risk of death within 2 weeks for corticosteroids compared with placebo was 1.18 (95% CI 1.09-1.27; p=0.0001). The final results of CRASH were reported in 2005 for outcomes on the GOS at 6 months after injury. The risk of death at 6 months was higher in the corticosteroid group than in the placebo group (1248 [25.7%] versus 1075 [22.3%] deaths; relative risk 1.15, 95% CI 1.07-1.24; p=0.0001)(2). Similarly, the risk of death or severe disability was higher in the corticosteroid group (1828 [38.1%] versus 1728 [36.3%] dead or severely disabled; relative risk 1.05, 95% CI 0.99-1.10; p=0·079)(2). The results reliably refute any reduction in mortality or severe disability with corticosteroids in the 6 months after head injury.

A Bayesian meta-analysis is used to re-synthesise the evidence on steroids following THI by updating the existing evidence to include the findings of CRASH. The meta-analysis is updated by including the results from CRASH as a fixed effect and using the evidence pre-dating CRASH as a random effects prior in the synthesis.[3] The annual incidence of THI is also updated from approximately 8,800 to 9,000 to correspond to the year 2005 (the year that the final results of CRASH were published).

B4.1 The effect of steroids on the primary endpoint of mortality

Figure B10 shows the forest plot with the updated evidence from CRASH for the primary outcome of death. When all previous trials are combined with CRASH, the odds ratio for death is 1.21 with 95% CrI 1.10 to 1.33. The updated evidence suggests that the use of steroids in THI increases the risk of death by 3.4% (i.e. an increase of about 3 deaths for every 100 people treated) when using the average death rate in the control arms of 22.3%. The credible interval indicates that steroids always increase the risk of death by a minimum of 2.2% up to 4.7% compared with no steroids. The CRASH trial result, judged either separately or in combination with the previous trial evidence, refutes any reduction in mortality with corticosteroids, although the size of the CRASH trial has a major influence on the result of the meta-analysis.

B4.1.1 Consequences of uncertainty in number of deaths averted per annum

On the basis of evidence on the endpoint of mortality, steroids should not be used routinely to treat THI. By clearly refuting a mortality benefit from steroid use, health outcomes can be improved by ensuring that the accumulated findings of research are implemented and has an impact on clinical practice. This requires the 12% of steroid use in clinical practice to be switched to no steroids. The value of implementation in this case is 37 deaths averted per annum. An assessment of the likely consequences of uncertainty in the updated odds ratio of death indicates that there are no consequences and nothing to be gained by further research (i.e. the decision not to use steroids is judged to be 100% certain).

[3] Although the trials prior to CRASH were more appropriately synthesised in a random effect meta-analysis, the results of CRASH (which was designed and commissioned to be of high quality and directly relevant to clinical practice and the target patient population) enter as a fixed effect (prior based on random effects is updated using a fixed effect). It would seem inappropriate to enter CRASH as random effect which would down weight its results by imagining that CRASH, like the previous trials, was randomly drawn from the same population of previous studies.

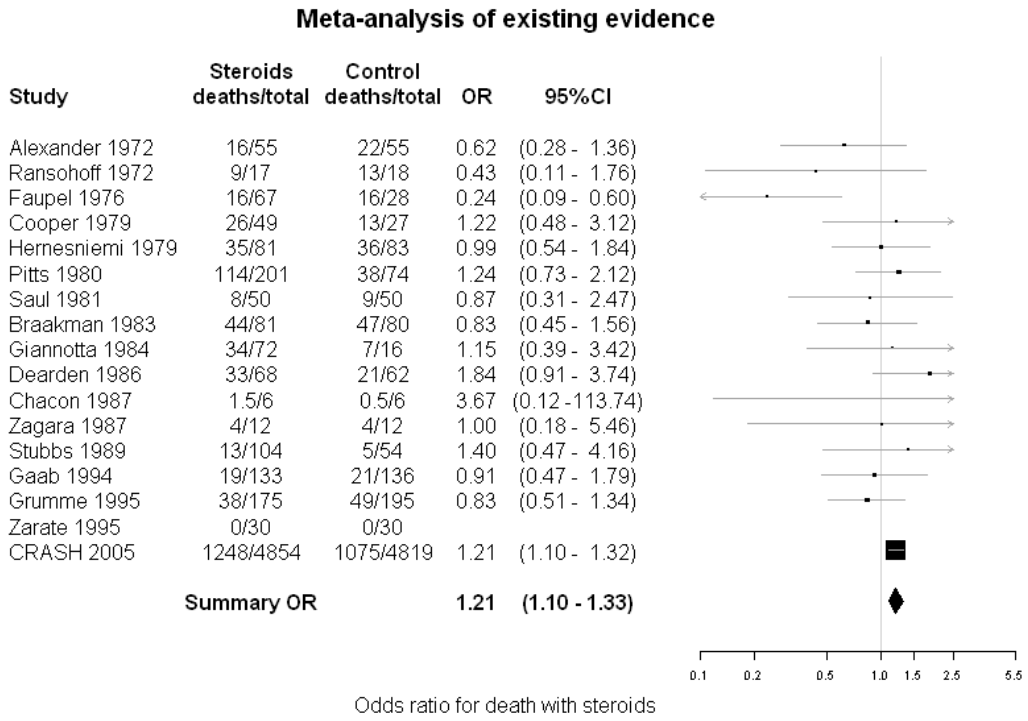


Figure B10: Updated meta-analysis of the effect of steroids on mortality

B4.2 The effect of steroids on other aspects of outcome

As discussed previously, the health impact of steroids on THI extends beyond their effect on death. The CRASH trial reported the outcomes of the GOS at 6 months after injury. Table B5 shows the proportion of individuals expected to be in each of the GOS outcomes by treatment based on the updated results of the evidence synthesis on all outcomes. The risk of severe disability is now lower with steroids compared with the control arms of the RCTs, while the risk of moderate disability is slightly higher and good recovery lower. When the evidence for the worse health outcomes of dead, vegetative and severely disabled are combined the odds ratio is 1.13, with 95% CrI 1.05 to 1.22.

Table B5: Distribution of GOS outcomes by treatment

GOS outcome	Percentage of individuals (95% CrI) by treatment	
	Steroids	No steroids
Dead	25.7 (24.5, 27.0)	22.3 (21.2, 23.5)
Vegetative	0.02 (0.00, 0.08)	0.02 (0.00, 0.08)
Severe disability	11.9 (10.1, 13.8)	13.5 (11.7, 15.3)
Moderate disability	17.9 (16.8, 19.0)	17.2 (16.2, 18.3)
Good recovery	44.5 (43.1, 45.9)	46.9 (45.5, 48.3)

CrI, credible interval

B4.2.1 Consequences of uncertainty in number of years lived in full health per annum

It is now possible to combine the evidence on the risk of being in a particular GOS outcome with the quality of life associated with that outcome in order to update the effect of steroids on the number of years lived in full health. If we assume that we can get the 12% of steroid use in clinical practice to be switched to no steroids, the value of implementation is 392 years of full health per annum. If steroids are not used in clinical practice after CRASH, an assessment of the likely consequences of uncertainty in the updated outcomes indicates that there is very little remaining uncertainty; the expected consequences are 3.2 years of full health per annum. Therefore, when the analysis of the potential value of additional evidence is updated, there are no expected benefits of acquiring additional evidence.

References

1. CRASH trial collaborators. Effect of intravenous corticosteroids on death within 14 days in 10008 adults with clinically significant head injury (MRC CRASH trial): randomised placebo-controlled trial. The Lancet. 2004;364:1321-28.
2. CRASH trial collaborators. Final results of MRC CRASH, a randomised placebo-controlled trial of intravenous corticosteroid in adults with head injury - outcomes at 6 months. Lancet. 2005;365(9475):1957-9.
3. Ghajar J, Hariri RJ, Narayan RK, Iacono LA, Firlik K, Patterson RH. Survey of critical care management of comatose, head-injured patients in the United States. Critical Care Medicine. 1995;23:560-7.
4. Jeevaratnam DR, Menon DK. Survey of intensive care of severely head injured patients in the United Kingdom British Medical Journal. 1996;312:944-7
5. Matta B, Menon D. Severe head injury in the United Kingdom and Ireland: A survey of practice and implications for management. Critical Care Medicine. 1996;24:1743-8.
6. McKeating EG, Andrews PJ, Tocher JI, Menon DK. The intensive care of severe head injury: a survey of non-neurosurgical centres in the United Kingdom. British Journal of Neurosurgery. 1998;12(1):7-14.
7. Alderson P, Roberts I. Corticosteroids in acute traumatic brain injury: systematic review of randomised controlled trials. British Medical Journal. 1997;314:1855-59.
8. Sauerland S, Maegele M. A CRASH landing in severe head injury. The Lancet. 2004;364:1291-92.
9. Yates D, Farrell B, Teasdale G, Sandercock P, Roberts I. Corticosteroids in head injury - the CRASH trial. Journal of Accident & Emergency Medicine. 1999;16:83-90.
10. Bracken MB, Shepard MJ, Collins WF, et al. A randomized, controlled trial of methylprednisolone or naloxone in the treatment of acute spinal-cord injury. Results of the Second National Acute Spinal Cord Injury Study. New England Journal of Medicine. 1990;322(20):1405-11.
11. Bracken MB, Shepard MJ, Holford TR, et al. Administration of methylprednisolone for 24 or 48 hours or tirilazad mesylate for 48 hours in the treatment of acute spinal cord injury. Results of the Third National Acute Spinal Cord Injury Randomized Controlled Trial. National Acute Spinal Cord Injury Study. Journal of the American Medical Association 1997;277(20):1597-604.
12. Alexander E. Medical management of closed head injuries. Clinical Neurosurgery 1972;19:210-50.
13. Braakman R, Schouten HJA, Blaauw-van Dishoeck M, Minderhoud JM. Megadose steroids in severe head injury. Journal of Neurosurgery. 1983;58:326-30.
14. Chacon L. Brain edema in severe head injury on children treated with and without dexametasone. Med Crit Venez 1987;2:75-9.
15. Cooper PR, Moody S, Clark WK, Kirkpatrick J, Maravilla K, Gould AL, et al. Dexamethasone and severe head injury. Journal of Neurosurgery. 1979;51:307-16.
16. Dearden NM, Gibson JS, McDowall DG, Gibson RM, Cameron MM. Effect of high dose dexamethasone on outcome from severe head injury. Journal of Neurosurgery. 1986;64:81-8.
17. Faupel G, Reulen HJ, Muller D, Schurmann K. Double-blind study on the effects of steroids on severe closed head injury. In: Pappius MM, Feindel W, editors. Dynamics of brain edema. Berlin: Springer-Verlag; 1976.
18. Gaab MR, Trost HA, Alcantara A, Karimi-Nejad A, Moskopp D, Schultheiss R, et al. "Ultrahigh" dexamethasone in acute brain injury. Zentralbl Neurochir. 1994;55:135-43.
19. Giannotta SL, Weiss MH, Apuzzo MLJ, Martin E. High dose glucocorticoids in the management of severe head injury. Neurosurgery. 1984;15:497-501.
20. Grumme T, Baethmann A, Kolodziejczyk D, Krimmer J, Fischer M, et al. Treatment of patients with severe head injury by triamcinolone: a prospective, controlled multicenter trial of 396 cases. Research in Experimental Medicine. 1995;195:217-29.
21. Hernesniemi J, Troupp H. A clinical retrospective and a double blind study of betamethasone in severe closed brain injuries. Acta Neurochrurgica. 1979:(Suppl. 28):499.
22. Pitts LH, Kaktis JV. Effect of megadose steroids on ICP in traumatic coma. In: Shulman K, Marmarou A, Millar JD, et al, editors. Intracranial pressure IV. Berlin: Springer-Verlag; 1980.

23. Ransohoff J. The effects of steroids on brain edema in man. In: Reulen HJ, Schurmann K, editors. Steroids and brain edema. New York: Springer-Verlag; 1972. p. 211-3.
24. Saul TG, Ducker TB, Salcman M, Carro E. Steroids in severe head injury. Journal of Neurosurgery. 1981;54:596-600.
25. Stubbs DF, Stiger TR, Harris WR. Multinational controlled trial of high-dose methylprednisolone in moderately severe head injury. In: Capildeo R, editor. Steroids in diseases of the central nervous system. Chichester: John Wiley & Sons; 1989. p. 163-8.
26. Zagara G, Scaravilli P, Carmen Belluci M, Seveso M. Effect of dexamethasone on nitrogen metabolism in brain-injured patients. Journal of Neurosurgery. 1987;31:207-12.
27. Zarate IO, Guerrero JG. Corticosteroids in paediatric patients with severe head injury. Practica Pediatrica 1995;4:7-14.
28. Jennett B, Bond M. Assessment of outcome after severe brain damage. Lancet. 1975;1(7905): 480-4.
29. Cooper NJ, Sutton AJ, Abrams KR, Turner D, Wailoo A. Comprehensive decision analytical modelling in economic evaluation: a Bayesian approach. Health Economics. 2004;13(3):203-26.
30. Smith TC, Spiegelhalter DJ, Thomas A. Bayesian approaches to random-effects meta-analysis: a comparative study. Statistics in Medicine. 1995;14(24):2685-99.
31. Shavelle RM, Strauss DJ, Day SM, Ojdana KA. Life expectancy. In: Zasler ND, Katz DI, Zafonte RD, editors. Brain injury medicine: principles and practice New York: Demos Medical Publishing; 2007.

