Paper 28

Power and sample size for cost-effectiveness analysis: a non-inferiority case study

Kathleen A. Boyd[1], Andrew H. Briggs[1], Paula Lorgelly[1], John Norrie[1] & Sarah Stock[2]

1. University of Glasgow
2. University of Edinburgh

Paper presented at HESG meeting, University of Sheffield, July 2009

Corresponding Author:
Kathleen A Boyd
Public Health & Health Policy
University of Glasgow
1 Lilybank Gardens
Glasgow
G12 8RZ
K.Boyd@clinmed.gla.ac.uk

Abstract

Randomised controlled trials (RCTs) which involve cost effectiveness evaluations rarely use health economic input when undertaking sample size calculations for the trial design; however, in such studies sample size calculations should be directly related to the cost-effectiveness result rather than to the effectiveness outcome alone. This paper reports on a rare case in which a clinical trial design sample size and power calculation was determined with regards to cost-effectiveness and discusses potential issues around non-inferiority trial designs in relation to economic analyses.

A cost-effectiveness decision tree model was developed to inform the sample size calculation for a proposed RCT of the effectiveness and acceptability of fetal fibronectin screening (fFN) of women with threatened pre-term labour. Cost savings due to fewer pre-term hospitalisations after the fFN test were balanced against potential morbidity and mortality concerns with false-negative results of the screening programme. A design based on hospitalisation admission rates as the superiority endpoint and complications of pre-term labour for the infant as the non-inferiority endpoint were proposed. Over 9 million trial participants would be required to prove that a policy of fFN screening increased mortality outcomes, while based on an infant morbidity outcome the sample size requirement falls to 250,000. By contrast, only 200 patients would be required to show that hospitalisation rates fall. Combining a superiority outcome in terms of hospitalisations avoided with a non-inferiority safety endpoint resulted in a final sample size of approximately 2000 women with preterm labour in order to demonstrate reduced hospitalisation without an unacceptable increase in morbidity.

Considerations of economic assessments alongside clinical trials can and should be used to guide conventional trial design. We demonstrate the feasibility of such calculations in a real life setting and demonstrate the role for economic considerations to guide non-inferiority margins.

1. Introduction

1.1 Sample size calculations for economic evaluations

Economic analyses are commonly undertaking alongside randomised controlled trials (RCTs), however, such trials are rarely designed with the economic outcomes in mind. RCTs are typically designed to show evidence surrounding the clinical effectiveness of a new treatment or health technology in comparison to the current standard of care, and the economic component is often then 'piggy-backed' (Kuntz and Weinstein 2001) on to the end of the trial. Design considerations are different for clinical and economic analyses (Briggs, 2000) and consequently economic evaluations that are conducted alongside a clinical trial with no input into the sample size calculation at the design stage of the trial are likely to be underpowered for both the cost analysis and cost-effectiveness analysis (Briggs, 2000).

If the purpose of the analysis is to inform decision making based on cost-effectiveness outcomes then sample size and power calculations should be directly related to the cost-effectiveness result rather than to the effectiveness outcome alone. Over the last decade there has been much discussion in the health economics community over alternative methods for undertaking such calculations (Drummond, 1994; Briggs & Gray, 1998; Briggs, et al. 2002; Glick, et al. 2007), and the methodology for such an approach is now well established. Therefore considerations of economic assessments alongside clinical trials can and should be used to guide conventional trial design. Despite the well established methodology, in practice, health economists are rarely given the opportunity to contribute to trial design (Briggs, et al. 2002).

This paper reports on a rare case in which a clinical and cost-effectiveness analysis trial design sample size and power calculation was determined with regards to cost-effectiveness. A brief introduction is given of the proposed RCT of fetal fibronectin screening (fFN) of women with threatened pre-term labour, followed by a discussion of the pre-trial model developed to inform the trial design calculation. The sample size calculations are then discussed. Initially standard calculations were used to estimate the required sample size based on the economic outcomes individually, in anticipation that the superiority sample size estimates would provide a good indication of the most appropriate parameters to use for the cost-effectiveness sample size calculations. The net monetary benefit (NMB) approach for estimating

sample size based on cost-effectiveness outcomes is discussed, but rejected in favour of a more appropriate non-inferiority sample size calculation to establish an acceptable power and sample size for this trial.

