

RESEARCH

Open Access



# Power and sample size calculation for incremental net benefit in cost effectiveness analyses with applications to trials conducted by the Canadian Cancer Trials Group

Louis Everest<sup>1</sup>, Bingshu E. Chen<sup>2</sup>, Annette E. Hay<sup>2</sup>, Matthew C. Cheung<sup>1,3,4†</sup> and Kelvin K. W. Chan<sup>1,3,4,5\*†</sup>

## Abstract

**Background** Historically, a priori power and sample size calculations have not been routinely performed cost-effectiveness analyses (CEA), partly because the absence of published cost and effectiveness correlation and variance data, which are essential for power and sample size calculations. Importantly, the *empirical* correlation between cost and effectiveness has not been examined with respect to the estimation of value-for-money in clinical literature. Therefore, it is not well established if cost-effectiveness studies embedded within randomized-controlled-trials (RCTs) are under- or over-powered to detect changes in value-for-money. However, recently guidelines (such as those from ISPOR) and funding agencies have suggested sample size and power calculations should be considered in CEAs embedded in clinical trials.

**Methods** We examined all RCTs conducted by the Canadian Cancer Trials Group with an embedded cost-effectiveness analysis. Variance and correlation of effectiveness and costs were derived from original-trial data. The incremental net benefit method was used to calculate the power of the cost-effectiveness analysis, with exploration of alternative correlation and willingness-to-pay values.

**Results** We identified four trials for inclusion. We observed that a hypothetical scenario of correlation coefficient of zero between cost and effectiveness led to a conservative estimate of sample size. The cost-effectiveness analysis was under-powered to detect changes in value-for-money in two trials, at willingness-to-pay of \$100,000. Based on our observations, we present six considerations for future economic evaluations, and an online program to help analysts include a priori sample size and power calculations in future clinical trials.

**Conclusion** The correlation between cost and effectiveness had a potentially meaningful impact on the power and variance of value-for-money estimates in the examined cost-effectiveness analyses. Therefore, the six considerations and online program, may facilitate a priori power calculations in embedded cost-effectiveness analyses in future clinical trials.

†Matthew C. Cheung and Kelvin K.W. Chan contributed equally to this work.

\*Correspondence:

Kelvin K. W. Chan

[kelvin.chan@sunnybrook.ca](mailto:kelvin.chan@sunnybrook.ca)

Full list of author information is available at the end of the article



## Highlights

- Analysts may use the online program presented in the present study to examine the a priori power of cost-effectiveness analyses.
- Analysts may potentially apply the considerations presented in this paper in the planning stage of future cost-effectiveness analyses.

## Background

The increasing cost of anticancer agents over the past two decades has generated discussion in the literature regarding the value of the novel anticancer therapies [1]. Specifically, concerns in literature have been presented due to the disproportionately modest survival benefits of novel cancer therapeutic agents, compared to the substantial increases in cost [2]. However, phase III anticancer trials are conventionally designed to detect improvements in efficacy and not necessarily changes in value-for-money [3, 4].

Currently, CEAs embedded in cancer trials are commonly completed without formal sample size and power calculations [5]. However, a recent ISPOR Good Research Practices Task Force report has suggested the inclusion of sample size calculations for CEAs embedded in clinical trials [5]. Additionally, research grant funding agencies may commonly request statistical sample size justifications even for secondary economic evaluation endpoints in cancer clinical trials [personal communication: Matthew Cheung, Co-chair of Committee on Economic Analysis, Canadian Cancer Trials Group, Nov 11, 2022]. However, there is currently a paucity of information in published literature with respect to empirical estimates of sample size parameters [5, 6]. Therefore, analysts may not be able to examine the power of CEAs embedded in clinical trials, despite its recognised importance in literature [5]. Additionally, the likelihood a new treatment is cost effective based on Bayesian methods, may be challenging to estimate because prior variance and covariance distributions are also not well established in literature [6, 7].

Importantly, because information regarding variances of cost and co-variances of cost and effectiveness are often not well established in the literature, a priori sample size or power calculations for cost-effectiveness analysis endpoints typically are not done or only partly done by making some assumptions regarding cost differences between the experimental and control groups and the corresponding variances [6, 8]. Additionally, because most new cancer treatments improve survival and increase costs compared to the standard, there is reasonable theoretical justification that covariance may be non-ignorable in these trials [9]. Therefore, in the absence of a priori power calculations, cost-effectiveness analyses may be under- or over-powered to detect changes in

value [9]. Additionally, because many new cancer agents enter the market with incremental cost effectiveness ratios near willingness-to-pay thresholds, cost-negotiations resulting in lower drug prices may reduce the associated power of cost-effectiveness analyses [10].

