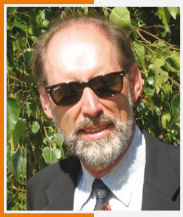
About the Author



Andrew R Willan

Andrew R Willan is an academic biostatistician and clinical trial methodologist, and currently holds the positions of Senior Scientist at SickKids Research Institute and Professor of Biostatistics in the Dalla Lana School of Public Health at the University of Toronto (ON, Canada). His contributions to statistical methodology include publications in the areas of cost–effectiveness analysis, value of information methods, management trials, crossover trials, non-nested regression analysis and bivariate response models.

Chapter 8

Bayesian methods for the design, analysis and interpretation of clinical studies

Andrew R Willan

Bayesian approaches to the design, analysis and interpretation of clinical studies have important advantages. They provide a transparent means of combining prior evidence with new data. They also allow direct statements of inference to be made regarding model parameters that are simple, relevant and intuitive. Furthermore, Bayesian methods provide the basis for value of information methods, which address questions such as: is the current evidence in favor of a new health intervention sufficient for its adoption and, if not, what is the optimal design for future research? In this chapter the author emphasizes the importance of these advantages for clinical research and decision-making. The author briefly discusses the issue of selecting prior evidence, arguing that the selection is subjective and that the subsequent inference should be considered conditional on the prior evidence selected. The author gives examples that illustrate how Bayesian methods provide statements of inference that are superior to those provided by frequentist methods. Lastly, the author gives a brief introduction to value of information methods.

doi:10.4155/EBO.13.431

Over the past 30 years, the pressure for healthcare providers to base their practice on empirical evidence has increased sharply. This has led to an enormous increase in the number of publications of clinical studies, and to the ever-expanding Cochrane database of systematic reviews. Such evidence is meant to inform clinicians regarding the relative effectiveness of competing healthcare interventions and play a role in treatment decision-making. The vast majority of clinical studies are designed and analyzed using frequentist statistical principles. The current author argues that the alternative Bayesian approach has a number of advantages that are especially relevant for the planning, analysis and interpretation of clinical research. The arguments in favor of a Bayesian approach are directly related to the Bayesian definition of probability. The first advantage, and the one that is immediately associated with Bayesian analyses, is the ability to incorporate prior knowledge transparently through the application of Bayes' theorem, which itself is an irrefutable statement of conditional probability, namely: $Pr(A|B) = Pr(B|A)Pr(A)/Pr(B)$. Using the Bayesian definition of probability, A can be a statement regarding model parameters while B is a statement regarding the observed data. Therefore, by applying Bayes' theorem, the prior knowledge of model parameters, given by Pr(A), can be updated, given the data, to yield the posterior knowledge, given by Pr(A|B). Applying Bayes' theorem is often challenging, not just the technical aspects of its solution, but also selecting the prior evidence and formulating the associated distributions for the model parameters. The technical challenges have been addressed recently with the advancement of computer algorithms for Gibbs sampling [1], but selecting prior evidence and formulating distributions can remain an issue, and is often used as an argument in favor of taking a frequentist approach.

The author argues that the prior distribution should be based on whatever evidence the researcher thinks is valid, and in that sense the associated inference and decision-making is transparently conditional on the evidence chosen [2]. However, for inference and decision-making to be convincing, the prior evidence selected must have credibility with those involved in the decision-making, as well as with those directly affected by it. Determining the appropriate prior evidence is an exercise in evidence synthesis and, therefore, the issue of publication bias needs to be addressed. An exhaustive search for relevant evidence is required, involving not just peer-reviewed journals, but also meeting abstracts and trial registration databases. The researcher may decide that there is no valid prior evidence, or that inference and decision-making should be based solely on the research data, in which case the author

Bayesian methods provide a transparent means of combining prior and new knowledge.

argues that the use of uninformative priors [3–5], although unsatisfactory from a purest point of view, is appropriate.

Bayesian methods lead to choerent, simple, intuitive and relevant statements of statistical inference.