1.2 fFN testing for women with threatened pre-term labour (TPTL)

Pre-term births account for approximately 37% of neonatal deaths in the UK, and are also associated with both short and longer term morbidities (ISD, 2005). Respiratory distress syndrome (RDS) is a major morbidity experienced by approximately 24% of pre-term infants and is the leading cause of death in pre-term infants. There is currently no effective way of treating pre-term labour; however, administering steroids to pregnant women who are threatened with pre-term labour (TPTL) can help reduce RDS morbidity and mortality experienced by pre-term infants.

In the UK approximately 15% of pregnant women become threatened with pre-term labour and under current practice all of those TPTL women are hospitalised and receive steroids to lower the risk of RDS. Accurate diagnosis of pre-term labour is challenging, and only 20% of women suspected of TPTL actually experience premature births (ISD, 2005). This is a concern, both for the women who are unnecessarily hospitalised, and frequently transferred between maternity units causing distress to them and their families, and to the NHS which incurs a substantial cost burden. Fetal fibronectin is a diagnostic test which is easily administered and is potentially an effective means of diagnosing symptoms of pre-term labour. A research study was proposed to explore whether fetal fibronectin testing (fFN) in women threatened with pre-term labour in the UK leads to an improvement in management and resource use through reducing hospital admission, transfer and intervention rates without increasing the risk of RDS related morbidity and mortality.

In effect this is a non-inferiority trial, designed to compare a new screening intervention (fFN testing) with the current standard of care (no test) to show that the intervention is no worse in terms of clinical effectiveness than the current standard of care. While infant death from RDS is rare, threatened pre-term labour is very common, with a resulting morbidity impact on approximately 5 in 1000 pregnancies in the UK[1]. Therefore the cost-effectiveness of the intervention relative to the control is

[1] Estimated based on preliminary calculations using annual number of UK pregnancies, the probability of experiencing TPTL, the probability of giving birth pre-term in the TPTL population, and the risk of infant morbidity in the pre-term labour population.

likely to be driven by the cost and resource use outcome (hospitalisations) and potentially the infant morbidity effectiveness outcome, as opposed to the infant mortality outcome.

2. A Pre-trial model for fFN testing

Power and sample size calculations were undertaken for the fFN trial protocol with the economic outcomes in mind. A pre-trial model was developed under considerable time constraint, utilising the most readily available evidence on the costs and effectiveness of fFN testing, to inform the trial design. This pre-trial model is therefore a preliminary estimation undertaken over a short time period, giving rise to considerable uncertainty in the parameter estimates. This uncertainty may have implications for the sample size estimation, and is an issue for discussion later in this paper.

The intervention involves fFN testing of women with TPTL. Those who test positive are admitted to hospital and given steroids which reduces the risk of RDS morbidity, and mortality, while those who test negative are not admitted to hospital. The control is the current standard of care in the UK, where there is no fFN testing and women suspected of TPTL are admitted to hospital and administered steroids. It is hypothesised that the fFN screening programme will improve the accurate diagnosis of pre-term labour amongst the TPTL population, resulting in fewer pre-term hospitalisations, however, there may be potential infant morbidity and mortality concerns with false-negative results of the screening programme, i.e. those women who test negative and are not hospitalised and therefore do not receive the necessary steroid treatment. The effectiveness end points of interest are infant RDS morbidity and infant RDS mortality, while hospitalisations are used to reflect the cost outcome.

2.1 Decision Tree

A decision tree was developed to illustrate the alternative pathways in the proposed trial and probability estimates were applied based on available secondary evidence. The decision tree is illustrated in Figure 1 with a list of corresponding parameter estimates and values in Table 1.

The pre-trial model population consists of all women in the UK who are threatened with pre-term labour. The decision tree has a control arm (No test), and an intervention arm (Test). Amongst the TPTL population only 20% will actually

experience a pre-term labour, with the remaining 80% experiencing a term labour. The decision tree illustrates these pathways indicating whether or not the test was given, whether or not a pre-term labour is experienced and the resultant morbidity and mortality outcomes. The intervention branch of the tree further splits to indicate the probability of test accuracy, based on the probability that the fFN test will correctly identify pre-term labour cases (sensitivity), and correctly identify those whose labour will run to term (specificity). Those identified as pre-term by the test (correctly and incorrectly) are hospitalised and receive steroids which reduces the risk of neonatal morbidity, while those identified as term by the fFN test (correctly and incorrectly) are not hospitalised and do not receive the morbidity reducing steroids. Those who experience false negative results consequently experience greater morbidity and mortality as they do not receive the risk reducing steroids. In the control arm (No test), all women with TPTL are hospitalised and receive treatment for pre-term labour.