An understanding of the relationship between the variances of and correlation between incremental cost and incremental effectiveness and their influence on power and sample size is not currently available for cancer trials with embedded economic evaluations [11]. To our knowledge, the Canadian Cancer Trials Group (CCTG) is currently the only oncology trial group with a standing committee, the Committee on Economic Analysis, dedicated to designing cost-effectiveness analyses to be embedded within randomized controlled trials with collection of cost and resource utilization prospectively. Therefore, we aimed to examine power calculations in cost-effectiveness analyses embedded within RCTs conducted by the CCTG, in order to better understand and facilitate future health technology assessments.

In this paper, we demonstrate the calculation of power and sample size of cost-effectiveness analyses based on the paradigm of incremental net benefit developed by Willan and Lin [11] using original trial cost and effectiveness individual patient data from phase III trials conducted by the CCTG [12–19]. The statistical model is summarized in the Methods section, followed by an application to cost-effectiveness analyses conducted by CCTG with respect to correlation coefficients and willingness-to-pay values in the Results section. We then discuss the practical implications of our demonstration in the Discussion section, providing guidance on the design of trial-based economic analyses and an online resource to enable sample size and power calculations.

## Methods

### Selection of Studies and Parameter Calculation

The present study examined all RCTs conducted by CCTG with an embedded cost-effectiveness analysis. Variance and correlation of effectiveness and costs were derived from original-trial data. The primary analysis examined the power of the cost-effectiveness analysis based on the

incremental net benefit method, with exploration of alternative correlation and willingness-to-pay values.

**Statistical Methods of Calculating the Sample Size and Power based on the Paradigm of Incremental Net Benefit**

Conventionally, incremental cost-effectiveness ratios (ICERs), and associated confidence intervals, are a common method of quantifying value and uncertainty respectively, in cost-effectiveness analyses of anticancer agents [4, 6]. In a two-arm randomized controlled trial, let the mean effectiveness of the treatment and standard arms be  $E_1$  and  $E_0$  respectively. Additionally, the mean in costs in the treatment and standard arms are represented as  $C_1$  and  $C_0$  respectively. The ratio between the change in mean cost and the change in mean effectiveness is the ICER, defined as:

$$ICER = \frac{C_1 - C_0}{E_1 - E_0} = \frac{\Delta C}{\Delta E}$$

where  $\Delta C = C_1 - C_0$  and  $\Delta E = E_1 - E_0$  are the cost and effectiveness difference between the treatment and the control groups, respectively.

The present analysis quantifies cost-effectiveness through the incremental net benefit (INB) method. This method was selected because sample size calculations based on the INB method and the willingness-to-pay value (or threshold), are well established in statistical literature [4], and are one of the commonly used methods, when sample size calculations are conducted in cost-effectiveness analyses [20]. In cost effectiveness analyses, the willingness-to-pay value is typically defined as the amount of money that a decision maker or society is willing to pay for a 1-unit improvement in

generality, the INB is used in the present analysis; however, the practical problems encountered when implementing the ICER method are similarly applicable [6, 11]. However, because ICER is a ratio of the cost and effectiveness, statistical inferences such as the corresponding standard error may be difficult to obtain. [6, 11]. Additionally, standard errors may be unreasonable in some cases (e.g., when the change in efficacy is close to zero, resulting in a standard error that is close to infinity).

The hypothesis test of the cost-effectiveness for a willingness-to-pay value may be examined as:

$$H_0 : b(\lambda) \leq 0, \text{ versus } H_1 : b(\lambda) > 0$$

where the alternative hypothesis may be interpreted as suggesting the treatment is demonstrated to be cost-effective. Additionally, failing to reject the null hypothesis implies that the cost-effectiveness of the treatment, when compared to the control, is at or below the willingness-to-pay threshold, barring the lack of statistical power to reject the null hypothesis. Further, in the present study, the hypothesis test  $H_0 : b(\lambda) \geq 0, \text{ versus } H_1 : b(\lambda) < 0$  is also examined in the CO.17 and CO.17 KRAS trials. This hypothesis test was included to examine the impact of the correlation coefficient on non-cost-effective trials. Importantly, the latter hypothesis test is unrealistic in CEAs, however the trends are generalizable to the former hypothesis test with respect to the impact of the correlation coefficient.