Appendix C

Probiotics in patients with severe acute pancreatitis (SAP)

Contents

C1. Introduction

The case study on the use of probiotics in patients with severe acute pancreatitis (SAP) is used to demonstrate the potential health consequences of performing a new study on this topic. This case study is an example of an assessment of the value of additional research in which it may be reasonable to combine existing evidence in a non-standard way. In particular, the limited evidence available and the suboptimal quality of certain previously performed studies support the notion that evidence may need to be weighted, in addition to standard weighting based on study size, to reflect aspects of quality and potential bias.

Existing studies on probiotics in severe acute pancreatitis have demonstrated improved patients outcomes as well as an increased risk of mortality. Consequently, the worth of and risk associated with a new study on probiotics have been heavily debated, but no formal assessment of the value of additional research has yet been performed. This value is assessed, for a UK setting, recognizing that evidence weighting may be required. A formal procedure for deriving evidence weights is not included here, rather, the direct impact of evidence weighting on the value of additional research is demonstrated and visualized.

C1.1 Background to the case study

Acute pancreatitis is an inflammatory process of the pancreas that is rapid in onset. Annual incidence rates for acute pancreatitis vary from around 10 to 70 per 100,000 people per country, and have been increasing in the United States and Europe.(1, 2) In around 20% of patients SAP occurs and their mortality rate may be as high as 10-30%.(3, 4) In 2005, more than 230,000 patients were treated for acute pancreatitis in hospitals in the United States, with a corresponding total annual cost of $2.2 billion.(5, 6) In the UK the annual incidence rate of acute pancreatitis is 22.4 per 100,000 persons.(7) In 2010, this amounted to around 14,000 cases of acute pancreatitis, and an estimated 2,800 patients with SAP (for a UK population estimate of 62.3 million).

Currently, evidence on the effectiveness of probiotics in the prevention of infectious complications is mixed and limited to only three studies, excluding studies of patients undergoing elective abdominal operations and studies that did not clearly report their inclusion criteria.(8-11) Two studies by Oláh et al. (2002 & 2007) reported (non-significant) improvements in outcomes, such as a lower risk of infected pancreatic necrosis, due to adding probiotics to enteral nutrition, while the study by Besselink et al. (2008) reported a significantly increased risk of mortality in the probiotics study arm. Table C1 shows the main characteristics of these three studies, including their outcomes. The conflicting results from these studies generated substantial debate on the benefits of probiotics in patients with severe acute pancreatitis and on the quality of the studies reporting these results.(12) A meta-analysis performed in 2009 stated that "Future large-scale, high-quality, placebo-controlled, double-blind trials are still required to clarify the issues of the effect of probiotic in severe acute pancreatitis", and a systematic review concluded in 2010 that "Current evidence does not support the application of immunoenhanced nutrients and probiotic supplements, and therefore none of them can be recommended in the management of acute pancreatitis at present."(13, 14) As current evidence on the effectiveness of probiotics in patients with SAP is limited and conflicting further investigation may be warranted but, at the same time, caution is needed as half the patients randomised into a new study are at risk of health loss (assuming equal allocation between treatment arms).(15, 16)

In this case study, the evidence from the three available studies is examined to assess the value of additional research on the effectiveness of probiotics in patients with SAP in the UK. The effectiveness is assessed based only on the mortality risk associated with probiotic treatment, other potentially beneficial or harmful effects are not included. Different scenarios are considered with respect to the additional weighting of the available evidence.

Table C1. Characteristics of the studies providing evidence used in this case study

Investigators Characteristics	Oláh et al.(10)	Oláh et al.(9)	Besselink et al.(8)
Study year	2002	2007	2008
Intervention	Enteral feeding with probiotics compared to enteral feeding without probiotics	Enteral feeding with probiotics compared to enteral feeding without probiotics	Enteral feeding with probiotics compared to enteral feeding without probiotics
N	45	62	296
Included patients with biliary tract disease	No	Yes	Yes
Usual care - Treated	23	29	144
Usual care - Deaths	2	6	9
Probiotics - Treated	22	33	152
Probiotics - Deaths	1	2	24
Odds ratio of			
- Death	0.50	0.25	2.81
- Infected pancreatic necrosis	0.23	0.25	1.49
- Requiring surgery	0.11	0.43	2.10
for probiotics vs no probiotics			
Randomization	Not described	Not described	Computer generated permuted block sequence
Blinding	Not described	Doubleblind	Doubleblind
Follow-up performed	No	No	Yes

C1.2 Current clinical practice and mortality rate

In this case study it is assumed that annually there are 2,800 new patients with SAP in the UK. All of these patients are assumed to receive usual care, which does not include providing probiotics given the increased risk of death demonstrated in the study by Besselink et al. (2008). Without the use of probiotics the mortality rate was 2/23, 6/29, and 9/144, in the Oláh (2002 & 2007), and Besselink (2008) studies, respectively. (8-10) In this case study, the mortality rate for usual care (no probiotics) was set to 8.7% (odds of 0.095), that is, the average mortality rate across the three studies considered.

C1.3 Setup of the analyses

All analyses are performed in R (v2.15.1) using the packages *rmeta* and *gplots*. Standard frequentist fixed effect and random effects meta-analyses are performed using the 'meta.summaries' function. This function computes a summary estimate and confidence interval from a collection of treatment effect estimates and standard errors, allowing for fixed or random effects. It is known that the Mantel-Haenszel method for fixed effect analysis and the DerSimonian-Laird method for random effects analysis may result in more accurate summary estimates, in particular for small study sizes, as these methods are based on actually observed numbers of (non)events. However, the 'meta.summaries' function also allows for optional quality weights, which are the main focus of this case study, and was therefore preferred over the other methods.[4] The 'forestplot' function (package *rmeta*) is adapted to improve visual clarity. The 'cummeta.summaries' function (package *rmeta*) is adapted to support the transfer of input weights supplied by the user to the 'meta.summaries' function.

[4] In practice, the Mantel-Haenszel and DerSimonian-Laird methods could still be used when applying quality weights, when these weights are used to adjust (recalculate) the number of (non)events in such a way that the adjusted number implicitly reflects the required weights.

C2. Standard meta-analysis

Figure C1 shows the forest plot visualizing existing evidence and the summary odds ratio (OR) of death for probiotics compared to usual care, along with 95% confidence interval (CI) derived from standard frequentist fixed and random effects analyses. In these analyses studies are weighted based only on their size. From Figure C1 it is apparent that the summary OR derived from a fixed effect analysis is very different from the summary OR derived from a random effects analysis. In fact, a fixed effect analysis provides indication of the harmful nature of probiotics whereas a random effects analysis indicates a beneficial effect of probiotics. However, both of the summary ORs are not statistically significant, indicating that the effectiveness of probiotics is still uncertain. Figure C2 shows the standard cumulative meta-analysis plots corresponding to Figure C1.

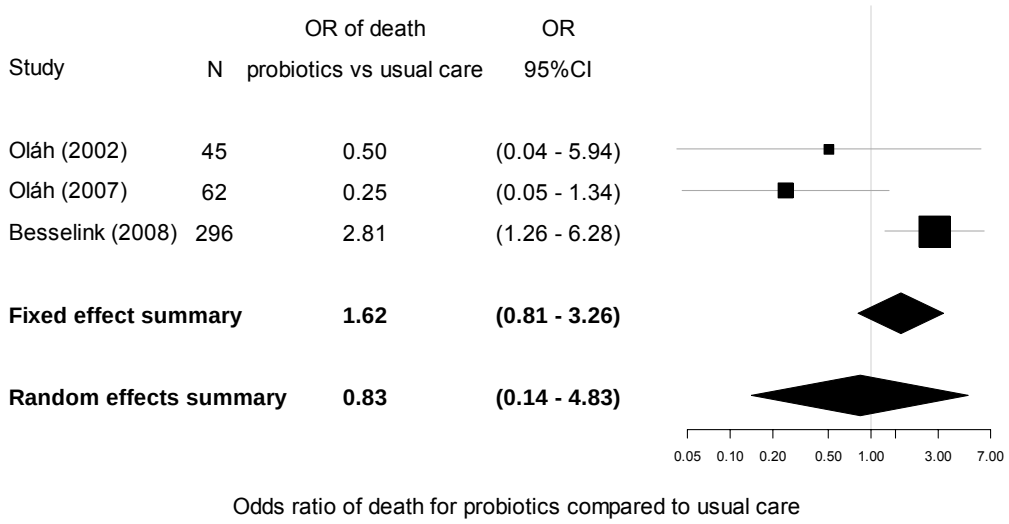


Figure C1. Forest plot visualizing existing evidence and results from standard meta-analysis

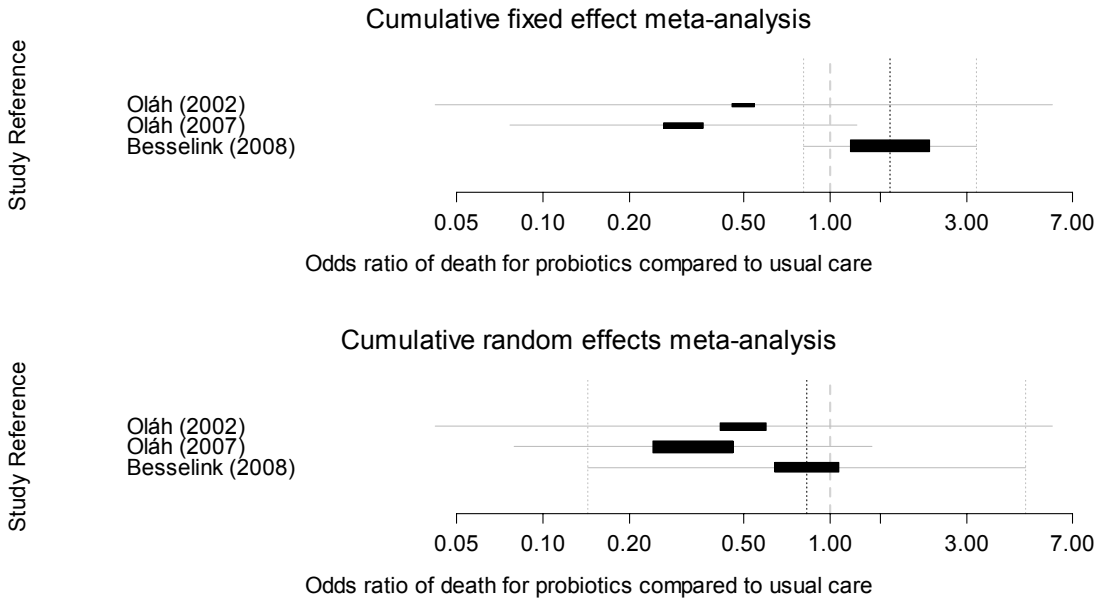


Figure C2. Standard cumulative meta-analysis plot

C3. Meta-analysis with evidence weighting

When performing a standard fixed or random effects meta-analysis the evidence collected from the included studies is weighted by the corresponding size of those studies. When studies are known or expected to vary in quality or bias, however, then additional weighting of evidence on top of weights derived from study size may be appropriate. In this case study on probiotics the study characteristics shown in Table C1 suggest that the quality of the 2002 and 2007 studies by Oláh et al. is likely to be lower than the quality of the 2008 study by Besselink et al. For example, the randomization method in the former two studies was not reported whereas the latter study used a computer-generated permuted-block sequence. In addition, the type of blinding was not clear for the Oláh 2002 study and was 'doubleblind' in the Oláh 2007 and Besselink 2008 studies. Finally, follow-up of patients to register all complications and events was performed only in the Besselink 2008 study and not in the Oláh (2002 & 2007) studies. Consequently, it may make sense to down weight the evidence from the studies by Oláh et al. (2002 & 2007) relative to the study by Besselink et al. (2008).
Evidence weighting was implemented by adjusting the weights due to study size from standard fixed and random effects analyses.[5] For each included study the standard weight may be adjusted and the resulting set of weights can then be rescaled again to sum to 1. In this case study the weights of the studies by Oláh et al. (2002 & 2007) are adjusted whereas the weight of the study by Besselink et al. (2008) is not modified.[6] Two scenarios are defined to illustrate the effect of additional weighting of evidence on the OR of death for probiotics compared to usual care.