The probability that pre-term labour births will experience infant morbidity and mortality are included in both arms of the trial. The risk of infant morbidity in the intervention arm is reduced in those women correctly identified as pre-term due to the treatment of steroids during hospitalisation, and it is also reduced for all pre-term births in the control arm as they are all hospitalised. The proportion of pre-term labour births that were identified as term (1-sensitivity) in the test arm, do not receive steroids and therefore are subject to the standard risk of infant morbidity and mortality. The model assumes that there is no risk of morbidity or mortality to term birth.

2.2 Model parameters

Table 1 details the parameters used in the model, such as the sensitivity and specificity of the fFN test, the probabilities of hospital transfer, infant morbidity, infant mortality and morbidity risk reduction when pre-term women receive steroids. The costs that can be incurred in the model include administration of the fFN test, hospital admissions and transfers between hospitals.

A monetary amount to reflect the value of a life at birth is required for the sample size calculations. The value of a statistical life was taken to be a useful monetary indicator for the value of preventing an RDS infant mortality and was estimated to be £1million. European Union 2001 recommendations suggest that a value between €0.9 –€3.5 million (In June 2001, 1€ ~ £0.6004) should be assigned to reflect the

value of a life (Johansson, 2002). The UK Department of Transport recommend a value of £1.4million (Department of Transport, 2007) while Abelson (2003) also report similar ranges in a review of studies that have estimated the value of a life but notes that some studies have indicated that the top end of these ranges to be overestimations. Therefore a value between £500,000 and £2.5million could be assigned, and it was felt that an estimate of £1million would be a realistic reflection of the value of a statistical life. The monetary value of avoiding or preventing an infant RDS morbidity was then determined by multiplying the value of a statistical life by the probability of experiencing RDS mortality.

2.3 Pre-trial model outcomes

The mean cost of hospitalisation per TPTL was calculated using the unit costs detailed in Table 1 and the associated probabilities from the pre-trial model. The mean cost of hospitalisation was estimated to be £1414, incorporating the cost of hospital admission, the cost of transfer (multiplied by the probability of being transferred) and the cost of the fFN test (multiplied by the probability of receiving the test).

Table 2 reports the pre-trial model outcomes in terms of the proportion of clients that were hospitalised, the proportion of infants that experienced RSD morbidity, and those that experienced an RSD related death in both the intervention and control arms. Given that all TPTL women are hospitalised in the control arm, these outcomes estimate that there are nearly 10% fewer hospitalisations in the intervention arm, without compromising effectiveness. The differences in infant RDS morbidity and mortality experienced between the intervention and control are negligible at 0.06% for morbidity and 0.002% mortality. So while fFN testing reduces hospitalisations by approximately 10%, it simultaneously increases RDS infant morbidity by 0.06% and mortality by 0.002%.

3. Sample Size Calculation for trial outcomes alone

As previously discussed this is a non-inferiority trial which aims to demonstrate that the intervention or treatment is no worse in terms of the clinical outcomes than the control. The pre-trial results detailed in Table 2 indicate that this is likely to be the case. There is a substantial reduction in the number of hospitalisations, which will

have considerable cost saving implications, without an unacceptable increase in infant morbidity and mortality.

In order to estimate an appropriate sample size for the trial protocol, the sample size requirements were initially estimated for each of the model outcomes individually. Individual calculations should provide a good indication of the most appropriate outcomes to use for the cost-effectiveness sample size calculations.

The method described in detail by Briggs & Gray (1998) was used to calculate sample size requirements for a single outcome. The sample size (N) depends on the required power (β) and significance level (α) of the result and uses the hypothesised differences in terms of the effectiveness outcome (ΔE) between the treatment (E_T) and control (E_C) interventions under study and their associated hypothesised variances. The null hypothesis assumes that there is no difference between the effectiveness outcome of the two arms while the alternative hypothesis assumes that there is a difference between the two. Equation 1.1 illustrates this sample size calculation.

(1.1)

$$N > \left[\frac{Z_{\alpha/2} * Var(\Delta E.null) + Z_{\beta} * Var(\Delta E.alt)}{\Delta E} \right]^2$$

If a normal distribution is assumed (as N approaches infinity) then the variance of the change in effect of the null is equal to the variance of the change in the alternative, simplifying this calculation as detailed in equation 1.2.