**Variance and Sample Size Calculations**

For a study with sample size  $n$ , where  $n_0$  and  $n_1$  represent the sample sizes of the control and treatments groups respectively, the variance of the INB  $b(\lambda)$  can be estimated as:

$$\sigma_{b(\lambda)}^2 = \frac{1}{n_0} \left( \lambda^2 \sigma_{E_0}^2 + \sigma_{C_0}^2 - 2\lambda\rho_0\sigma_{E_0}\sigma_{C_0} \right) + \frac{1}{n_1} \left( \lambda^2 \sigma_{E_1}^2 + \sigma_{C_1}^2 - 2\lambda\rho_1\sigma_{E_1}\sigma_{C_1} \right) \tag{2}$$

efficacy [21, 22]. For a two-arm randomized controlled trial, the INB is defined as

$$b(\lambda) = \lambda \cdot (\Delta E) - (\Delta C) \tag{1}$$

Importantly, the ICER may be calculated as a special case of the INB. Specifically, the ICER may be calculated as the horizontal intercept of the plot of  $b(\lambda)$  (y axis) and  $\lambda$  (x axis). Therefore, when the INB is calculated as  $\lambda \cdot (\Delta E) - (\Delta C) = 0$ , willingness-to-pay value  $\lambda$  is equivalent to the ICER. Furthermore, the confidence limits for  $b(ICER)$  cross the horizontal axis at the Fieller limits for a specified ICER, allowing for further inferences [6, 11, 23]. Therefore, without the loss of

where  $\sigma_{E_j}^2$  and  $\sigma_{C_j}^2, j = 0, 1$  are the variance of effectiveness and costs in the control and experimental groups, respectively. Additionally,  $\rho_j$  is the correlation coefficient between effectiveness and cost in the group  $j, j = 0, 1$ . Conventionally, the variance of ICER is calculated through bootstrap methods or Fieller’s theorem in some cases. A derivation of the Fieller’s confidence limits when  $\hat{b}(\lambda)$  crosses the horizontal axis is presented in Willan (2006) [24] as well as Zethraeus and Löthgren [23]. Variance and correlation of effectiveness and costs were derived from original-trial data using intent-to-treat analyses, over the entire trial time horizon. Detailed derivations and formulas of these parameters are presented by Willan [9]. Additionally, details with respect to the

parameter estimates are presented in the associated CEAs of the examined trials [16–19]. The smallest important difference in incremental net benefit is defined as  $b(\lambda)_\delta = \lambda \times (\Delta E) - (\Delta C)$ . Therefore, as examined by Willan and Lin (2001), to test the one-sided hypotheses at the  $\alpha$  level and  $(1 - \beta) \times 100\%$  power, the total sample size is given by:

$$n = \frac{(z_{1-\alpha} + z_{1-\beta})^2 \cdot \sigma_{b(\lambda)}^2}{b(\lambda)_\delta^2} \tag{3}$$

where the type I error,  $\alpha$  in Eq. (3), is the probability of claiming that the treatment cost-effective if the null hypothesis is true, which is usually set to be  $\alpha = 0.05$ . The type II error,  $\beta$  in Eq. (3), is the probability of failing to reject the null hypothesis, when the true INB is equal to or less than  $b(\lambda)_\delta$ . Therefore,  $z_{1-\alpha}$  and  $z_{1-\beta}$  are quantiles of a standard normal distribution with respect to  $(1 - \alpha)$  and the power, respectively.

In general, a priori assumptions based on existing literature regarding  $\Delta E$  and variance of  $\Delta E$ , are used to inform the primary endpoint of the efficacy sample size calculations of the clinical trial. Specifically, in phase III clinical trials, efficacy sample size calculations are typically based on the target hazard ratio and the assumption of an exponential survival distribution [3]. Therefore, in the present analysis, the sample size calculation is based on the assumption that survival time follows an exponential distribution. In comparison, the

These assumptions have been made to facilitate illustrative comparisons with respect to the impact of the correlation coefficient on sample size calculation and should be applied to external cost-effectiveness analyses cautiously. Based on these assumptions, we can rewrite variance formula Eq. 2 as

$$\sigma_{b(\lambda)}^2 = \frac{4}{n} \left( \lambda^2 \sigma_E^2 + \sigma_C^2 - 2\lambda\rho\sigma_E\sigma_C \right) \tag{4}$$