Scenario 1. The weight of the evidence from the studies by Oláh et al. (2002 & 2007) is reduced with a factor 0.5 (that is, considered less important) relative to the study by Besselink et al. (2008) on top of weighting based on study size. This scenario acknowledges the lower quality of the former two studies compared to the latter study.

Scenario 2. The weight of the evidence from the studies by Oláh et al. (2002 & 2007) is increased with a factor 1.5 (that is, considered more important) relative to the study by Besselink et al. (2008) on top of weighting based on study size. This scenario addresses the hypothetical situation in which the latter study would be of lower quality than the former two studies.

Although these scenarios suffice to demonstrate the significance and impact of evidence weighting on assessment of the value of additional research, the weights of 0.5 and 1.5 applied here are not derived using formal procedures. Generic and formal procedures to bias modelling for evidence synthesis, of varying complexity, are, however, available and these should be used in meta-analyses whenever appropriate.(17-20)

For scenario 1 the original and adjusted weights, both unscaled and scaled, for the three studies considered are shown in Table C2. For scenario 2 the corresponding weights are shown in Table C3.

Table C2. Changes in study weights for scenario 1.

	Fixed effect analysis				***Random effects analysis***			
	Original weights		**Adjusted weights**		**Original weights**		**Adjusted weights**	
Study	Unscaled	Scaled	Unscaled	Scaled	Unscaled	Scaled	Unscaled	Scaled
Oláh (2002)	0.627	0.079	0.313	0.045	0.302	0.243	0.151	0.170
Oláh (2007)	1.347	0.170	0.673	0.097	0.408	0.328	0.204	0.230
Besselink (2008)	5.952	0.751	5.952	0.858	0.532	0.428	0.532	0.600

[5] Although the clear differences in study quality shown in Table C1 would advocate the use of a random effects analysis over the use of a fixed effect analysis in this case study, it should be recognized that the number of included studies is too low to accurately estimate the between-study variance. In a Bayesian setting, which is not considered here, accuracy may be improved through the use of a prior distribution for the between-study variance. In this frequentist setting results from both a fixed effect analysis and a random effects analysis will be shown for completeness, together with the impact of additional evidence weighting.

[6] The studies by Oláh et al. (2002 & 2007) are given similar weight adjustment based on the fact that they were conducted in the same setting and are of roughly the same quality. In general, each included study may be assigned a separate weight adjustment.

Table C3. Changes in study weights for scenario 2.

	Fixed effect analysis				*Random effects analysis*			
	Original weights		**Adjusted weights**		**Original weights**		**Adjusted weights**	
Study	Unscaled	Scaled	Unscaled	Scaled	Unscaled	Scaled	Unscaled	Scaled
Oláh (2002)	0.627	0.079	0.940	0.105	0.302	0.243	0.454	0.284
Oláh (2007)	1.347	0.170	2.020	0.227	0.408	0.328	0.611	0.383
Besselink (2008)	5.952	0.751	5.952	0.668	0.532	0.428	0.532	0.333

Figures C3 and C4 show the results of the (cumulative) meta-analysis performed for scenario 1 and have layout similar to Figures C1 and C2, respectively.

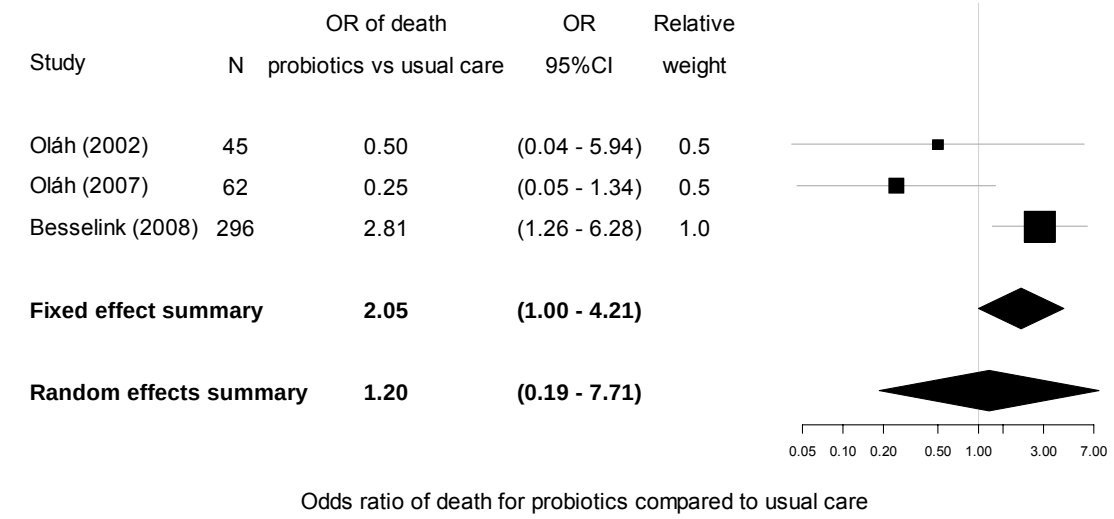


Figure C3. Forest plot visualizing existing evidence and meta-analysis results for scenario 1 in which the weight of the evidence from Oláh (2002 & 2007) is reduced with a factor 0.5

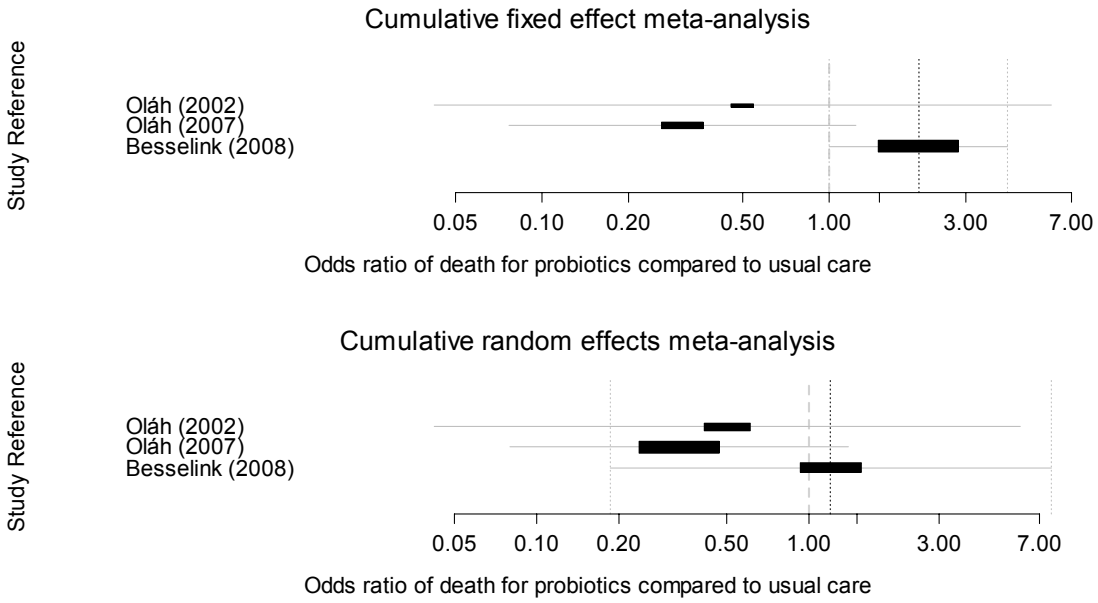


Figure C4. Cumulative meta-analysis plot for scenario 1 in which the weight of the evidence from Oláh (2002 & 2007) is reduced with a factor 0.5

In scenario 1, Figures C3 and C4 indicate that, due to the reduction in weights of the evidence by Oláh et al. (2002 & 2007), the summary OR of death for probiotics resulting from a fixed effect analysis is now (borderline) statistically significant. The summary OR from a random effects analysis, on the other hand, increases from less than 1 to more than 1, but still has a large CI encompassing the value of 1.

Figures C5 and C6 show the results of the (cumulative) meta-analysis performed for scenario 2 and also have layout similar to Figures C1 and C2, respectively.

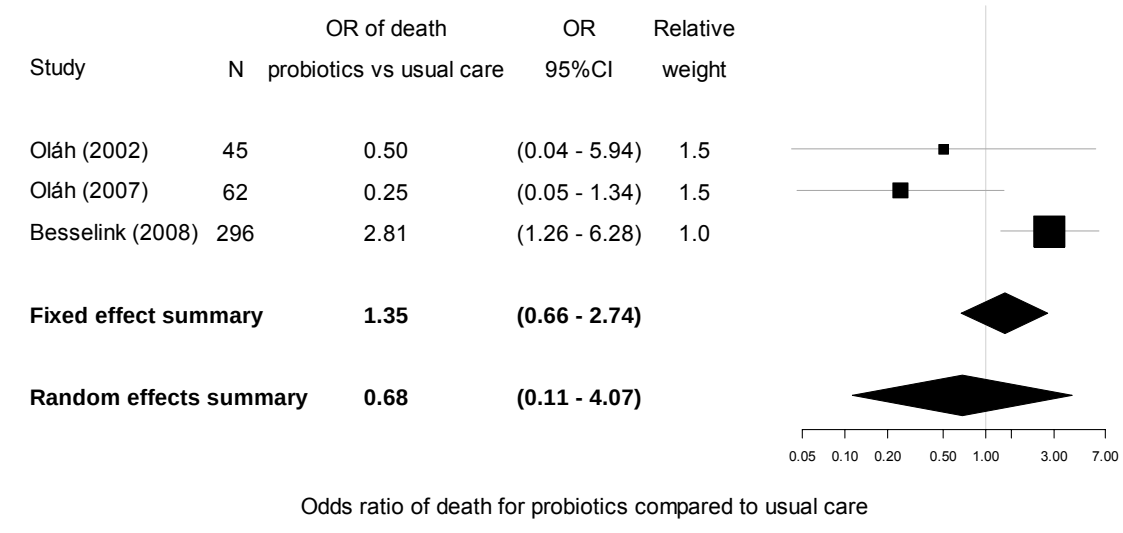


Figure C5. Forest plot visualizing existing evidence and meta-analysis results for scenario 2 in which the weight of the evidence from Oláh (2002 & 2007) is increased with a factor 1.5

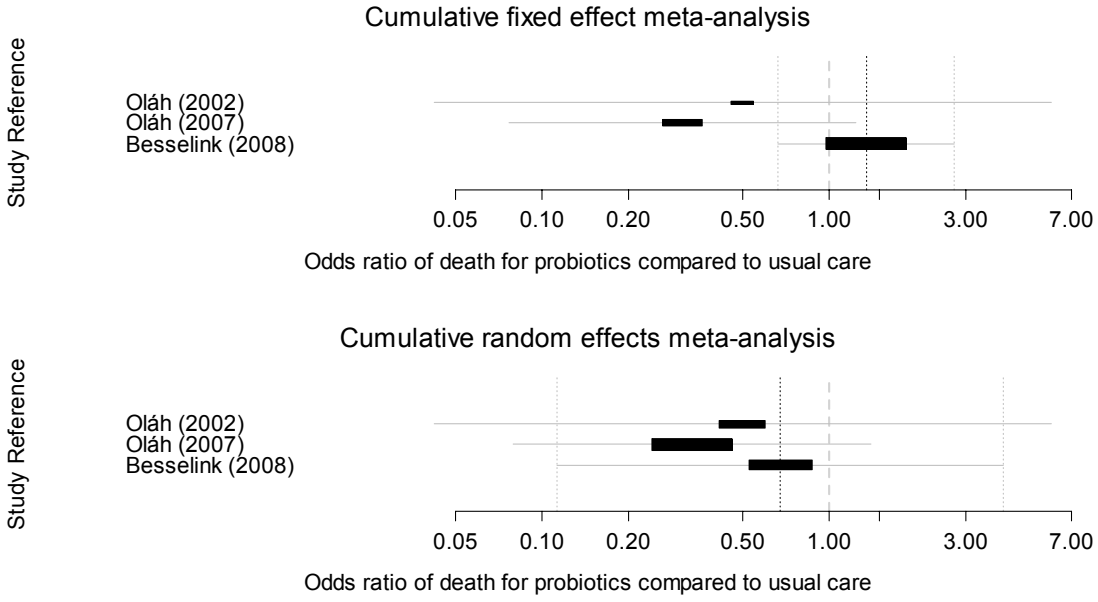


Figure C6. Cumulative meta-analysis plot for scenario 2 in which the weight of the evidence from Oláh (2002 & 2007) is increased with a factor 1.5

In scenario 2 the summary ORs of both a fixed effect and a random effects analysis decrease, with the OR from the fixed effect analysis still remaining larger than 1, and both ORs not reaching statistical significance.