The author further argues that transparently conditional inference and decision-making is better than frequentist inference, which is always opaquely conditional on ignoring all prior evidence. Some authors recommend a sensitivity analysis, in which various priors, ranging from skeptic to enthusiastic, are used [2]. Having selected the prior evidence to include, even if it is none, and accepting that the inference and decision-making is conditional, the other advantages of the Bayesian approach are realized. One advantage is that a Bayesian approach permits probability statements regarding model parameters, allowing for simple, intuitive and relevant statistical inference, and avoiding the misunderstandings often associated with the frequentist approach. Another advantage is the ability to apply Bayesian decision theory. Bayesian decision theory answers questions such as: is the current evidence in support of a new healthcare intervention sufficient for its adoption and at what level of reimbursement, and, if not, what research study is optimal for gathering further evidence?

Numerous authors have argued in favor of a Bayesian approach to statistical analysis [2–13]. A broad summary and good review is given by Spiegelhalter *et al.* [7]. Goodman states that the frequentists approach is "an amalgam of incompatible elements" [6]. In the same article, he states that the use of p-values is based on "the mistaken idea that a single number can capture both the long-run outcomes of an experiment and the evidential meaning of a single result". In an accompanying article, Goodman argues that Bayesian approaches "make the distinction clear between experimental evidence and inferential conclusions while providing a framework in which to combine prior with current evidence" [8]. Spiegelhalter *et al.* state, "that a Bayesian approach allows a formal basis for using external evidence" and Bayesian methods provide a rational way to deal with the ethics of randomization, equivalence trials and the monitoring of accumulating evidence [2]. Kadane argues that the Bayesian approach is more flexible and ethical, and provides an appropriate means for addressing the problem of multiple testing [10]. Cornfield, and Lilford and Braunholtz argue that the Bayesian framework is the only coherent approach for healthcare policy decision-making [10,11].

In this chapter, in the section 'Definitions of probability & statements of inference', two definitions of probability are given. The first supports the frequentist approach, while the other supports the Bayesian. The ability of

the two approaches to interpret the results of clinical studies is compared. In a following section, a number of real examples are given demonstrating how a frequentist approach can lead to faulty inference and how a Bayesian approach provides inference that is more consistent with the evidence. Lastly, a brief discussion of the role of Bayesian decision theory in the planning, analysis and interpretation of clinical research is given.

Definitions of probability & statements of inference

Frequentist and Bayesian inferences are fundamentally different since they are based on two entirely different definitions of probability. For the frequentist, the probability of an event is the limiting relative frequency of its occurrence in a series of repeated observations of a chance outcome in which it could occur. In this definition, the probability that a tossed coin will come up heads is 0.5, because if it was flipped an arbitrarily large number of times the proportion of times it would come up heads is 0.5. For the Bayesian, probability is the subjective expression of the uncertainty, or degree of belief, regarding the unknown. This is the definition in common usage. When someone takes an umbrella with them on a day in which the weather broadcast reports an 80% probability of rain, they are not thinking that if the day was relived an arbitrarily large number of times, it would rain on four-fifths of them. They are thinking there is a good chance of getting wet today if they do not have an umbrella with them. The day only occurs once, and the weather broadcaster is expressing his or her subjective belief about whether or not it will rain that day.

The different definitions of probability lead to differing statements of inference. Consider the results of a randomized clinical trial (RCT) designed to compare treatment with standard, with respect to the probability of a bad outcome, in which the frequentist p-value is 0.035, and where a one-sided test of the null hypothesis that the relative risk (RR) is 1 is applied at the 5% level. The frequentist statement of inference would read as follows: 'we can reject the null hypothesis that the RR is ≥1 (i.e., treatment is equivalent or inferior to standard) in favor of the alternative hypothesis that the RR is <1 (i.e., treatment is superior to standard) with a probability of being wrong of less than 5%. This means that if the null hypothesis is true (i.e., treatment is equivalent or inferior to standard) and the trial was repeated an arbitrarily large number of times, the proportion of times that the results of these replications will be at least as inconsistent with the null hypothesis in favor of the alternative as the data from the trial under consideration is less than 5%.' This is not a direct statement of inference about the RR. It merely facilitates the deductive reasoning, in which one rejects the null hypothesis and concludes that the RR <1, because the probability of falsely do so is small.