Since  $b(\lambda)_\delta = \lambda \cdot (\Delta E) - (\Delta C)$ , let  $z = (z_{1-\alpha} + z_{1-\beta})^2$ , for example, when one-sided  $\alpha = 0.05$  and power = 80%, we have  $z = 6.18$  and  $4 \times z = 24.72$ . Therefore, we can replace the quantities in the sample size Eq. 3 with the corresponding terms for cost differences and the effectiveness differences in incremental net benefit analysis and obtain the following sample size formula

$$n = 4 \cdot z \cdot \frac{\lambda^2 \pi \sigma_E^2 + \sigma_C^2 - 2\lambda\rho\sqrt{\pi}\sigma_E\sigma_C}{\lambda^2(\Delta E)^2 + (\Delta C)^2 - 2\lambda(\Delta E)(\Delta C)} \tag{5}$$

where  $\pi = \frac{n_e}{d_e}$ , and denotes the ratio of the total expected sample size ( $n_e$ ) and the expected number of events ( $d_e$ ). The parameter  $\pi$  will be fixed once the design for the primary effectiveness endpoint is finalized. For example, in a trial with sample size  $n_e = 500$  and the final analysis will be triggered when  $d_e = 400$  events are observed, then  $\pi = \frac{500}{400} = 1.25$ . Furthermore, as examined by Willan & Lin [11], the corresponding power function is given by:

$$P(\delta) = \Phi \left( \frac{\widehat{b}(\lambda)_\delta}{\sqrt{V[\widehat{b}(\lambda)]}} - Z_{1-\alpha} \right) = \Phi \left( \frac{\lambda \cdot (\Delta E) - (\Delta C)}{\sqrt{\frac{4}{n} \times (\lambda^2 \sigma_E^2 + \sigma_C^2 - 2\lambda\rho\sigma_E\sigma_C)}} - Z_{1-\alpha} \right) \tag{6}$$

treatment effectiveness  $\Delta E$  and variance  $\sigma_{\Delta E}^2$  were calculated analytically, based on individual patient data.

To have a better understanding on what roles the correlation coefficients will play in the sample size and power determination, we make some simplification based on the following assumptions:

1. The study will be balanced between the control and experimental arms  $n_0 = n_1 = \frac{n}{2}$ .
2. The costs in the control and experimental arms have the same variance  $\sigma_{C_0}^2 = \sigma_{C_1}^2 = \sigma_C^2$
3. The effectiveness in the control and experimental arms have the same variance  $\sigma_{E_0}^2 = \sigma_{E_1}^2 = \sigma_E^2$
4. The correlation coefficients in the control and experimental arms are the same:  $\rho_0 = \rho_1 = \rho$

where  $\Phi(\cdot)$  is defined as the cumulative distribution function for a standard normal random variable. The power curve gives the probability of rejecting the hypothesis  $H_0 : b(\lambda) \leq 0$  in favour of the hypothesis  $H_1 : b(\lambda) > 0$ , at the level  $\alpha$ , for a given smallest important difference in incremental net benefit value,  $b(\lambda)_\delta$  [11, 25]. Therefore, in the present study, the smallest important difference may be conservatively calculated as a function of the observed  $\Delta E$  and  $\Delta C$  values, because discussion in literature exists regarding the best method of defining  $b(\lambda)_\delta$  [24]. However, because the intent of the present analysis was to examine the practical implications of the correlation coefficient between effectiveness and cost, with respect to a priori sample size or power calculations for cost-effectiveness analysis, the results may be generalizable to other methods of defining  $b(\lambda)_\delta$ . Additionally, as a

**Table 1** Summary of Cost and Survival differences for selected CCTG trials

Trial	n	Survival $E_1$ (Years)		Cost $C_1$ (CAD)		$\rho$
		SE	$\sigma$	SE	$\sigma$	
BR.10	172	0.2930	2.717	3,757	34,840	0.124
BR.21	731	0.0325	0.622	558	10,668	0.203
LY.12 <sup>a</sup>	519	0.0045	0.073	1,807	29,103	0.042
CO.17 (all patients)	567	0.0184	0.309	958	16,130	0.44
CO.17 (KRAS subgroup)	226	0.0296	0.316	1,598	16,987	0.43

$E_1$  Experimental arm effectiveness,  $C_1$  experimental arm cost,  $\rho$  Correlation Coefficient, SE Standard Error,  $\sigma$  Standard deviation, CAD Canadian Dollars

<sup>a</sup> LY.12 is a non-inferiority design. For LY.12, survival is defined as the restricted mean QALY from randomization to stem cell mobilization

**Table 2** Variance of the incremental net benefit estimates by correlation coefficient, assuming a willingness-to-pay threshold of \$100,000