In this case study, the studies by Oláh et al. (2002 & 2007) are given similar weight adjustment. Varying this weight adjustment while keeping the weight of the study by Besselink et al. (2008) fixed (unadjusted) allows calculation of the summary OR over a range of weight adjustments. Figure C7 shows the summary OR and corresponding 95% CI from a fixed effect analysis when the studies by Oláh et al. receive weight adjustments in the range of [0, 4] while the study by Besselink et al. has its weight adjustment fixed at 1. Here, a relative adjustment of 1 indicates 'no adjustment', that is, a standard fixed effect analysis. Figure C8 shows similar results, but now for a random effects analysis instead of a fixed effect analysis.

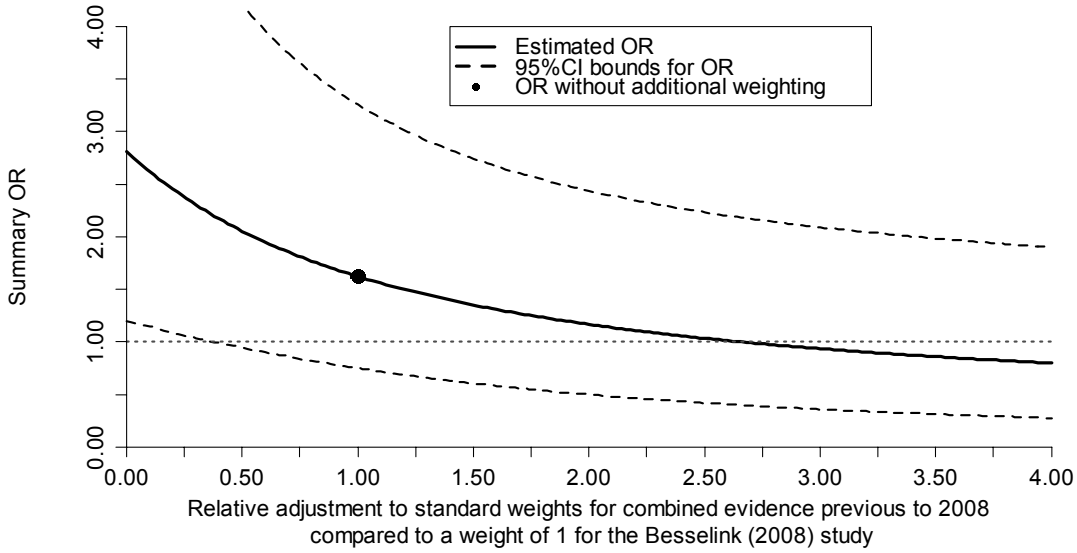


Figure C7. Impact of differential weighting of evidence on results from a fixed effect analysis

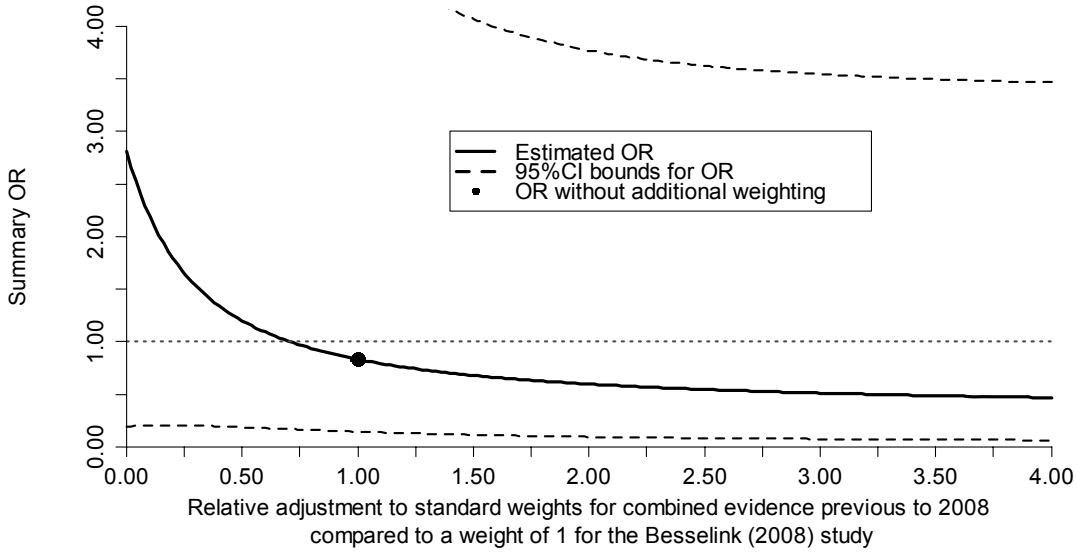


Figure C8. Impact of differential weighting of evidence on results from a random effects analysis

Figures C7 and C8 may be used to roughly assess visually for which set(s) of weight adjustments the resulting summary OR would be statistically significant greater or less than 1. When ones believes about study quality match such a set of weights then further research would not be worthwhile, assuming that a statistically significant OR is all that is required to ensure nationwide implementation of the optimal treatment.

Figure C7 confirms that if the weight of the studies by Oláh et al. is reduced by a factor of 0.5 or more, then a fixed effect analysis would return a summary OR statistically significant greater than 1, indicating that probiotics are harmful. In the range [0,4] there is no weight adjustment for the studies by Oláh et al. for which the summary OR would be statistically significant less than 1. When a random effects analysis is performed, Figure C8 shows that there is no weight adjustment in the range of [0,4] for the studies by Oláh et al. for which the summary OR would be statistically significant greater than or less than 1. When there are more than 2 (sets of) studies for which weight adjustments may be defined, visualizing the summary result corresponding to particular weight adjustment values, such as in Figures C7 and C8, may become difficult. However, it is still possible to report the combinations of weight adjustment values for which the summary OR would be statistically significant. Expert would then need to determine the extent to which certain weight adjustments can be justified, or alternatively, would need to define a plausible set of weight adjustment values prior to calculation of the summary results.

C4. Health consequences of resolving current uncertainty

As long as the summary OR of a new treatment based on all existing evidence can be greater than or less than 1 the optimal treatment of patients is not known with certainty. Hence, implementing that treatment, based on an estimated OR less than 1, will result in suboptimal health outcomes if the actual, true OR would be greater than 1. Similarly, not implementing the new treatment based on an estimated OR greater than 1, will result in suboptimal health outcomes if the actual, true OR would be less than 1. In this case study, assessing the health consequences of resolving uncertainty, in terms of the expected number of deaths averted annually in the UK, is done as follows. First, a log-normal distribution is assumed for the OR of death for probiotics, that is, $\ln(OR^{D}_{PROB}) \sim N(\mu, \sigma^2)$, with μ and σ the mean and standard deviation on the log scale. Consequently the probability density function for the OR of death for probiotics equals

$$\mathrm{PDF}^{\mathrm{OR}}(x,\mu,\sigma) = \frac{1}{x\sqrt{2\pi\sigma^2}} e^{-\frac{\ln(x-\mu)^2}{2\sigma^2}} \quad [1]$$

The probability of death given usual care (P^{D}_{USUAL}) is 8.7%, with corresponding odds ($ODDS^{D}_{USUAL}$) of 0.095, therefore, the probability of death with probiotics P^{D}_{PROB}, given the OR of death for probiotics compared to usual care OR^{D}_{PROB} equals

$$\mathrm{P^D_{PROB}} \mid \mathrm{OR^D_{PROB}} = \frac{\mathrm{OR^D_{PROB}} * \mathrm{ODDS^D_{USUAL}}}{1+\mathrm{OR^D_{PROB}} * \mathrm{ODDS^D_{USUAL}}} \quad [2]$$

In this case study it is assumed that usual care in patients with SAP does not include the use of probiotics. Therefore, treatment would not change, and health consequences would be zero, for any value of the true OR of death for probiotics compared to usual care greater than 1. Conversely, for any value of the true OR less than 1 there would be negative health consequences as patients do not yet receive optimal care, that is, probiotic treatment. The health consequences *H*, in terms of the number of deaths averted, given the probability of death with usual care and with probiotics are defined as

$$H \mid \mathrm{P^D_{USUAL}}, \mathrm{P^D_{PROB}} = \begin{cases} \left(\mathrm{P^D_{USUAL}} - \mathrm{P^D_{PROB}}\right) * IR^{SAP} & if\ \left(\mathrm{P^D_{PROB}} < \mathrm{P^D_{USUAL}}\right) \\ 0 & if\ \left(\mathrm{P^D_{PROB}} \geq \mathrm{P^D_{USUAL}}\right) \end{cases} \quad [3]$$

with IR^{SAP} the incidence rate of SAP, in this case the annual incidence rate of 2,800 for the UK. After filling in the values for $ODDS^{D}_{USUAL}$ and P^{D}_{USUAL} in equation [2] and [3] the health consequences *H* are a function only of the OR of death for probiotics compared to usual care. Given the lognormal distribution for the OR of death the total expected health consequences are now estimated as

$$E[Deaths\ averted] = \int_{x=0}^{\infty} \mathrm{PDF}^{\mathrm{OR}}(x,\mu,\sigma) * H(x) \quad [4]$$

Note that equation [4] integrates over the entire distribution of the OR of death but health consequences only accrue over the range [0,1]. In addition, equation [4] assumes that any OR of death for probiotics less than 1 would cause probiotics to be implemented in clinical practice.

C4.1 Minimal clinical difference

In clinical practice, a new treatment is likely to be introduced only when it provides substantial health benefits compared to existing treatments. Thus, a threshold is introduced for the minimum clinical difference, in terms of a minimum absolute reduction in the risk of death, below which implementation will not take place. To assess the health consequences of resolving uncertainty in view of the minimum clinical difference only equation [4] needs to be adapted. The upper bound for the OR of death for probiotics that would warrant actual implementation is estimated as

$$\text{maxOR}^{\text{D}}_{\text{PROB}} = \frac{\left(\frac{\text{P}^{\text{D}}_{\text{USUAL}} - MCD}{1-(\text{P}^{\text{D}}_{\text{USUAL}} - MCD)}\right)}{\text{ODDS}^{\text{D}}_{\text{USUAL}}} \quad [5]$$

with MCD the minimum absolute reduction in the risk of death. Now the range of the integral in equation [4] becomes limited to [0, $\text{maxOR}^{\text{D}}_{\text{PROB}}$] with $\text{maxOR}^{\text{D}}_{\text{PROB}} < 1$ for any $MCD > 0$ and equal to 1 when MCD = 0.

Figure C9 shows the impact of resolving the current uncertainty in OR of death for probiotics based on a fixed effect analysis. When the minimum clinical difference is set to 0% treatment without probiotics is assumed to be replaced by treatment with probiotics for any OR of death for probiotics less than 1. For higher thresholds, for example, a minimum clinical difference of 2%, treatment with probiotics will only be implemented if it would reduce the current mortality risk of 8.7% by at least 2% to 6.7% or less.