This is not a statement about falsely rejecting the null hypothesis in favor of the alternative for this particular trial, but rather it is a statement about the proportion of an arbitrarily large number of null hypotheses that would be falsely rejected using the same criterion. Based on the author's 35 years of experience, the proportion of people who interpret and apply the results of clinical research that fully understand this is very low. Although this is not a criticism of frequentist approaches *per se*, it does mean that frequentist inference is often misunderstood, misapplied and leads to faulty clinical decision-making. The argument made in this chapter is that Bayesian approaches provide statements of inference that are more intuitive and relevant, and are less likely to be misapplied. Furthermore, Bayesian approaches facilitate decision theory, which provides optimal clinical decision-making in the face of uncertainty.

By contrast, Bayesian methods do allow for direct statements of inference regarding the RR, based on its posterior distribution function. Ignoring for the moment the issue of incorporating prior information, the Bayesian statement of inference is the $Pr_{post}(RR < 1) = 0.035$, is simply, 'the probability that treatment is superior to standard is 96.5%.' Such simple statements of inference are not available in the frequentist framework, where probability statements are restricted to observations of the data. Clearly, the Bayesian statement inference is more intuitive, easier to interpret and addresses the specific issue in question, namely 'what is the evidence that treatment is superior to standard?'

For the same example consider the frequentist confidence interval for RR of 0.493–0.917. This is interpreted as follows: 'if the trial was conducted an arbitrarily large number of times, in the limit, the proportion of the confidence intervals from these replications that include the true RR is 95%.' It is hard to understand how this provides inference regarding the RR. If the interval excludes the null hypothesis, one claims that it can be rejected, but this has more to do with the way the interval is constructed than its definition, and in any case, the confidence interval is not required to test the null hypothesis. A Bayesian statement regarding a credible interval with the same limits is 'there is a 95% probability that the RR lies in the interval from 0.493 to 0.917.' Again, the Bayesian approach provides more intuitive and interpretable statements of inference.

Examples

Antenatal corticosteroids for women at risk for early delivery

Numerous RCTs have found that a single course of antenatal corticosteroids (ACS) for pregnant women who are at risk of early delivery, reduces the risk of respiratory distress syndrome and other bad

outcomes in the babies. In a 2006 Cochrane review four RCTs were identified to have reported on cerebral palsy in the infants (the review has since been updated with an additional trial) [14]. This evidence is vital since it relates to a lawsuit and its interpretation has important consequences. The data are shown in **Table 8.1**. The Cochrane review reported a RR of 0.63, with an associated 95% confidence limits of 0.36, 1.08 and a two-sided p-value of 0.08. Assessing this evidence in an expert opinion, an academic clinician, wrote, "there is no evidence from this review of a difference of incidence of later cerebral palsy". This opinion presumably follows from the lack of significance at the traditional 0.05. Quite apart from the fact that this is in fact a one-sided question and recognition of this might have lead to a different conclusion, the term 'no evidence' in reference to this evidence represents a complete misinterpretation, which is made far too frequently. Frequentist methods, which have become locked into the use of 0.05 as the threshold level for significance, have no way of assessing this evidence in an intuitive and meaningful way. The prespecification of a level of significance of 0.05, or sometimes 0.01, apparently became the convention because Fisher, due to copyright issues, was prevented from reproducing the complete tables for a standard normal random variable, but not from citing the levels for 0.05 and 0.01 [10].