(1.2)

$$N > \left[\frac{Z_{\alpha/2} * Var(E_T) + Z_{\beta} * Var(E_C)}{\Delta E} \right]^2$$

These standard sample size calculations were undertaken individually for the three economic outcomes (morbidity and mortality reflecting effectiveness and hospitalisations reflecting the cost), to estimate the sample sizes required to show a significant difference between the intervention and control for each. This provided three superiority estimates that would be required to detect a significant difference based on hospitalisations, then mortality and then morbidity alone. It was anticipated that the superiority sample size estimates would provide a good indication of the

most appropriate parameter to use for the cost-effectiveness sample size calculation. Table 3 details the parameters used for these calculations and their values, where M represents the effectiveness outcome for mortality, H represents the hospitalisation outcome and S represents the morbidity outcome.

The following equation 2.1 was used to calculate the required sample size for a power of 90% based on the mortality outcome (M), significant at the 5% level. The null hypothesis assumes there is no difference between the RDS infant mortality (M_T=M_C) of the two arms and the alternative hypothesis assumes that there is a difference between the two ($M_T \neq M_C$). $Z_{\alpha/2}$ is multiplied by the variance of the change in mortality under the null hypothesis, and Z_β is multiplied by the variance of the change in mortality under the alternative hypothesis. Equation 2.2 provide further detail, with the null using the pooled mortality value of M, and the alternative using the mortality effect under the intervention arm M_T and the control M_C. These are summed, divided by the actual difference observed and then squared.

(2.1)

$$N > \left[\frac{Z_{\alpha/2} * Var(\Delta M.null) + Z_\beta * Var(\Delta M.alt)}{\Delta M} \right]^2$$

(2.2)

$$N > \left[\frac{(Z_{\alpha/2} * \sqrt{2M(1-M)} + (Z_\beta * \sqrt{M_T(1-M_T) + M_C(1-M_C)})}{M_T - M_C} \right]^2$$

Under the null hypothesis M reflects the pooled values for mortality as it assumes there is no difference, while M_T represents the proportion of mortalities from the intervention arm (Test), and M_C the proportion of mortalities from the control arm (No test). This equation was solved using the values detailed in Table 3 to estimate the required sample size based on a significant difference in mortality. The calculation was then repeated solving for hospitalisations (using the H proportions) and for morbidity (using the S proportions). The sample sizes derived from each of these calculations are detailed in Table 4.

Using this approach it can be seen that the mortality outcome measure predicts an enormous and unrealistic sample size requirement of over 9 million participants to show a significant difference between the intervention and control. The sample size based on the morbidity measure is much smaller but also unrealistic at over 247,000;

while the hospitalisation measure provides a more reasonable sample size requirement of just less than 200 participants.

These outcomes indicate that it is possible to base the sample size calculation on the hospitalisation measure; however, this is only an indicator of cost and does not provide any information about safety or mortality outcomes. It is clear from the Table 2 and Table 4 results that the mortality outcome will not be of use due to the negligible difference in the intervention and control arms, however, the morbidity outcome measure indicates a slightly larger difference between the two arms (0.06%) and could potentially be a meaningful measure of effectiveness. Ultimately we are concerned with calculating the sample size based on the cost-effectiveness outcome rather than the cost or effectiveness outcomes individually, however, the individual calculation results detailed in Table 4 indicate that the hospitalisation and morbidity outcomes can be used in the cost-effectiveness sample size calculations.

4. Sample Size Calculations for cost-effectiveness

4.1 NMB approach

Over the last decade there has been much discussion on appropriate methods for calculating sample sizes based on expected cost-effectiveness outcomes. Most typically this can be done following the net monetary benefit (NMB) approach, which is a re-arrangement of the incremental cost-effectiveness ratio calculation incorporating a monetary value of willingness to pay. The sample size and power calculations can be calculated based on the expected change in NMB, i.e. the change in monetarised effect minus the change in cost between the two alternatives. Briggs, et al. (2002) provides a thorough explanation of the theory and equations, which are briefly summarised here.