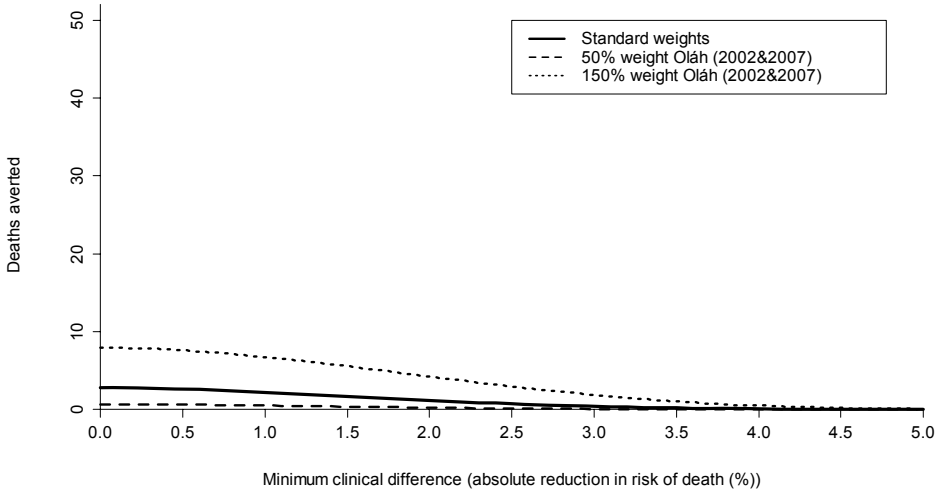


Figure C9. Potential consequences of resolving the current uncertainty in OR of death for probiotics based on a fixed effect analysis and a minimum clinical difference

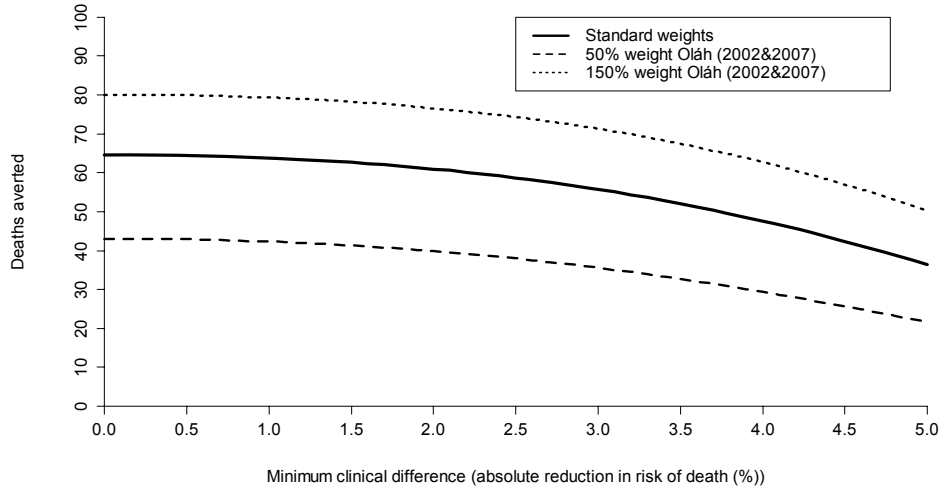


Figure C10. Potential consequences of resolving the current uncertainty in OR of death for probiotics based on a random effects analysis and a minimum clinical difference

As the implementation of probiotics becomes less likely with increasing values for the minimum clinical difference the potential health consequences of resolving current uncertainty are reduced, that is, the expected number of deaths averted annually in the UK decreases.

Any weight adjustments applied to the considered studies will change the distribution for the OR of death for probiotics and therefore also the number of deaths averted as function of the minimal clinical difference. For scenario 1, reduced weights for the evidence from Oláh et al., and for scenario 2, increased weights for the evidence from Oláh et al., the number of deaths averted is also shown in Figure C9. This Figure shows that in a fixed effect analysis the health consequences of resolving current uncertainty are zero when a minimum clinical difference of ≥ 4.0% reduction in absolute risk of death is required for actual implementation, regardless of the weight adjustments for the studies by Oláh et al. (2002 & 2007). Figure C10 has layout similar to Figure C9 and shows the expected number of deaths averted annually in the UK based on a random effects analysis. From Figure C10 it is apparent that in a random effects analysis the number of deaths averted annually in the UK ranges from 43 to 80 when the minimum clinical difference is 0%, and from 22 to 50 when it is 5.0%. Table C4 provides a summary overview of the results visualized in Figures C9 and C10. The (rounded) expected number of deaths averted annually in the UK is given, when uncertainty is resolved, for fixed and random effects analyses and standard weighting as well as weighting according to scenarios 1 and 2, for a minimal clinical difference of 0-5% absolute risk reduction in mortality.

Table C4. Potential consequences in number of deaths averted of resolving the current uncertainty in OR of death for probiotics at specific values for the minimum clinical difference

Minimum clinical difference in absolute risk of death	**Standard fixed effect**	**Fixed effect**	**Fixed effect**	**Standard random effects**	**Random effects**	**Random effects**
Weight adjustment Oláh 2002 & 2007	**NA**	**0.5**	**1.5**	**NA**	**0.5**	**1.5**
0% (any difference)	3	1	8	65	43	80
1% risk reduction	2	0	7	64	42	79
2% risk reduction	1	0	4	61	40	77
3% risk reduction	0	0	2	56	36	71
4% risk reduction	0	0	0	48	29	63
5% risk reduction	0	0	0	36	22	50

Based on the distribution of the OR of death for probiotics the likelihood of the true OR falling into any specific category of values can also be assessed, as well as the corresponding potential health consequences. Again assuming a log-normal distribution for the OR of death for probiotics, $\ln(OR^{D}_{PROB}) \sim N(\mu, \sigma^2)$, with μ and σ the mean and standard deviation on the log scale, the probability that the true OR (OR_{true}) falls between OR_{low} and OR_{high} equals:

$$P\left(OR_{low} \leq OR_{true} \leq OR_{high}\right) = \Phi\left(\frac{\ln(OR_{high})-\mu}{\sigma}\right) - \Phi\left(\frac{\ln(OR_{low})-\mu}{\sigma}\right) \quad [6]$$

with Φ the cumulative distribution function of the standard normal distribution. Specific categories of interest for the OR are predefined as (OR ≥ 1.0), (1.0 > OR ≥ 0.9), (0.9 > OR ≥ 0.8), (0.8 > OR ≥ 0.7), and (0.7 > OR), for standard and weighted fixed effect analyses and for standard random effects analysis. In addition, separate categories were defined for weighted random effects analysis as (OR ≥ 1.0), (1.0 > OR ≥ 0.8), (0.8 > OR ≥ 0.6), (0.6 > OR ≥ 0.4), (0.4 > OR ≥ 0.2) and (0.2 > OR) to better represent the wide distribution of the OR from weighted random effects analysis. Estimation of the expected number of deaths averted for each of these categories is performed by integration over the corresponding boundaries of the OR interval using equation 4.

Figure C11 shows the categorized consequences of resolving current uncertainty in the OR of death for probiotics for standard fixed and random effects analyses. The size of each bar indicating the likelihood of the true OR falling into the corresponding category is indicated (in %) at the top. The top line of the x-axis label indicates the expected number of deaths averted conditional on the true OR falling into each category. Based on three studies only, the between study variability cannot be estimated accurately in the random effects analysis, resulting in a very wide distribution for the resulting summary OR. This is reflected in the high (but probably not robust) probability of the true OR having a value less than 0.7, which is 42% for the random effects analysis and only 1% for the fixed effect analysis.

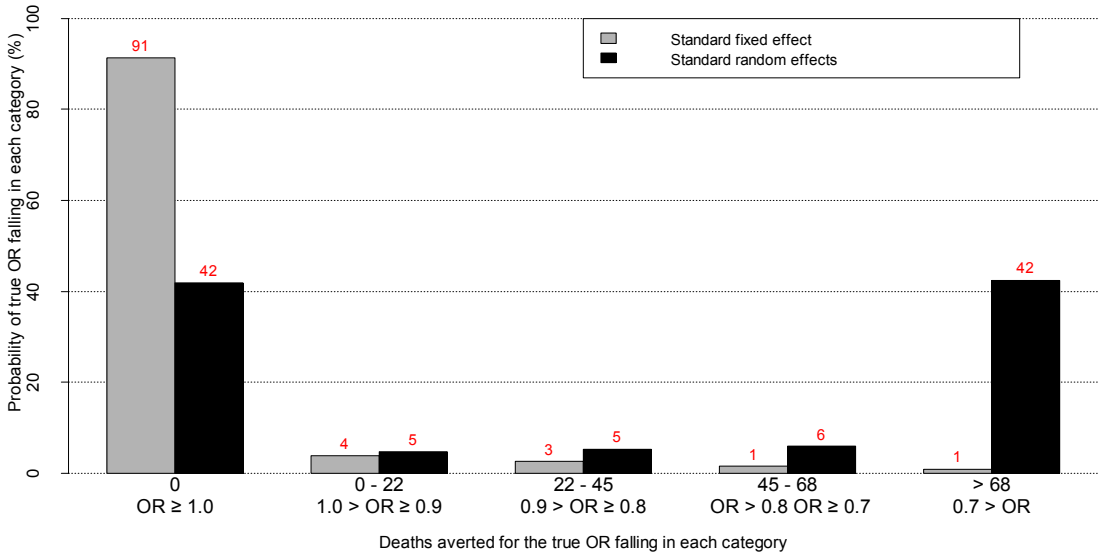


Figure C11. Categorization of the consequences of resolving current uncertainty in the OR of death for probiotics for standard fixed and random effects analyses. Overall, the fixed and random effects analyses result in respectively 3 and 65 expected deaths averted.

Figures C12 and C13 also show the health consequences of resolving current uncertainty but now for weighted fixed effect analysis (Figure C12) and weighted random effects analysis (Figure C13).

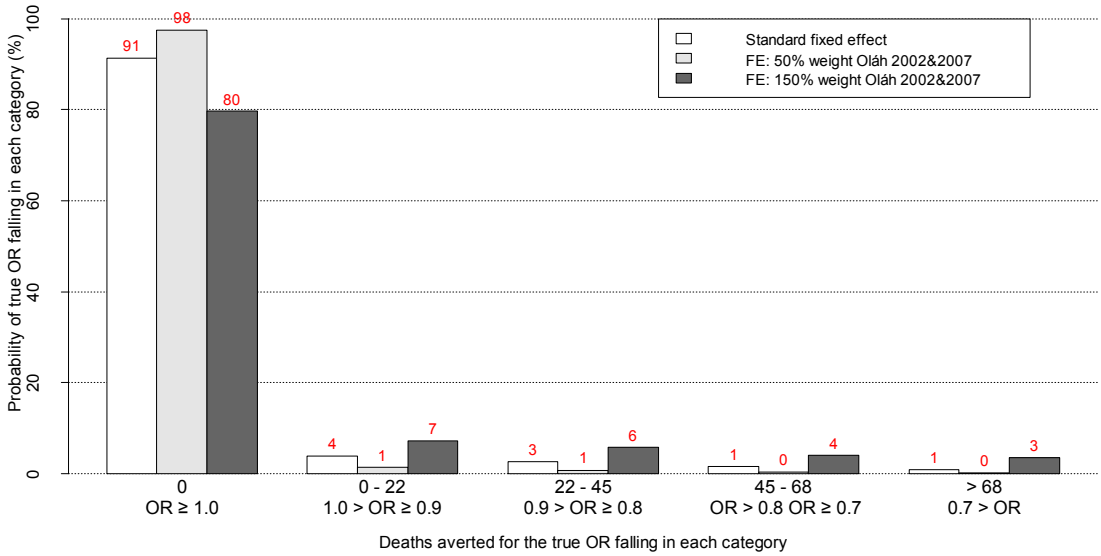


Figure C12. Categorization of the consequences of resolving current uncertainty in the OR of death for probiotics for standard and weighted fixed effect analyses. Overall, the 50% weight, standard weight, and 150% weight for the studies by Oláh et al. result in respectively 1, 3, and 8 expected deaths averted.

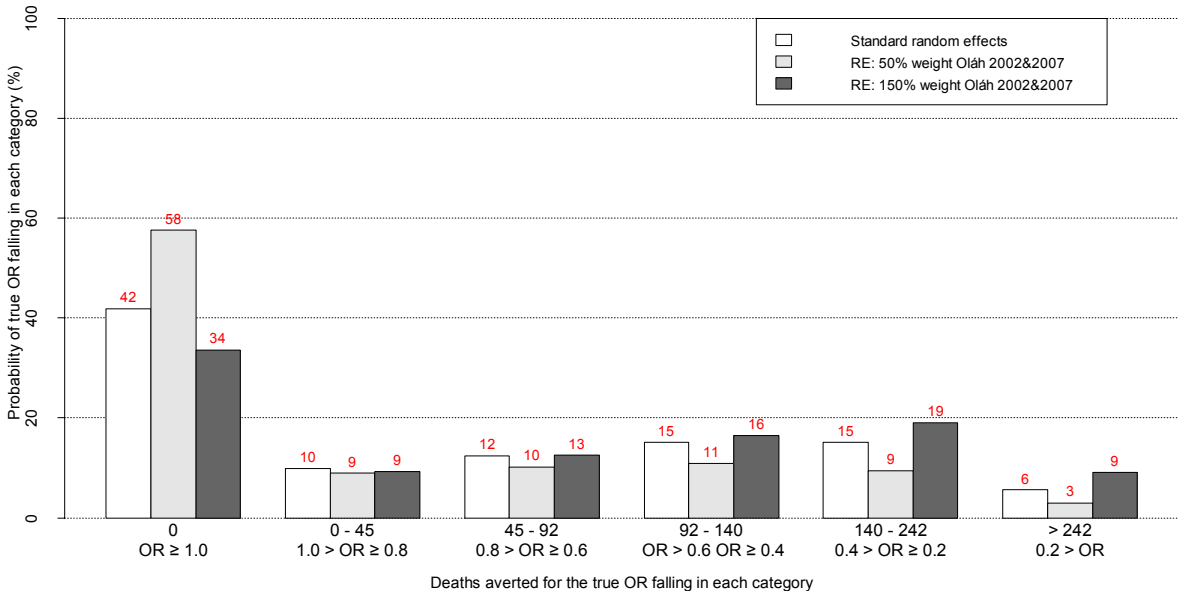


Figure C13. Categorization of the consequences of resolving current uncertainty in the OR of death for probiotics for standard and weighted random effects analyses. Overall, the 50% weight, standard weight, and 150% weight for the studies by Oláh et al. result in respectively 43, 65, and 80 expected deaths averted.