Insisting that a null hypothesis be rejected with a low probability of error may have some appeal when one is attempting to limit the probability of incurring the opportunity loss of adopting a new intervention in the face of uncertainty; however, that is not at issue here, since the use of ACS in this circumstance is standard practice. The issue is what is the evidence that failing to administer ACS to a pregnant woman at risk of early delivery increases the risk of cerebral palsy? This is an issue that the frequentist approach is ill equipped to address. A Bayesian approach, in which an uninformative prior is used for the log odds ratio, provides a means of making inference using this data to address the issue in an intuitive and

Table 8.1. Distributions for early cephalic version example.

Arm	Distribution $\beta(a,b)$	Mean $a/(a+b)$	Variance $ab/\{(a+b)^2(a+b+1)\}$
EECV	$\beta(474,409)$	0.53681	0.00028127
ECV	$\beta(514,371)$	0.58079	0.00027480
Difference	Normal (-0.043985, 0.00055607)		

ECV: External cephalic version; EECV: Early external cephalic version

The opportunity loss of decision taken is the difference between the utility of the best decision and the utility of the decision taken.

meaningful way. By using an uninformative prior we are explicitly stating that the inference is based solely on the data from the four trials, and that it is a reasonable approach given we are including all known relevant RCTs. Using the WinBUGS code given in the supplementary information available online (see online at www.future-science.com/doi/suppl/10.4155/EBO.13.431), the posterior probability that the log odds ratio is negative is 0.9495, leading to the inference that the probability that a single course of ACS reduces the risk of cerebral palsy is 94.95% and provides strong evidence in favor of a single course of ACS.

Early external cephalic version

For pregnant women in the breech position at or beyond 37 weeks gestation it is standard practice to undergo a procedure referred to as external cephalic version (ECV) in an attempt to manipulate the fetus into a vertex position and, thus, avoiding the need for a caesarean delivery (CD). A randomized pilot study was performed to determine if, applying the procedure earlier between 33 and 36 weeks', early external cephalic version (EECV) may reduce the need for CD [15]. In total, 116 patients were randomized to each procedure and the observed numbers of CDs were 83 and 75 in the ECV and EECV arms, respectively. Based on these results a larger RCT was performed, in which 767 patients were allocated to the ECV arm and 765 patients to the EECV arm [16]. The authors, reporting the results of the large trial, stated that there were no differences in rates of cesarean section (398/765 [52.0%] for EECV vs 430/768 [56.0%] for ECV; RR: 0.93; 95% CI: 0.85, 1.02; p = 0.12) and conclude that EECV "does not reduce the rate of caesarean section" [16].

If we assume the uninformative prior $\beta(1,1)$ for the probability of a CD for ECV and EECV, then the posterior distributions following both the pilot and larger trial would be:

- $\beta(1 + 75 + 398, 1 + [116 - 75] + [765 - 398]) = \beta(474,409)$ for EECV
- $\beta(1 + 83 + 430, 1 + [116 - 83] + [765 - 430]) = \beta(514,371)$ for ECV

Using the WinBUGS code given in the supplementary information available online, to derive the posterior distribution for the difference in the probabilities, yields an expected difference of -0.044, where a probability that the difference is less than 0 is 0.9687 (i.e., the probability that EECV reduces the risk of CD is 96.87%). With this statement of inference, it is very hard to support the conclusion that EECV "does not reduce the rate of caesarean section". It is important to note that, by assuming an

uninformative prior, we are basing our inference solely on the data from the two trials, and that by taking a Bayesian approach we are able to make direct inference regarding the difference in the between-arm probabilities of a CD. Assuming a normal approximation for the β distributions yields an expected difference in the probabilities of -0.044 and a probability that EECV reduces the risk of CD of 96.89%.