The sample size depends on the required power (β) and significance level (α) of the results and uses the hypothesised differences in cost (ΔC) and effect (ΔE) between the treatment and control interventions under study and their associated hypothesised variances and covariance along with a set monetary willingness to pay value (λ). The null and the alternative hypotheses are used, the null assuming there is no difference between the costs and effects of the interventions, and the alternative assuming there is a difference between the costs and effects. Equation 3.1 illustrates the standard equation to show a hypothesised NMB as significant,

which is then re-arranged incorporating the standard NMB expression (equation 3.2) and its variance (equation 3.3). As illustrated previously in equation 1.2, if a normal distribution is assumed (as N approaches infinity) then the variance of the change in NMB of the null is equal to the variance of the change in the NMB of the alternative. The variance expressions for both costs and effects are divided by the required number of participants in the control and treatment arms and equation 3.3 can then be rearranged to solve for the required sample size (N) to show a significant difference in NMB (equation 3.4)[2].

(3.1)

$$NMB > Z_{\alpha/2} \ * Var(\Delta NMB.null) + Z_{\beta} * Var(\Delta NMB.alt)$$

(3.2)

$$NMB = \lambda(\Delta Effect) - \Delta Cost$$

(3.3)

$$VarNMB = \lambda^2 * Var(\Delta E) + Var(\Delta C) - 2\lambda \operatorname{cov}(\Delta E, \Delta C)$$

(3.4)

$$N > \left[\frac{(Z_{\alpha/2} * \sqrt{\lambda^2 Var(E.null) + Var(C.null) - 2\lambda * \operatorname{cov}(Enull, Cnull)} + (Z_{\beta} * \sqrt{\lambda^2 (Var(E.alt) + Var(C.alt) - 2\lambda * \operatorname{cov}(E.alt, C.alt))})}{\lambda(\Delta E - \Delta C)} \right]^2$$

This is the standard methodological approach for sample size calculations on cost-effectiveness, based on the expected change in net monetary benefit (NMB), i.e. the change in monetary effect minus the change in cost between the two alternatives. In the case of the fFN trial, it may be more appropriate to use a non-inferiority approach, i.e. specifying a non-inferiority margin based on morbidity, rather than using the standard NMB approach for sample size calculation. Table 2 indicates a negligible difference between the treatment and control arms for the mortality outcome and a small difference for the morbidity outcome; while the results of the individual sample size calculations in Table 4 established that mortality is likely to be an inappropriate outcome for deriving a realistic sample size. Due to the non-inferiority characteristics of the effectiveness outcomes, the NMB approach is unlikely to be appropriate and therefore it was decided that a non-inferiority approach based on the morbidity outcome should be taken. This approach will combine the superiority outcome in

[2] Briggs & Gray (1998) and Briggs et al. (2002) provide a thorough discussion of the methods and equations used to derive these calculations.

terms of hospitalisations avoided with the non-inferiority morbidity endpoint, where the non-inferiority margin is calculated from the cost-savings that might accrue.

4.2 Non-inferiority approach

The non-inferiority (NI) margin is the pre-specified amount by which the treatments non-inferiority can be judged in comparison to the control. The NI margin for morbidity will be determined from the cost savings that accrue through reduced hospitalisations in the intervention arm. The difference in hospitalisations between the intervention and control arms is multiplied with the cost of hospitalisation, to give the cost savings which accrue through fewer unnecessary hospitalisations in the fFN test arm. This is then divided by the value of avoiding infant morbidity to provide the non-inferiority margin. Table 3 details the values used in the NI margin for morbidity equation 4.1.

(4.1)

$$\text{NI margin} = (H_1 - H_0) * \frac{C}{L_s}$$

Once calculated, this specified non-inferiority margin can then incorporated into the basic sample size calculation, utilising the morbidity outcomes (S probabilities) from Table 3. Equation 4.2 details the sample size calculation based on the morbidity non-inferiority margin, and is expanded in equation 4.3 to detail the variance under the null and alternative hypotheses.

(4.2)

$$N > \left[\frac{Z_{\alpha/2} * Var(\Delta S.null) + Z_{\beta} * Var(\Delta S.alt)}{\Delta S + \text{NI margin}} \right]^2$$

(4.3)

$$N > \left[\frac{(Z_{\alpha/2} * \sqrt{2S(1-S)} + Z\beta * \sqrt{S_T(1-S_T) + S_C(1-S_C)}}{(S_T - S_C) + \text{NI margin}} \right]^2$$

The resultant non-inferiority sample size for morbidity is detailed below in Table 5, indicating the variation in sample size requirements to changes in the power specification. Table 5 shows that when the non-inferiority approach is used with the morbidity outcome measure, a realistic sample size in the region of 2000 participants

are required. In this way the fetal fibronectin testing has been designed with the cost-effectiveness outcome in mind and is adequately sampled to show a significant difference between the two arms based on the non-inferiority of the morbidity outcome.