As the distribution of the OR from the random effects analysis has a mean of 0.83, standard deviation of 2.45 and mode of 0.37 on the normal scale, the probability of a low value for the true OR only decreases for OR categories including values less than 0.4 in Figure C13. According to the standard random effects analysis in Figure C13 the true OR would fall in range of, for example, 0.6 - 0.8 with probability 12% and in that case switching to treatment with probiotics would result in 45 - 92 deaths averted annually in the UK. The expected number of deaths averted aggregated over all OR categories in Figures C11-C13 matches the outcomes shown in line 1 of Table C4 (any difference is relevant).

In conclusion, the analyses performed in this case study indicate that the effectiveness of probiotics in patients with SAP is still uncertain, unless one applies a fixed effect analysis and is also willing to almost discard the evidence from the studies by Oláh et al. (2002 & 2007). Therefore, a new study on the effectives of probiotics in patients with SAP may be worthwhile. However, the actual health consequences expected from such a new study vary substantially with the type of meta-analysis performed and the weight adjustments applied to existing evidence. When a particular type of meta-analysis and specific weight adjustments (if any) have been chosen and can be justified, it should be assessed whether the expected deaths averted annually by resolving current uncertainty outweigh the expected health loss within a new study on probiotics in patients with SAP.

References

1. Lowenfels AB, Maisonneuve P, Sullivan T. The changing character of acute pancreatitis: epidemiology, etiology, and prognosis. Curr Gastroenterol Rep 2009 Apr;11(2):97-103.
2. Frossard JL, Steer ML, Pastor CM. Acute pancreatitis. Lancet 2008 Jan 12;371(9607):143-52.
3. Whitcomb DC. Clinical practice. Acute pancreatitis. N Engl J Med 2006 May 18;354(20):2142-50.
4. Forsmark CE, Baillie J. AGA Institute technical review on acute pancreatitis. Gastroenterology 2007 May;132(5):2022-44.
5. DeFrances CJ, Cullen KA, Kozak LJ. National Hospital Discharge Survey: 2005 annual summary with detailed diagnosis and procedure data. Vital Health Stat 13 2007 Dec;(165):1-209.
6. Fagenholz PJ, Fernandez-del CC, Harris NS, Pelletier AJ, Camargo CA, Jr. Direct medical costs of acute pancreatitis hospitalizations in the United States. Pancreas 2007 Nov;35(4):302-7.
7. Roberts SE, Williams JG, Meddings D, Goldacre MJ. Incidence and case fatality for acute pancreatitis in England: geographical variation, social deprivation, alcohol consumption and aetiology--a record linkage study. Aliment Pharmacol Ther 2008 Oct 1;28(7):931-41.
8. Besselink MG, van Santvoort HC, Buskens E, Boermeester MA, van Goor H, Timmerman HM, et al. Probiotic prophylaxis in predicted severe acute pancreatitis: a randomised, double-blind, placebo-controlled trial. Lancet 2008 Feb 23;371(9613):651-9.
9. Oláh A, Belagyi T, Poto L, Romics L, Jr., Bengmark S. Synbiotic control of inflammation and infection in severe acute pancreatitis: a prospective, randomized, double blind study. Hepatogastroenterology 2007 Mar;54(74):590-4.
10. Oláh A, Belagyi T, Issekutz A, Gamal ME, Bengmark S. Randomized clinical trial of specific lactobacillus and fibre supplement to early enteral nutrition in patients with acute pancreatitis. Br J Surg 2002 Sep;89(9):1103-7.
11. Li YM. Adjuvant therapy for probiotics in patients with severe acute pancreatitis: An analysis of 14 cases. World Chinese Journal of Digestology 2007;15:302-4.
12. Stapleton JR, McClave SA. Controversial results with use of probiotics in critical illness: contradictory findings from large multicenter trial. Curr Gastroenterol Rep 2009 Aug;11(4):259-62.
13. Sun S, Yang K, He X, Tian J, Ma B, Jiang L. Probiotics in patients with severe acute pancreatitis: a meta-analysis. Langenbecks Arch Surg 2009 Jan;394(1):171-7.
14. Oláh A, Romics L, Jr. Evidence-based use of enteral nutrition in acute pancreatitis. Langenbecks Arch Surg 2010 Apr;395(4):309-16.
15. Besselink MG, van Santvoort HC, van der Heijden GJ, Buskens E, Gooszen HG. New randomized trial of probiotics in pancreatitis needed? Caution advised. Langenbecks Arch Surg 2009 Jan;394(1):191-2.
16. Besselink MG, van Santvoort HC, Boermeester MA, Buskens E, Akkermans LM, Gooszen HG. Probiotic prophylaxis in acute pancreatitis: prudence required. Nat Clin Pract Gastroenterol Hepatol 2009 Mar;6(3):E3-E6.
17. Turner RM, Spiegelhalter DJ, Smith GC, Thompson SG. Bias modelling in evidence synthesis. J R Stat Soc Ser A Stat Soc 2009 Jan;172(1):21-47.
18. Spiegelhalter DJ, Best NG. Bayesian approaches to multiple sources of evidence and uncertainty in complex cost-effectiveness modelling. Stat Med 2003 Dec 15;22(23):3687-709.
19. Eddy DM, Hasselblad V, Schachter R. Meta-analysis by the Confidence Profile Method: the Statistical Synthesis of Evidence. San Diego: Academic Press; 1992.
20. Wolpert R, Mengersen KL. Adjusted likelihoods for synthesizing empirical evidence from studies that differ in quality and design: effects of environmental tobacco smoke. Statist Sci 2004;19:450-71.

Appendix D

Topotecan, PLDH and paclitaxel for second-line treatment of advanced ovarian cancer

Contents

D1. Introduction

There may be numerous alternative, mutually exclusive interventions that could be used to treat a given patient. This means that evidence from research must be applied to a decision problem in which clinicians and patients must discriminate between multiple alternatives. Where evidence is generated on the basis of a pairwise comparison of a subset of the relevant alternative interventions, this must be put into the context of a simultaneous comparison of all of the relevant alternatives in order to fully inform decisions. This may be achieved informally based on some weighing up of separate pairwise comparisons, but it is also possible to use statistical methods to make this simultaneous comparison.

This case study considers the use of chemotherapy in the treatment of advanced ovarian cancer. It considers the decision problem addressed by the National Institute for Clinical Excellence (NICE), which is reported in the technology appraisal guidance TA91 issued in May 2005 and in Main et al. 2006.(1, 2) The decisions taken by NICE are based on estimates of clinical effectiveness and cost-effectiveness. In contrast we focus only on the impact of treatment on overall survival as the final endpoint, where treatment effects are expressed as hazard ratios. At the time of the original analysis three alternative interventions were available for platinum resistant patients not previously treated with paclitaxel: topotecan, paclitaxel and PLDH. Hence clinicians and patients were faced with a choice between three alternative treatment options. The available evidence consisted of three trials that each made a different pairwise comparison, as shown in Table D1.

Table D1. Trial evidence on treatments for advanced ovarian cancer

	Treatments compared		
Trial	**Paclitaxel**	**Topotecan**	**PLDH**
039	53.0 (n=114)	63.0 (n=112)	-
30-49	-	59.7 (n=235)	62.7 (n=239)
30-57	56.3 (n=108)	-	46.6 (n=108)

Median weeks survival (number of patients analysed)

None of the trials included all three treatments. This case study illustrates how the available evidence can be used to estimate the health impacts of the three alternative treatments in a way that can inform the choice faced by patients and clinicians. The hazard ratios for death are taken from analyses of the overall patient population. The results of the three trials relating to the decision problem are as follows:

- trial 039 compared topotecan to paclitaxel (hazard ratio 0.914, 95% CI 0.68 to 1.226);
- trial 30-49 compared topotecan to PLDH (hazard ratio 1.216, 95% CI 1 to 1.478);
- and trial 30-57 compared paclitaxel to PLDH (hazard ratio 0.931, 95% CI 0.702 to 1.234).

Separate pair wise comparisons starting with trial 30-49 would indicate PLDH might be preferred to topotecan. However, a patient faced with the choice between PLDH and topotecan would also have the option of treatment with paclitaxel. The difference between paclitaxel and PLDH in trial 30-57 is not statistically significant but might favour paclitaxel. The difference between topotecan and paclitaxel in trial 039 is not statistically significant but favours topotecan. In fact, based on the point estimates we would conclude that the separate pair wise comparisons were inconsistent as they do not display transitivity. However, each hazard ratio is estimated with uncertainty and so it is not possible to come to conclusions about the extent of any inconsistency without some formal, simultaneous comparison of all three sources of evidence.

D1.1 Indirect or mixed treatment comparison for evidence synthesis

Table D1 shows that there is no common comparator between all three trials. Pooling trials on the basis of a common comparator would permit the inclusion of only two of the three studies, and the results

could be used to make an indirect comparison of the pair of treatments from the excluded study.(3, 4) Techniques for network meta-analysis such as a mixed treatment comparison would allow all evidence from all three trials to be combined.(5, 6)

Trial 30-57 was stopped early when NICE issued guidance for paclitaxel to be used as first-line treatment for advanced ovarian cancer as this led to slowed recruitment. As a consequence the length of follow up was considerable shorter in 30-57 compared to 039 and 30-49, which meant that the hazard ratio was calculated at an earlier time point. This could lead to a concern that the hazard ratio from trial 30-57 should not be pooled directly with the hazard ratios from trials 039 and 30-49 if it is believed that the hazard ratio varies over time. The options for evidence synthesis put forward to the NICE technology appraisal committee included two scenarios:

> (i) An indirect comparison on the basis of trials 039 and 30-49, effectively ignoring any evidence provided by trial 30-57.
>
> (ii) A mixed treatment comparison (MTC) on the basis of all three trials, assuming that the hazard ratio from trial 30-57 was exchangeable with that from trials 039 and 30-49.

All of the evidence syntheses were conducted using a Bayesian approach with non-informative priors.

Table D2 compares the hazard ratios reported in each clinical trial to the hazard ratios estimated in the indirect evidence synthesis (based on evidence from only two trials 039 and 30-49) and the MTC (based on evidence from all three trials 039, 30-49 and 30-57). The order of comparison is reversed to that originally reported in trials 30-49 and 30-57 and so the hazard ratios and confidence intervals are inverted.

Table D2. Hazard ratio: observed and estimated from indirect and MTC evidence syntheses

Hazard ratio (95% interval)	Observed	Indirect	MTC
Topotecan compared to paclitaxel	0.91 (0.68 to 1.23)	0.91 (0.68 to 1.22)	1.06 (0.85 to 1.33)
PLDH compared to topotecan	0.82 (0.68 to 1.00)	0.82 (0.68 to 1.00)	0.88 (0.74 to 1.05)
PLDH compared to paclitaxel	1.08 (0.81 to 1.43)	0.75 (0.54 to 1.06)	0.93 (0.75 to 1.16)

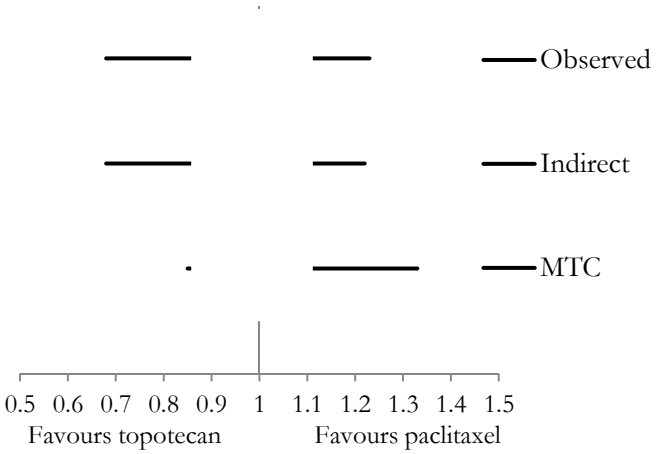
It can be seen that the indirect treatment comparison reflects exactly the hazard ratios from the two included studies, but implies that the hazard ratio for PLDH compared to paclitaxel is much lower than was observed in the omitted study 30-57. The hazard ratio of 0.75 from the indirect comparison lies outside the confidence interval for the observed hazard ratio for PLDH compared to paclitaxel. The credible interval from the indirect comparison is much wider than the observed confidence interval for PLDH compared to paclitaxel, as would be expected.(3)

The hazard ratios from the MTC do not lie outside the confidence intervals for the corresponding observed treatment effects. However, in reconciling all of the available evidence the point estimates of the hazard ratio for topotecan versus paclitaxel is estimated to be greater than 1 in the MTC when it is less than one based on the observed evidence. The hazard ratio for PLDH compared to paclitaxel is estimated to be less than 1 in the MTC when it is greater than one based on the observed evidence. In the MTC the credible interval for PLDH compared to topotecan includes 1.