Trial	Variance		Relative Changes
	$\rho = 0$	$\rho = \text{observed}$	
BR.10	1,744,993	1,690,399	3%
BR.21	21,792	20,318	7%
LY.12	6,922	6,800	2%
CO.17 (all patients)	8,571	5,477	36%
CO.17 (KRAS subgroup)	22,780	14,610	36%

$\rho$  Correlation Coefficient

sensitivity analysis the present study also examined a frequentist method of  $b(\lambda)_\delta$ . Briefly, this frequentist method is characterized as the minimum  $b(\lambda)_\delta$  value that satisfies Eq. 6 [6]. Further information on this frequentist method of  $b(\lambda)_\delta$  is available in Lachin [26].

Additionally, the methods examined in the present study may be extended to non-censored data based on the mean parameter estimates, as examined in Willan [6, 11].

**Results**

**Summary of included trials**

In the present analysis, we included all the trials for which CCTG had completed a cost-effectiveness analysis, including empirically calculated variances (of means of cost and effectiveness), correlation coefficients, as well as means of the  $\Delta C$  and  $\Delta E$ . In total, we identified four trials, and one retrospective subgroup analysis from one of the four trials for inclusion. The present analysis found that the standard deviation for the cost of the experimental arm ranged from \$10,000 to \$35,000. The standard deviation for the effectiveness of the experimental arm was observed to range from 0.1-years to 2.72-years. The correlation coefficient between  $\Delta C$  and  $\Delta E$  were

observed to be low to moderate in all the included trials (0.042 to 0.44) (Table 1).

**Correlation coefficient analysis**

The correlation coefficient analysis of the present study examined value-for-money estimates changed, as a function of original-trial cost and effectiveness correlation. We applied the proposed variance formula for health economic to four different CCTG studies: BR.10, BR.21, CO.17 (all patients and a sub study of CO.17 KRAS wild type) and LY.12. First, we examined the impact of the correlation between the  $\Delta C$  and  $\Delta E$  on the variance of  $b(\lambda)$ . The results of the correlation coefficient analysis are summarized in Table 2. Additionally, contour plots examining the impact of  $\rho$  and  $b(\lambda)$  with respect to the variance of  $b(\lambda)$  are presented in Fig. 4 of the Additional file 1: Appendix. Below we use the CO.17 trial (all patients) to demonstrate how the variance of  $b(\lambda)$  changes as a function of the correlation coefficient  $\rho$ , which can vary from -1 (minimum) to 1 (maximum).

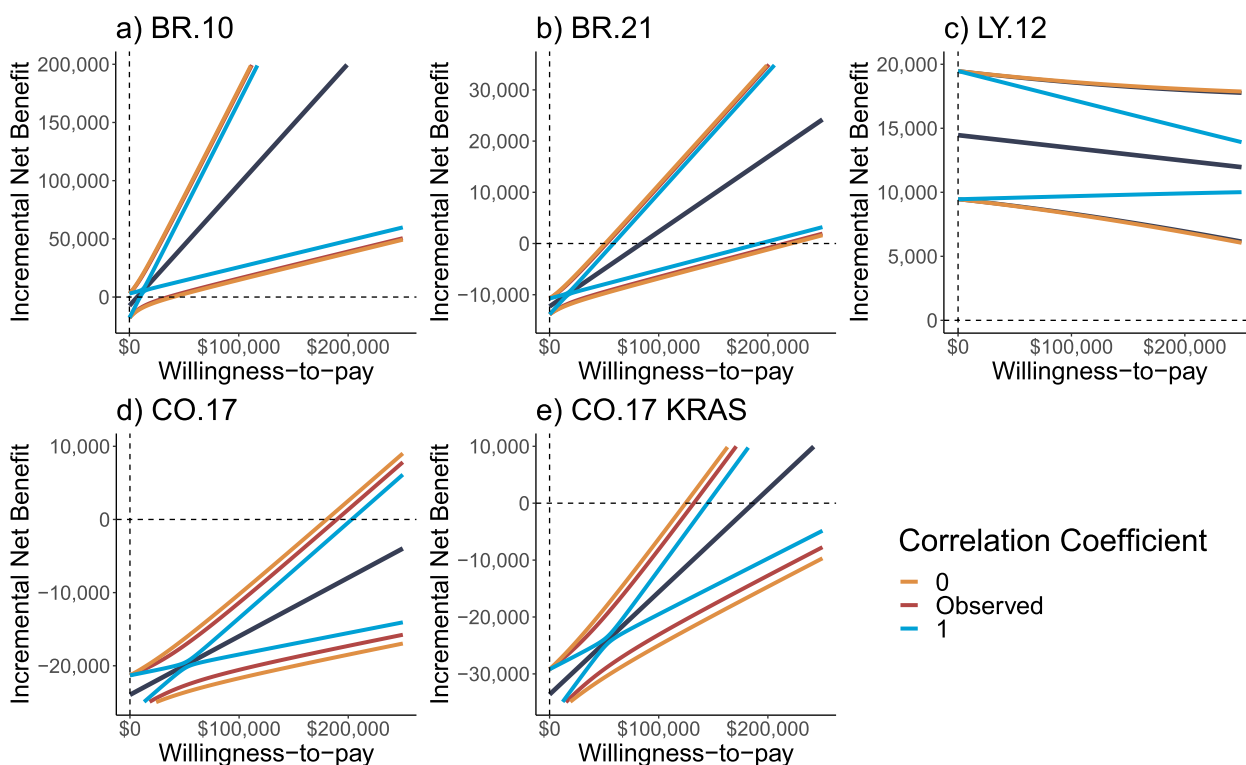
The trial CO.17 is a randomized controlled trial comparing the efficacy of cetuximab plus best supportive care ( $n=287$ ) versus best supportive care ( $n=285$ ) in patients with refractory advanced colorectal cancer. Assuming that the willingness-to-pay is \$100,000 per life-year gained, then the variance of  $b(\lambda)$  is given by