5. Points for Discussion

This real life application of a sample size calculation based on a cost-effectiveness outcome demonstrates the feasibility of this approach, leading to appropriately sampled trials with respect to the cost and cost-effectiveness outcomes. As has been demonstrated the standard net monetary benefit calculation can be re-arranged to derived the required sample size with respect to a significant difference in NMB; but it can also be adapted to incorporate a non-inferiority margin for studies which involve a non-inferiority outcome. Some points of interest for discussion are as follows:

5.1 Uncertainty

The design and development of trial protocols are typically undertaken in short periods of time in order to meet protocol submission deadlines, resulting in potential uncertainty in the parameter estimates and sample size calculations. In the fFN case study discussed in this paper a pre-trial model was developed over a short time period utilising the most readily available secondary evidence in order to undertake sample size calculations. Naturally there is considerable uncertainty surrounding the parameter point estimates in the pre-trial model, and therefore there will also be uncertainty surrounding the sample size estimates. What are the implications of sample size uncertainty? The role of uncertainty in sample size calculations has been given some attention in published literature and some have proposed the use of probabilistic sensitivity analysis to deal with this uncertainty (O'Hagan, T.). While this is of considerable value, the time constraints imposed by protocol submission deadlines may interfere and dealing with this uncertainty may only be feasible after protocol submission.

5.2 Recruitment Feasibility

There is scope for assessing how variations in required sample sizes will affect recruitment costs and feasibility. As demonstrated in the fFN case study, Table 2, using mortality as an effectiveness endpoint was unrealistic, due to the enormous

sample size required to show a statistical difference between the intervention and control.

The feasibility and cost of recruitment may also become an issue if a regulatory or funding body dictate what the power endpoint of interest should be for a trial. This was not the case in the fFN trial design, and morbidity could be used as an acceptable non-inferiority endpoint where mortality was found to be inappropriate; however this may become an issue in other interventions if a pre-specified endpoint yields an unrealistic sample size.

5.3 Value of a statistical life

The value of a statistical life was used in the fFN pre-trial model to reflect the value of preventing an RDS infant mortality, and used to derive the value of preventing an RDS infant morbidity. This value was set at £1million, however there is considerable debate regarding the approaches used to derive this value (Guria, et al. 2005) and the resultant monetary amount that appropriately reflects this value (Department of Transport, 2007; Abelson 2003). Indeed, it may be appropriate to incorporate the monetary value of a QALY (Dolan, et al. 2008; Mason, et al. 2008) rather than the monetary value of a statistical life. Different amounts used to reflect the value of a statistical life will generate different sample sizes. The uncertainty surrounding which approach should be used to determine this value therefore has uncertainty implications for the sample size estimate.

References

Abelson, P. (2003) 'The value of life and health for public policy', *The Economic Record*, Vol. 79, Special Issue S2-S13

Briggs, A.H. & Gray, A. (1998) 'Power and sample size calculations for stochastic cost-effectiveness analysis', *Medical Decision Making,* Vol. 18, S81

Briggs, A.H. (2000) 'Economic evaluation and clinical trials: size matters', *BMJ*, Vol. 321, pp. 1362 – 1363.

Briggs, A.H., O'Brien, B. & Blackhouse, G. (2002), 'Thinking outside the box: Recent advances in the analysis and presentation of uncertainty in cost-effectiveness studies', *Annual Review of Public Health*, Vol. 23, pp. 377-401

Department of Transport (2007), 'Highway Economics Note no.1: 2005 Valuation of the Benefits of Prevention of Road Accidents and Casualities'. Available online at: http://www.dft.gov.uk/pgr/roadsafety/ea/pdfeconnote105.pdf

Dolan, P., Metcalfe, R., Munro, V., & Christensen, M. (2008), 'Valuing lives and life years: anomalies, implications, and an alternative', *Health economics, Policy & Law*, Vol. 3, pp. 277-300

Drummond, M. (1994), *Economic analysis alongside clinical trials*, Dept of Health, London.