Topical paromomycin & gentamicin for cutaneous leishmaniasis

In a recent article in the *New England Journal of Medicine*, Ben Salah *et al.* reported the results of a randomize trial of topical paromomycin and gentamicin versus a control, in the treatment of cutaneous leishmaniasis [17]. Of the 125 patient randomized to active, 101 achieved the threshold for cure, compared with 73 of the 125 patients randomized to control. Based on these results, the authors rejected the null hypothesis using a two-sided test at the 0.05 level, and concluded in the abstract that "this trial provides evidence of the efficacy of paromomycin–gentamicin" [17]. It is worth noting that the stated conclusion is actually just a restating of the results. It says nothing about the whether or not paromomycin–gentamicin increases the probability of cure. This is appropriate because frequentist methods can only provide probability statements about the data, given a particular hypothesis, not about the difference in the probabilities of cure, given the data. To make direct inferential statements, appropriate prior distributions are required. A randomized Phase II pilot study was reported earlier, by some of the same authors [18]. In that study, of the 50 patients randomized to active, 47 achieve the threshold for cure, compared with 30 of the 42 patients randomized to control. If we assume the uninformative prior $\beta(1,1)$ for the probability of a cure for active and control prior to the pilot then the posterior distributions following both the pilot and larger trial would be:

- $\beta(1 + 47 + 101, 1 + [50 - 47] + [125 - 101]) = \beta(149,28)$ for active
- $\beta(1 + 30 + 73, 1 + [42 - 30] + [125 - 73]) = \beta(104,65)$ for control

Using WinBUGS with the same code used in the 'Early external cephalic version' section, to derive the posterior distribution for the difference in the probabilities, yields a difference of 0.206, with a probability that the difference is >0 of 0.9999 (the same result is found using normal approximations). Using a Bayesian approach and the pilot data to formulate the prior, the authors could have concluded that the probability that active increases the probability of cure is 99.99%, rather than concluding that 'this trial provides evidence' that active increases the probability of cure, which is only to state that using the data from the

trial, a two-sided, 0.05% level test of the null hypothesis can be rejected, which is only to say that if the same criteria was similarly applied to an arbitrarily large number of tests of the hypothesis, only 5% of them would be falsely rejected.

Trial of planned cesarean or vaginal delivery for twin pregnancy

In a recent article in the *New England Journal of Medicine,* by Barrett *et al.* [19], the sample size was determined using a frequentist approach to achieve an 80% power for rejecting the null hypothesis using a two-sided, 5% level test if the new intervention reduced the probability of a bad outcome from 4 to 2%. The proportion of bad outcomes was observed to be approximately 2% in both arms and the null hypothesis was not rejected. That the observed proportion of bad outcomes in the standard group was lower than expected, led to the question, 'are these negative results underpowered' and 'is the conclusion of no difference valid?' These are very hard questions to answer in the frequentist framework. The 95% CI excludes a RR of 0.5, which does not really lead to a satisfactory statement of inference. The best that can be said is that we can 'exclude with 95% confidence that the RR is <0.5'; however, that is simply restating the observation that the lower limit of the 95% confidence interval exceeds 0.5. Adopting a Bayesian approach, using whatever prior evidence is deemed relevant, a direct statement of inference could have been made regarding the probability that the RR is less <0.5, along the lines of, 'the probability that the new intervention reduces the risk of a bad outcome sufficiently to justify its adoption is 1%', for the sake of argument.

The examples in this section are not given to imply that Bayesian approaches arrive at a truth that frequentist approaches fail to uncover. They are given to illustrate that Bayesian statements of inference are more intuitive and relevant, and appear to be more consistent with the empirical observations. The reader can decide if this has been successful.

Decision theory & value of information methods

The direct statement 'there is a very high probability that a new intervention improves health outcomes' does not necessarily imply that the intervention should be adopted. The following questions still need to be answered. Is the evidence sufficient and if the evidence is insufficient, what is the optimal design for further research? These questions can be addressed by Bayesian decision theory, also referred to as value of information

Bayesian methods support value of information methods, which can be used to determine the adequacy of current evidence and the value of additional research.