$$\sigma_{b(\lambda)}^2 = 8,574 - 3,516 \cdot \rho$$

When  $\rho = 0$ , the variance  $\sigma_{b(\lambda)}^2 = 8,571$ . When  $\rho = 0.44$ , the variance  $\sigma_{b(\lambda)}^2 = 5,477$ , which is a decrease of 36.1%. When  $\rho = -1$  or 1, the variance will be increased or decrease by 82.0%, respectively. Assuming that the willingness-to-pay is \$200,000 per life-year gained, then the variance of  $b(\lambda)$  is given by

$$\sigma_{b(\lambda)}^2 = 28,778 - 7,032 \cdot \rho$$

Which is a linear function of the correlation coefficient  $\rho$ . When  $\rho = 0$ , the variance is 28,778. When  $\rho = 0.44$ ,



**Fig. 1** Impact of the correlation coefficient on the confidence intervals of the incremental net benefit across willingness-to-pay values

the estimated correlation coefficient from the trial data, the variance  $\sigma_{b(\lambda)}^2 = 21,746$ , which is a decrease of 21.5%. When  $\rho = -1$  or 1, the variance will be increased or decrease by 48.8%, respectively. The relative changes of the variance when  $\rho = 0$  and  $\rho = \text{observed}$ , at a willingness-to-pay value of \$100,000, across all the examined trials is presented in Table 2. In the examined trials, the correlation coefficient accounted for a reduction range of 2% to 36% of the INB variance. Further, based on the CO.17 example, we appreciate that the larger the correlation coefficient, the greater the reduction of INB variance.