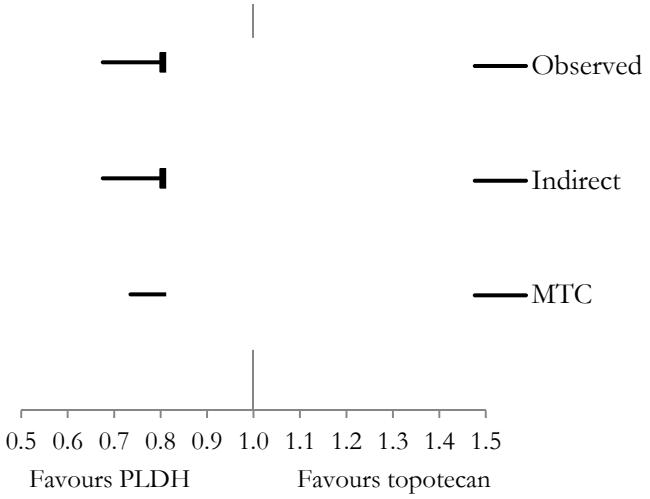
The fact that the MTC and indirect comparison produce different results is due to the fact that they incorporate different evidence. The choice presented here requires that the evidence from trial 30-57 either be ignored completely (indirect comparison) or incorporated on equal footing with the evidence provided by trials 039 and 30-49 (MTC). Another option for evidence synthesis would be to allow evidence from trial 30-57 to be incorporated with an intermediate weight, as demonstrated in appendix C.

The forest plots in Figure D1 show the data from Table D2. These are not traditional forest plots as the first point represents the observed hazard ratio. The second point represents the hazard ratio from the indirect comparison of trials 039 (topotecan compared to paclitaxel) and 30-49 (PLDH compared to topotecan). The third point represents the hazard ratio from the MTC which incorporates evidence from all three trials.

(a) Topotecan compared to paclitaxel



(b) PLDH compared to topotecan



(c) PLDH compared to paclitaxel

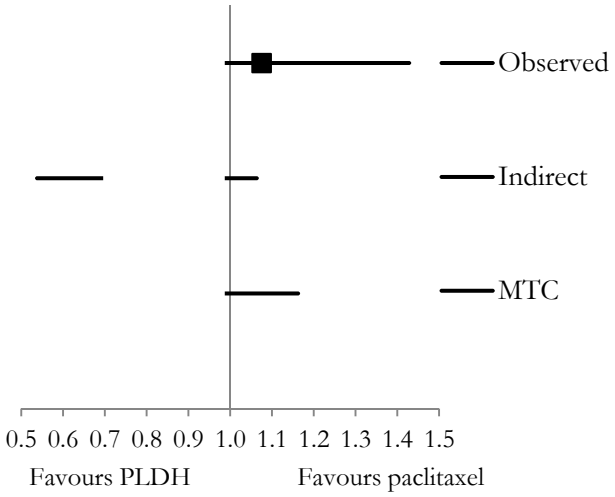


Figure D1. Forest plots comparing observed and pooled hazard ratios

D2. Information required to estimate absolute health impacts

In analyses based on the results of the evidence syntheses the baseline hazard of death is modelled using the topotecan arm from trial 30-49, which was the largest trial and reported results by each sub-group of interest in the original analysis. The hazard rate was estimated from the reported median survival by assuming that survival times followed an exponential distribution. The variance in the baseline hazard was similarly estimated using the estimated hazard and the number of events (which at the point of median survival is equal to half the number of patients (n) allocated to that arm). This is in line with the assumptions used in Main et al.(1) For the separate pairwise comparisons presented in this appendix the three different baseline hazards are estimated by the same method based on median weeks survival from the topotecan arm from trial 30-49 (59.7 weeks; n=235), the topotecan arm from trial 039 (63 weeks; n=112) and the paclitaxel arm from trial 30-57 (56.3 weeks; n=108).

The age-standardised incidence per 100,000 women was reported to be 17.9 in England and 20.6 in Wales in the year 2000. Ovarian cancer is often asymptomatic in the early stages and over 75% of cases are diagnosed with advanced disease. Between 55% and 75% of women whose tumours respond to first-line therapy relapse within 2 years of completing treatment and may be treated with second-line chemotherapy.(1) We therefore assume an incidence of 20 per 100,000 women and assume 65% are eligible for second line treatment, which gives an annual UK patient population of 8,143.

The Bayesian meta-analyses undertaken in WinBUGS provide 10,000 simulations from the posterior distributions for the baseline hazard and the hazard ratios.(7) These are used to estimate the absolute hazard of death with each of the three comparators. These hazards are then converted into a probability by assuming that the hazard is constant over one year (i.e. that it follows an exponential distribution). This probability of death is then multiplied by the number of eligible patients in order to calculate the number of deaths. In order to evaluate uncertainty in the separate pair wise comparisons we assume that the baseline hazards and the hazard ratios are normally distributed on the log scale. Using the mean and variance on the log scale to characterise Normal distributions, we draw 10,000 samples before converting back to the natural scale and calculating the number of deaths per year using the same method previously described. The indirect evidence synthesis and the MTC both provide 10,000 correlated samples of the absolute hazard of death with each treatment. Each difference pair wise comparison provides 10,000 correlated samples of the absolute hazard of death for the two directly compared treatments, and no information on the third. The number of deaths expected based on perfect information corresponding to each analysis is calculated by taking the minimum hazard from each of the 10,000 samples.

At the time that the appraisal was undertaken paclitaxel was in widespread use in the NHS. We therefore make the simple assumption that paclitaxel was current practice in the UK. However, some results are presented to allow comparisons against all three comparators in order to illustrate the impact of assuming one of the other alternative treatments is current practice.

Table D3 shows the number of deaths that would be expected based on each of the trial results separately. This shows the number of deaths expected with current information on each of the treatments (Current deaths), where current information is restricted to only the single trial that directly compared the pair of treatments in each row. The difference in the expected number of deaths between each pair of treatments (Best v worst) shows the reduction in the number of deaths that could be made by switching from the least effective to the most effective treatment on the basis of the corresponding trial. The difference in the expected number of deaths with perfect information compared to the number of deaths expected with the most effective treatment shows the maximum value of additional information to reduce uncertainty in the effectiveness of that treatment. The value of perfect information compared to the least effective treatment in each pair shows the reduction in the number of deaths that would be expected by switching to the more effective treatment and by reducing uncertainty in how effective it is.

Table D3. Number of deaths over one year based on separate pair wise comparisons

Trial	Treatment	Current deaths		Best v worst	Deaths perfect info	
		trt 1	trt 2		v trt 1	v trt 2
039	1=topotecan, 2=paclitaxel	3564	3801	237	67	305
30-49	1=PLDH, 2=topotecan	3201	3695	495	3	497
30-57	1=PLDH, 2=paclitaxel	4078	3865	213	287	74

If paclitaxel is current practice then the results of trial 039 suggest that there is value in switching implementation to topotecan (avoiding 237 deaths) and value in additional information to reduce uncertainty in the effectiveness of topotecan relative to paclitaxel (avoiding 67 deaths). The results of trial 30-57 suggest that there is no value in switching implementation to PLDH but there is value in perfect information to establish the effectiveness of PLDH relative to paclitaxel (avoiding 74 deaths). When paclitaxel is current practice the results of trial 30-49 cannot be used in isolation to inform the benefits of switching treatment to either topotecan or PLDH. The results suggest there is little value in perfect information to reduce uncertainty in the effectiveness of PLDH relative to topotecan. A crude summation of the value of perfect information for topotecan versus paclitaxel (67) and for PLDH versus paclitaxel (74) would suggest that 141 deaths could be avoided by a further trial that incorporated all three comparators.

Table D4 provides the same information as Table D3 but is based on the results of the evidence syntheses. The first column in each section (Current deaths) shows the number of deaths per year expected with each of the comparators. It can be seen that regardless of whether the indirect or MTC evidence synthesis is selected, PLDH is always expected to result in the fewest deaths. As such, the choice of method for evidence synthesis does not alter the conclusion about which treatment is expected to be most effective on the basis of current evidence. The second column in each section (Current v PLDH) compares the number of deaths expected with each treatment to the number of deaths expected with PLDH (the most effective treatment), and this is the value in switching practice from each comparator to PLDH. The third column in each section (Perfect info v current) compares the number of deaths expected with perfect information to the number of deaths expected with current information on each comparator. The comparison of perfect information to the number of deaths expected with PLDH gives the maximum value of reducing uncertainty in the effectiveness of PLDH.

Table D4. Number of deaths over one year for simultaneous comparison of all three alternatives

	Indirect			MTC		
	Current deaths	Current v PLDH	Perfect info v current	Current deaths	Current v PLDH	Perfect info v current
Topotecan	3694	496	508	3694	331	384
Paclitaxel	3953	755	766	3537	174	227
PLDH	3198	0	12	3363	0	53

The numbers differ because indirect evidence synthesis makes use of two trials, 039 and 30-49, and omits the information provided by trial 30-57. The MTC makes use of the information provided by all three trials.

Current evidence suggests that PLDH is the optimal treatment that would produce the fewest number of deaths regardless of the method of evidence synthesis adopted. The value of improving implementation to utilise the best comparator on the basis of current evidence is consistently lower for the MTC compared to the indirect evidence synthesis. The value of perfect information is estimated to be lower for the MTC compared to the indirect evidence synthesis assuming that paclitaxel is current practice. However, the value of perfect information relative to the best current practice (12 and 53 deaths avoided for the indirect and MTC respectively) is much lower than that implied by the crude summation of the

separate pair wise comparisons (141 deaths avoided). This is perhaps not surprising given that each separate pair wise comparison ignores the evidence available from two existing trials in determining the number of deaths expected with 'current' information on each comparator.

If paclitaxel is not the most effective treatment then providing paclitaxel will be associated with excess deaths . The figures in Tables D3 and D4 show the expected excess deaths, but this can be broken down to show the likelihood of observing different levels of excess deaths. Figure D2 shows the distribution of excess deaths underlying the comparison of perfect information with the number of deaths expected by providing paclitaxel in current practice, and is based on the separate pair wise comparisons.

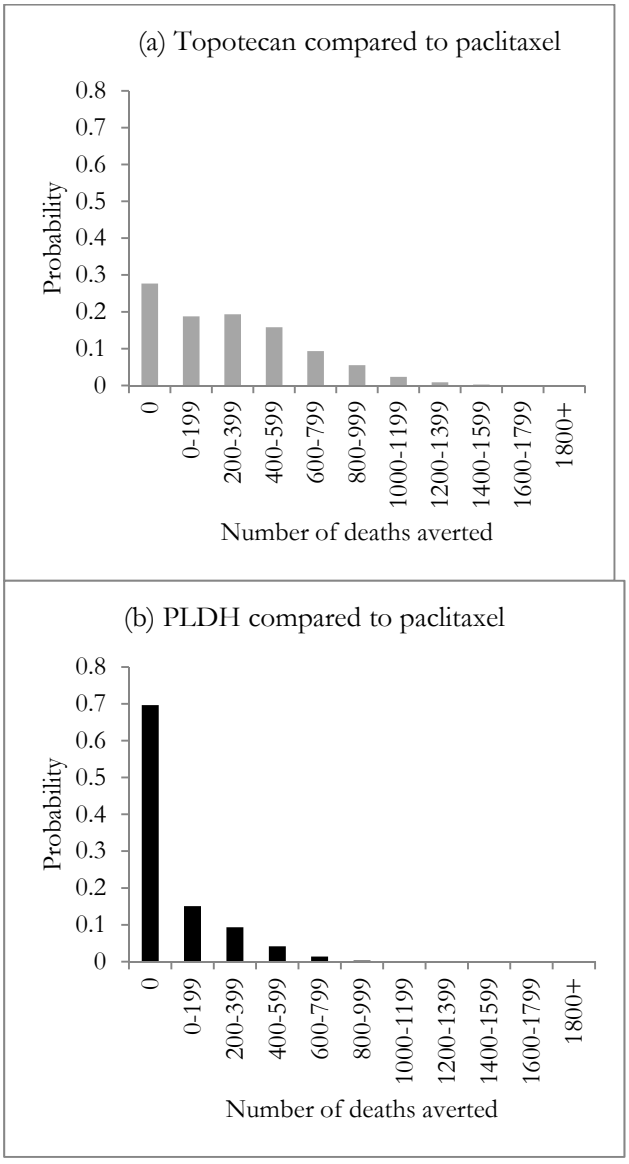


Figure D2. Consequences of uncertainty in the hazard ratio for death: separate pair wise comparisons

The comparison of topotecan to paclitaxel on the basis of trial 039 suggests that there is a 27.68% chance that paclitaxel is the most effective treatment compared to topotecan. The comparison of PLDH to paclitaxel on the basis of trial 30-57 suggests that there is a 69.65% chance that paclitaxel is the most effective treatment compared to PLDH.