Aa Expected value of sample information: the reduction in the expected opportunity loss from additional evidence.

methods. To apply value of information methods a mean utility is ascribed to each intervention in the form of net benefit (NB), defined as $NB = e\lambda - c$, where: e is the mean health outcome (effectiveness), where larger values of e are preferred; λ is the threshold value for the willingness-to-pay for a unit of health outcome; and c is the mean of the total healthcare costs associated with treating a patient with the intervention. Let incremental NB (INB) be the difference between the NB for treatment and the NB for standard. If the current mean of INB is positive, and no further evidence is to be sought, then the utility (i.e., NB) maximizing decision for future patients is to adopt treatment. However, if there is positive probability that INB is <0, the decision to adopt treatment is associated with an opportunity loss. The opportunity loss is a function of INB, and its expectation can be taken with respect to the current evidence. The amount by which the expected opportunity loss is reduced by performing a future study is the expected value of sample information (EVSI) of that study. If the cost of the future study exceeds the EVSI for all sample sizes, then the current evidence is sufficient and the optimal decision is to adopt treatment. On the other hand, if the EVSI exceeds trial cost for some sample size, then we are in a true state of equipoise and the optimal decision is to delay the adoption of treatment and perform the study with a sample size that maximizes the difference between the value and the cost. Therefore, by adopting a Bayesian approach, the truly relevant issues regarding the assessment of current evidence in treatment decision-making can be addressed. Details and extension can be found in [20–36].

Disclaimer

The opinions expressed in this chapter are those of the author and do not necessarily reflect the views of Future Science Ltd.

Financial & competing interests disclosure

A Willan is funded by the Discovery Grant Program of the Natural Sciences and Engineering Research Council of Canada (grant number 44868–08). The author has no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Summary.

- Bayesian methods have a number of advantages that are especially relevant for the planning, analysis and interpretation of clinical research.
- Bayesian methods provide a transparent means of updating prior evidence when new data become available.
- Bayesian approaches permit probability statements regarding model parameters, allowing for simple, intuitive and relevant statistical inference, and avoiding the misunderstandings often associated with the frequentist approach.
- Bayesian decision theory answers questions such as: is the current evidence in support of a new healthcare intervention sufficient for its adoption and at what level of reimbursement, and, if not, what research study is optimal for gathering further evidence?

References

1 Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBUGS – a Bayesian modelling framework: concepts, structure, and extensibility. *Stat. Comput.* 10(4), 325–337 (2000).

2 Spiegelhalter DJ, Freedman LS, Parmar MKB. Bayesian approaches to randomized trials. *J. Roy. Stat. Soc. A Sta.* 157(3), 357–416 (1994).

3 Jeffreys H. *Theory of Probability (2nd Edition).* Oxford University Press, Oxford, UK (1961).

4 Kass R, Raftery A. Bayes factors. *J. Am. Stat. Assoc.* 90(430), 773–795 (1995).

5 Carlin C, Louis T. *Bayes and Empirical Bayes Methods for Data Analysis*. Chapman and Hall, London, UK (1996).

6 Goodman SN. Toward evidence-based medical statistics. 1: the p value fallacy. *Ann. Intern. Med.* 130(12), 995–1004 (1999).

7 Spiegelhalter DJ, Myles JP, Jones DR, Abrams KR. An introduction to bayesian methods in healthtechnology assessment. *Br. Med. J.* 319(7208), 508–512 (1999).

8 Goodman SN. Toward evidence-based medical statistics. 2: the Bayes factor. *Ann. Intern. Med.* 130(12), 1005–1013 (1999).

9 Kadane JB. Prime time for Bayes. *Contr. Clin. Trials* 16(5), 313–318 (1995).

10 Cornfield J. Recent methodological contributions to clinical trials. *Am. J. Epid.* 104(4), 408–421 (1976).

11 Lilford RJ, Braunholtz D. The statistical basis of public policy – a paradigm shift is overdue. *Br. Med. J.* 313(7057), 603–607 (1996).

12 Berry DA, Stangl DK. *Bayesian Biostatistics*. Berry DA, Stangl DK (Eds). Dekker, NY, USA (1996).

13 Breslow N. Biostatistics and Bayes. *Stat. Sci.* 5(3), 269–284 (1990).

14 Roberts D, Dalziel SR. Antenatal corticosteroids for accelerating fetal lung maturation for women at risk of preterm birth. *Cochrane Database Syst. Rev.* 3, CD004454 (2006).