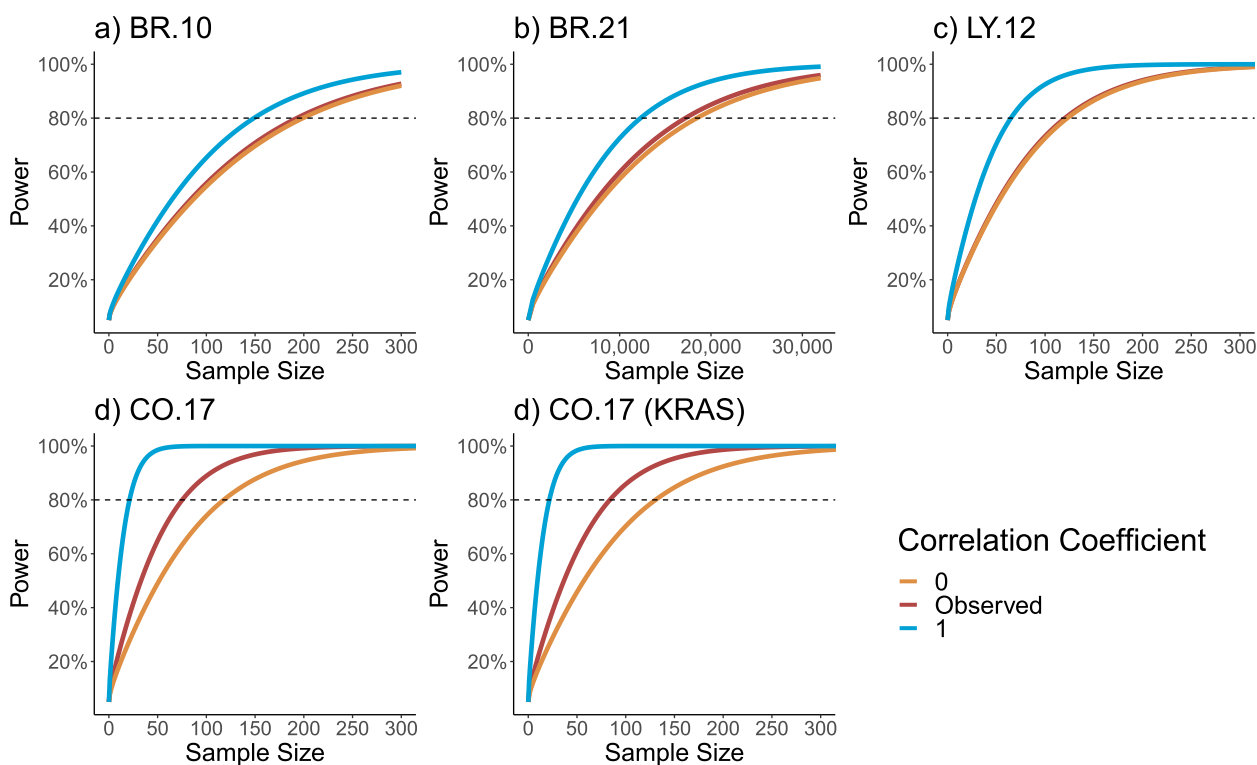
The relationship between the correlation coefficient and confidence intervals of the INB estimate, across a range of willingness-to-pay ( $\lambda$ ) values (\$0 to \$250,000), are examined in Fig. 1. In all the included trials, when the willingness-to-pay increased from \$100,000 to \$200,000, the width of the confidence intervals of the INB estimate also increased. Additionally, in all the included trials when the correlation coefficient was assumed as  $\rho = 0$ , the confidence intervals of the INB estimate, were slightly more conservative (i.e., wider), compared to the observed correlation coefficient value. Further, in all the included trials when the correlation coefficient was assumed as  $\rho = 1$ , the confidence intervals of the INB estimate underestimated the observed confidence interval (i.e.,

too narrow). Inferences with respect to the ICERs of the examined trials are also possible, as the ICER and the Fieller limits are represented as the intersection at the horizontal axis of the INB estimate and confidence intervals, respectively [11, 23]. Additionally, the  $\Delta E$  and  $\Delta C$  values used to derive the INB estimates based on the cost-effectiveness analysis in all the examined trials, are presented in the (Additional file 1: Appendix Table 1).

**Sample size and power calculations**

In order to examine the relationship between the correlation coefficient and the power and sample size of the examined trials, we applied the power formula presented in Eq. 6, to the four CCTG trials (and one subgroup analysis), when  $\rho = 0$ ,  $\rho = \text{observed}$ , and  $\rho = 1$ . This analysis examines the probability of rejecting the null hypothesis  $H_0 : b(\lambda) \leq 0$ , versus  $H_1 : b(\lambda) > 0$  (and  $H_0 : b(\lambda) \geq 0$ , versus  $H_1 : b(\lambda) < 0$  in the CO.17 and CO.17 KRAS trials), at willingness-to-pay threshold of \$100,000 (Fig. 2).

In all the examined trials, when the correlation coefficient was assumed to equal the correlation coefficient observed in the associated cost-effectiveness analysis, the sample size needed to detect a change in the INB value decreased compared to  $\rho = 0$ . In trials with correlation coefficients close to zero (LY.12:  $\rho = 0.042$ ) the



**Fig. 2** Impact of the correlation coefficient on the sample size power function curve, at the willingness-to-pay threshold of \$100,000