Glick, H. Doshi, J., Sonnad, S. & Polsky, D. (2007), *Economic Evaluation in Clinical Trials*, Oxford University Press, Oxford.

Guria, J., Leung, J., Jones-Lee, M. & Loomes, G. (2005) 'The willingness to accept value of statistical life relative to the willingness to pay value: evidence and policy implications', *Environmental and Resource Economics*, Vol. 32, pp. 113-127

Information Services Division (2005), *Births and Babies,* NHS National services Scotland. Available online at: http://www.isdscotland.org/isd/1022.html

Johansson, P. (2002), 'On the definition and age-dependency of the value of a statistical life', *The Journal of Risk and Uncertainty*, Vol. 25, No. 3, pp. 251–263

Kuntz, K., & Weinstein, M. (2001), *Modelling in economic evaluation*. In 'Economic Evaluation in Health Care: merging theory with practice', ed. M. Drummond & A. McGuire, pp.141-171. Oxford: Oxford University Press

Mason, H., Jones-Lee, M. & Donaldson, C. (2008), 'Modelling the monetary value of a QALY: A new approach based on UK Data', *Health Economics*, Early View online publication 14th October 2008.

National Institute for Health & Clinical Excellence. 2008. Guide to Methods of Technology Appraisal. UK: National Institute for Health & Clinical Excellence. Available online at: http://www.nice.org.uk/media/B52/A7/TAMethodsGuideUpdatedJune2008.pdf

Table 1: Pre-trial model parameter estimates

Item	Unit	Description
COSTS		
C. Test	£50	Cost of administering test
C. Admission	£1,068	Maternity Inpatient cost per stay (av 2.2days) incl drugs/treatment
C. Transfer	£1,000	Estimate to reflect NHS cost of transfers between different hospital CLUs
C.saving life	£1,000,000	Statistical value of a life at birth
C.saving morbidity avoided	£25,700	Statistical value of avoiding morbidity
PROBABILITIES		
ProbPreterm	0.20	20% of TPTL. Prevalence is 8% amongst entire pop of pregnancies
fFN test Sensitivity	0.82	Average sensitivity from 7, 14 & 21 days sensitivity outcomes
fFN test Specificity	0.79	Average specificity from 7, 14 & 21 days sensitivity outcomes
R Transfer	0.32	Risk of TPTL admittied patients being transferred to another hospital
Pmorbid	0.24	Probability of respiratory related morbidity in Preterm births (ICU admissions only)
Pmortal	0.026	Probability of respiratory related mortality in Preterm births
RRs	0.54	Relative Risk morbidity reduction with steroids
ProbTest.fFN	0.50	Probablity of getting test - randomised

Table 2: Pre-trial model outcomes

	Hospitalised **Prop Hospitalised**	Mortality **Mortality**	Morbidity **Morbidity**
Test (intervention)	0.05022	0.0001170	0.00455
No test (control)	0.15000	0.0001016	0.00395
Difference	-0.09978	0.00002	0.00060

Table 3: Sample size calculation parameters

Abbreviation	Value	Description
M	0.000109289	Mortality outcome pooled
M_T	0.000116991	Mortality outcome Intervention
M_C	0.000101587	Mortality outcome Control
H	0.10011	Hospitalised outcome pooled
H_T	0.05022	Hospitalised outcome Intervention
H_C	0.15	Hospitalised outcome Control
S	0.00425	Morbidity outcome pooled
S_T	0.00455	Morbidity outcome Intervention
S_C	0.00395	Morbidity outcoome Control
L	£1,000,000	Value of a life saved
Ls	£25,700	Value of avoiding morbidity
C	£1,414	Cost Hospitalisation
Power	90%	Beta
Significance	5%	Alpha
$Z_{\alpha/2}$	1.96	
Z_{β}	1.28	