15 Hutton EK, Kaufman K, Hodnett E *et al;* the Early External Cephalic Version Trial Group. External cephalic version beginning at 34 weeks' gestation versus 37 weeks' gestation: a randomized multicenter trial. *Am. J. Obstet. Gynecol.* 189(1), 245–54 (2003).

16 Hutton E, Hannah M, Ross S *et al*; the Early ECV2 Trial Collaborative Group. The Early External Cephalic Version (ECV) 2 Trial: an international multicentre randomised controlled trial of timing of ECV for breech pregnancies. *BJOG* 118(5), 564–577 (2011).

17 Ben Salah A, Messaoud NB, Guedri E *et al.* Topical paromomycin with or without gentamicin for cutaneous leishmaniasis. *N. Engl. J. Med.* 368(6), 524–532 (2013).

18 Ben Salah A, Buffet PA, Morizot G *et al.* WR 279,396, a third generation amino-glycoside ointment for the treatment of Leishmania major cutaneous leish-maniasis: a Phase 2, randomized, double blind, placebo controlled study. *PLoS Negl. Trop. Dis.* 3(5), e432 (2009).

19 Barrett JFR, Hannah ME, Hutton EK *et al*. A randomized trial of planned cesarean or vaginal delivery for twin pregnancy. *N. Engl. J. Med.* 369(14), 1295–1305 (2013).

20 Grundy PM, Healy MJR, Rees DH. Economic choice of the amount of experimentation. *J. Roy. Stat. Soc. B Met.* 18(1), 32–48 (1956).

21 Claxton K, Posnett J. An economic approach to clinical trial design and research priority-setting. *Health Econ.* 5(6), 513–524 (1996).

22 Claxton K. The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies. *J. Health Econ.* 18(3), 341–364 (1999).

23 Cheng Y, Su F, Berry DA. Choosing sample size for a clinical trial using decision analysis. *Biometrika* 90(4), 923–936 (2003).

24 Willan AR, Pinto EM. The expected value of information and optimal clinical trial design. *Stat. Med.* 24(12), 1791–1806 (2005).

25 Willan AR. Clinical decision making and the expected value of information. *Clin. Trials J.* 4(3), 279–285 (2007).

26 Eckermann S, Willan AR. Expected value of information and decision making in HTA. *Health Econ.* 16(2), 195–209 (2007).

27 Eckermann S, Willan AR. Time and EVSI wait for no patient. *Value Health* 11(3), 522–526 (2008).

28 Eckermann S, Willan AR. The option value of delay in health technology assessment. *Med. Decis. Making* 28(3), 300–305 (2008).

29 Willan AR. Optimal sample size determinations from an industry perspective based on the expected value of information. *Clin. Trials J.* 5(6), 587–594 (2008).

30 Willan AR, Kowgier ME. Determining optimal sample sizes for multi-stage randomized clinical trials using value of information methods. *Clin. Trials J.* 5(4), 289–300 (2008).

31 Eckermann S, Willan AR. Globally optimal trial design for local decision making. *Health Econ.* 18(2), 203–216 (2009).

32 Willan AR, Eckermann S. Optimal clinical trial design using value of information methods with imperfect implementation. *Health Econ.* 19(5), 549–561 (2010).

33 Willan AR, Eckermann S. Value of information and pricing new health care interventions. *PharmacoEconomics* 30(6), 447–459 (2010).

34 Willan AR, Eckermann S. Accounting for between-study variation in incremental net benefit in value of information methodology. *Health Econ.* 21(10), 1183–1195 (2012).

35 Willan AR, Goeree R, Boutis K. Value of Information methods for planning and analyzing clinical studies optimize decision making and research planning. *J. Clin. Epidemiol.* 65(8), 870–876 (2012).

36 Chen MH, Willan AR. Determining optimal sample sizes for multistage adaptive randomized clinical trials from an industry perspective using value of information methods. *Clin. Trials J.* 10(1), 54–62 (2013).