Figure D3 shows the distribution underlying the comparison between perfect information and paclitaxel on the basis of the indirect evidence synthesis. The combined height of the stacked bars shows the probability that the number of deaths per year when treating with paclitaxel exceeds the number of deaths that would be observed when treating with the most effective treatment by the amount shown on the x-

axis. Each stacked bar is divided up to show which of the alternative options is the most effective treatment. There are no excess deaths only if paclitaxel is more effective than topotecan or PLDH.

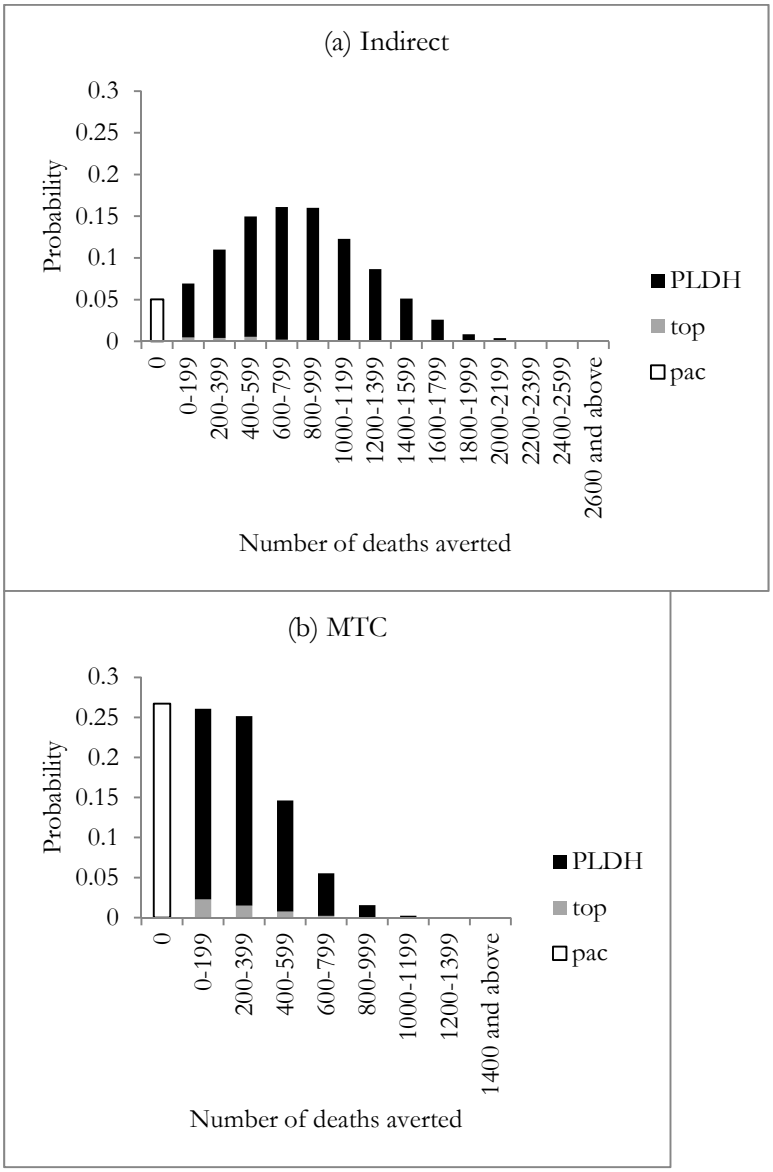


Figure D3. Histogram for consequences of uncertainty in the hazard ratio of death: evidence syntheses

The probability that paclitaxel is the most effective treatment is shown by the white bar: 5.03% on the basis of the indirect comparison and 26.72% on the basis of the MTC.

There is 94.97% probability that paclitaxel is not the most effective treatment based on the results of the indirect comparison, with a 93.12% chance that it is PLDH and a 1.85% chance it is instead topotecan. There is a 73.23% chance that paclitaxel is not the most effective treatment based on the results of the MTC, with a 68.45% chance that it is PLDH and a 4.78% chance that it is topotecan.

The choice of method for evidence synthesis does not alter the conclusion about which treatment is most likely to be the most effective on the basis of current evidence or the ranking of alternative treatments.

D2.1 Minimum clinical difference

The calculations so far have estimated the number of deaths with perfect information by taking the minimum hazard of death. Incorporating a minimum clinical difference in the calculation means that the minimum hazard is selected only if it is lower than the hazard associated with current practice by some

specified amount. Figures D4 and D5 show how the reduction in number of deaths with perfect information relative to current information and treatment with paclitaxel falls if we only switch practice from paclitaxel once the improvement in one year hazard of death with the most effective treatment exceeds ever larger minimum clinical differences. Figure D4 shows the separate pair wise comparisons on the same graph and Figure D5 shows the alternative evidence syntheses on the same graph. Figure D6 combines the evidence syntheses with a crude summation of the separate pair wise comparisons. The numbers corresponding to a minimum clinical difference of 0 are those for perfect information compared to paclitaxel in Tables D3 and D4.

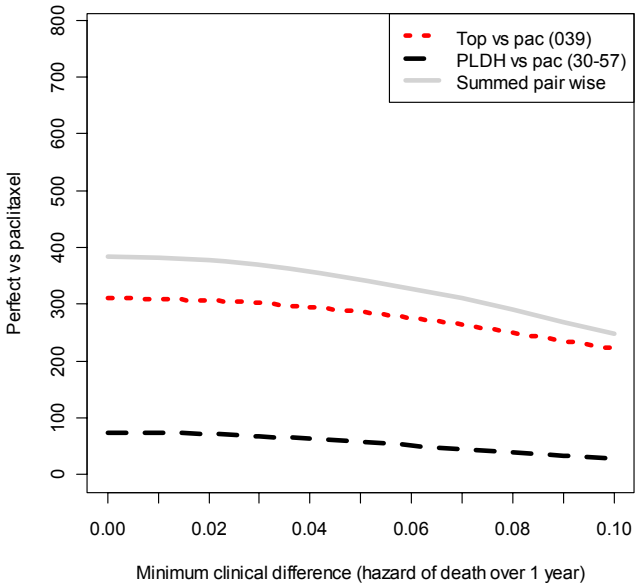


Figure D4. Deaths averted by perfect information with minimum clinical difference: separate pair wise comparisons

Figure D4 shows the number of deaths averted if current users of paclitaxel would switch to best treatment indicated by an additional trial versus topotecan (dotted line) or an additional trial versus PLDH (dashed line) if the comparator were shown to reduce the hazard of death by more than a minimum clinical difference. The estimated value of perfect information for each pair wise comparison is based only on the evidence provided by the single direct trial of each pair of comparators.

Figure D5 shows the number of deaths averted if current users of paclitaxel would switch to best treatment indicated by perfect information if it were showed to reduce the hazard of death by more than a minimum clinical difference. The estimated value of perfect information for the indirect comparison is based the evidence provided by two trials, whereas for the MTC it is based on the evidence provided by all three trials.

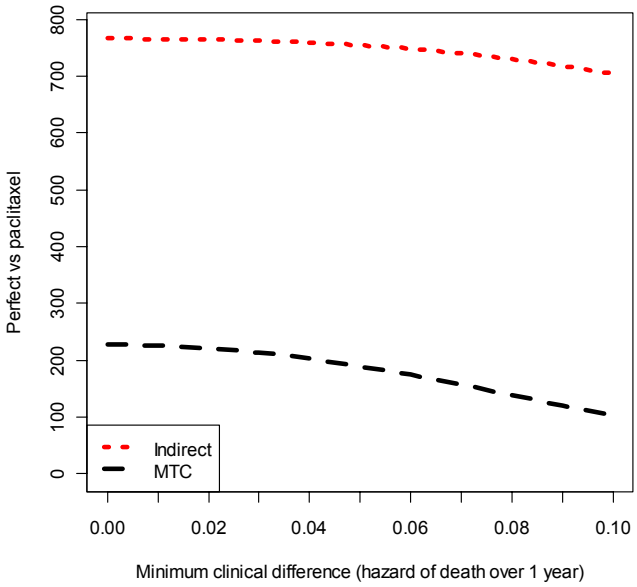


Figure D5. Deaths averted by perfect information with minimum clinical difference: evidence syntheses

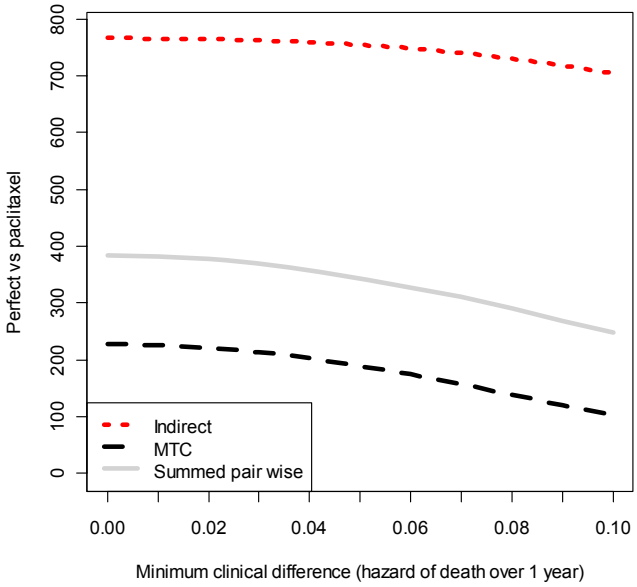


Figure D6. Deaths averted by perfect information with minimum clinical difference: combined graph

Figure D6 shows that the number of deaths averted by perfect information would be overestimated by a crude summation of the separate pair wise comparisons compared to that suggested by the MTC (which is the only methods that accounts for all of the current available evidence).

The minimum clinical difference in the reduction in the hazard of death before treatment switches from paclitaxel can also be expressed as a minimum expected reduction in the number of deaths annually. Table D5 shows how the minimum clinical difference in the absolute hazard of death over one year

translates into a minimum reduction in the number of deaths for the separate pair wise comparisons and for the alternative evidence syntheses.

Table D5. Minimum clinical difference expressed as reduction in number of deaths per year

Minimum clinical difference	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1
Pair wise: top v pac	0	43	87	132	176	222	267	314	360	407	455
Pair wise: pac v PLDH	0	43	86	130	176	219	265	310	356	403	450
Indirect	0	42	85	128	171	215	259	304	349	395	441
MTC	0	46	93	140	188	236	285	334	384	434	484

Because the baseline hazard of death when treating with paclitaxel is estimated using different information in each analysis the implied number of deaths averted annually for a given absolute reduction in the hazard of death varies. However, they are all broadly in line.

References

1. Main C, Bojke L, Griffin S, Norman G, Barbieri M, Mather L, et al. Topotecan, pegylated liposomal doxorubicin hydrochloride and paclitaxel for second-line or subsequent treatment of advanced ovarian cancer: a systematic review and economic evaluation. Health technology assessment (Winchester, England). 2006;10(9):1-132 iii-iv. Epub 2006/03/21.
2. The National Institute for Health and Clinical Excellence. TA91 Ovarian cancer (advanced) - paclitaxel, pegylated liposomal doxorubicin hydrochloride and topotecan (review): Guidance. London: NICE, 2005 25 May 2005. Report No.
3. Bucher HC, Guyatt GH, Griffith LE, Walter SD. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. Journal of clinical epidemiology. 1997;50(6):683-91.
4. Higgins JP, Whitehead A. Borrowing strength from external trials in a meta-analysis. Statistics in medicine. 1996;15(24):2733-49. Epub 1996/12/30.
5. Ades AE, Cliffe S. Markov chain Monte Carlo estimation of a multiparameter decision model: consistency of evidence and the accurate assessment of uncertainty. Medical decision making : an international journal of the Society for Medical Decision Making. 2002;22(4):359-71. Epub 2002/08/02.
6. Griffin S, Bojke L, Main C, Palmer S. Incorporating direct and indirect evidence using bayesian methods: an applied case study in ovarian cancer. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2006;9(2):123-31. Epub 2006/04/22.
7. Spiegelhalter D, Thomas A, Best N, Gilks W. BUGS: Bayesian inference using Gibbs sampling, Version 0.50. MRC Biostatistics Unit, Cambridge. 1995.