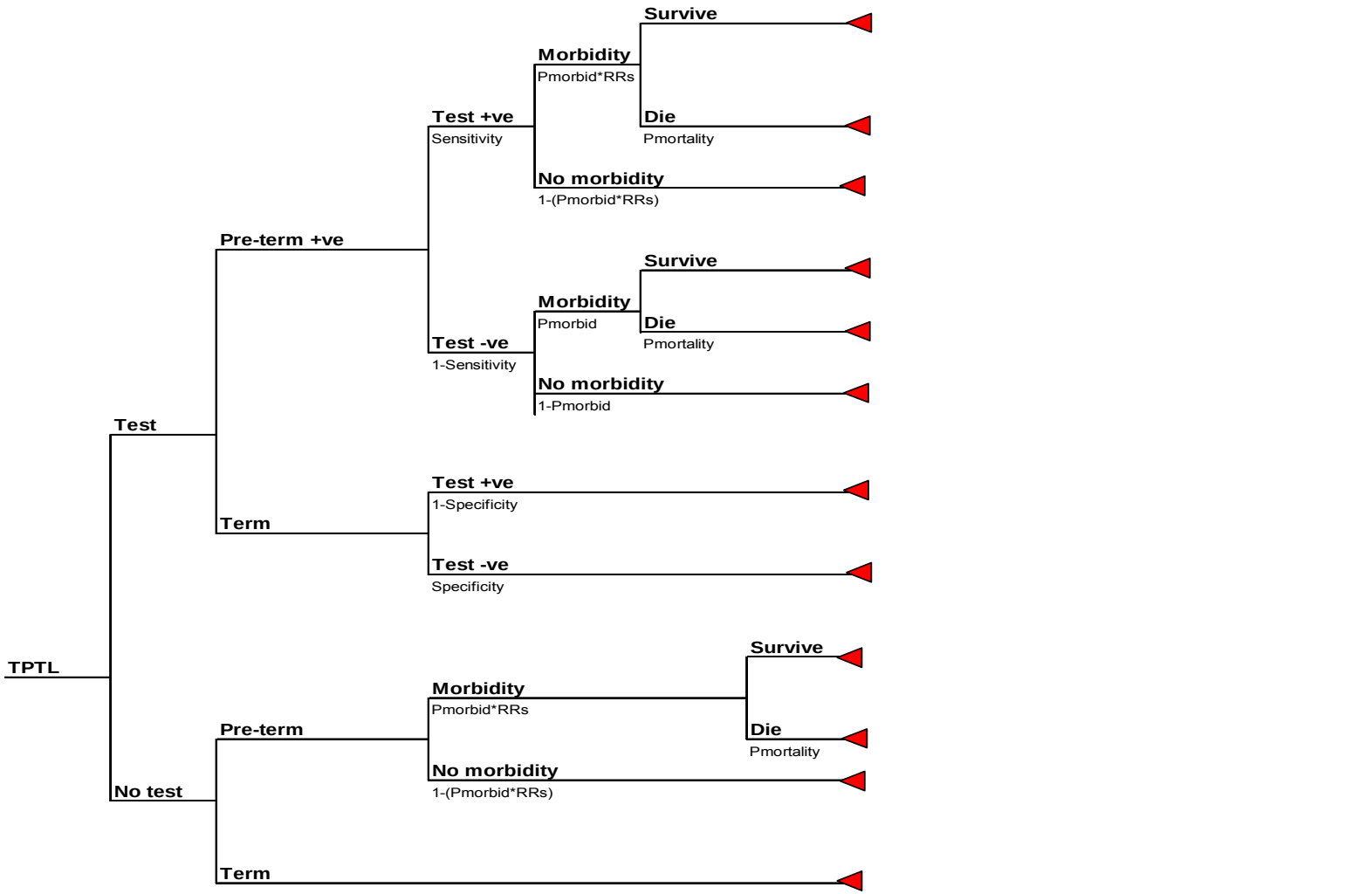
Table 4: Standard power calculations sample sizes

	Superiority sample size (N)
Mortality (M)	9678005
Morbidity (S)	247706
Hospitalisation (H)	188

Table 5: Non-inferiority power calculation sample sizes

Power	Non-inferiority sample size (N)
90%	2400
85%	2051
80%	1793

Figure 1


Hospitalisations
RDS morbidity
RDS mortality
TPTL
Test
No test
Pre-term +ve
Term
Pre-term
Term
Test +ve
Sensitivity
Test -ve
1-Sensitivity
Test +ve
1-Specificity
Test -ve
Specificity
Morbidity
Pmorbid*RRs
No morbidity
1-(Pmorbid*RRs)
Morbidity
Pmorbid
No morbidity
1-Pmorbid
Morbidity
Pmorbid*RRs
No morbidity
1-(Pmorbid*RRs)
Survive
Die
Pmortality
Survive
Die
Pmortality
Survive
Die
Pmortality