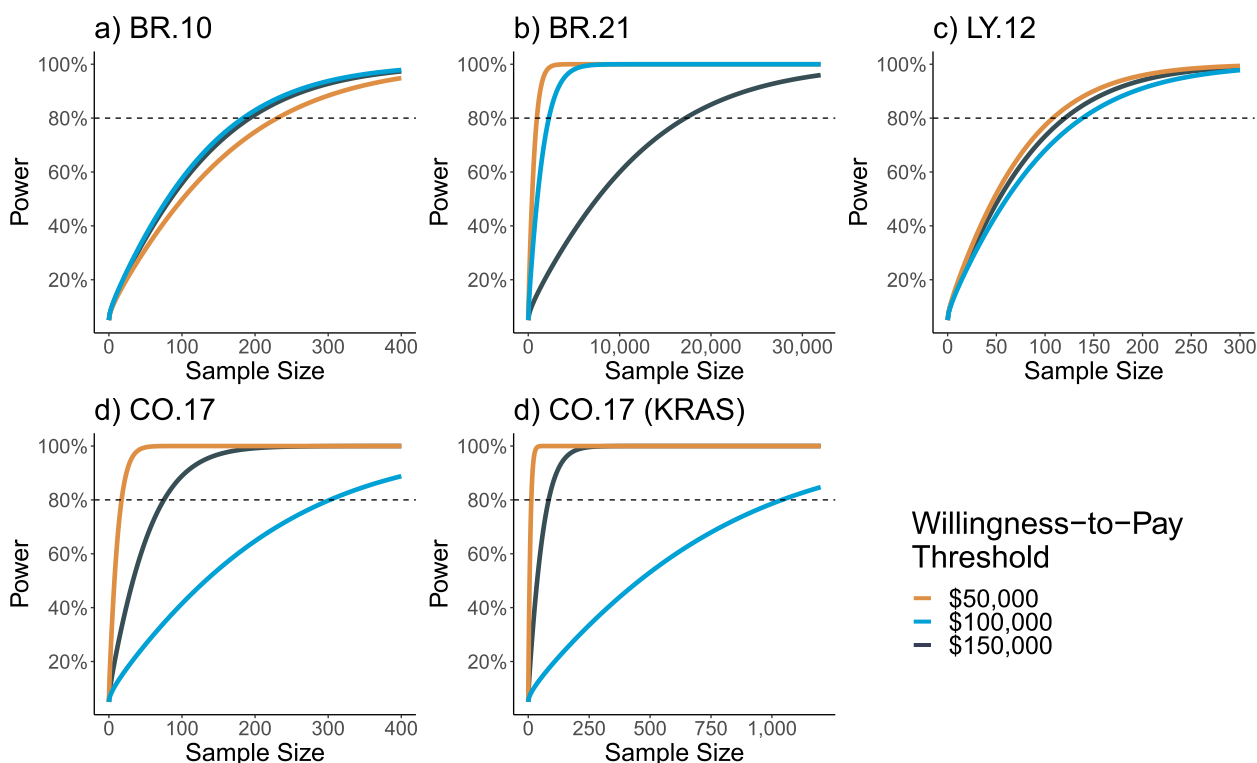
Legend: Panels **a)**, **b)**, and **c)** were examined based on the hypothesis test:  $H_0 : b(\lambda) \leq 0$ , versus  $H_1 : b(\lambda) > 0$ . Panels **d)** and **e)** were examined based on the hypothesis test  $H_0 : b(\lambda) \geq 0$ , versus  $H_1 : b(\lambda) < 0$

sample size required to have 80% power to reject the null hypothesis, changed minimally. However, in trials with relatively higher correlation coefficients (CO.17 (all patients):  $\rho = 0.44$ ) the magnitude of the sample size reduction, in order to have 80% power to reject the null hypothesis, was also relatively higher. Additionally, this trend was also observed in all trials when the  $b(\lambda)_\delta$  value was defined using a frequentist method, (Additional file 1: Appendix Fig. 1). This observation was expected because, based on formula 4, it is straightforward to appreciate the relationship between the sample size and the correlation coefficient  $\rho$ . As the correlation coefficient increases, the required sample size decreases. The relationships between the correlation coefficient and confidence intervals of the INB estimate, across a range of willingness-to-pay ( $\lambda$ ) values (\$0 to \$250,000), are examined in Fig. 1.

In order to examine the relationship between the power and sample size of the included trials across willingness-to-pay thresholds, we also applied the power function presented in Eq. 6, to the four CCTG trials (and one subgroup analysis) at willingness-to-pay thresholds of \$50,000, \$100,000, and \$150,000. Figure 3 examines the probability of rejecting the null

hypothesis  $H_0 : b(\lambda) \leq 0$ , versus  $H_1 : b(\lambda) > 0$  (and  $H_0 : b(\lambda) \geq 0$ , versus  $H_1 : b(\lambda) < 0$  in the CO.17 and CO.17 KRAS trials), assuming the correlation coefficient observed in the cost-effectiveness analysis. The observed cost effectiveness analyses were under-powered (<80%) to reject the null hypothesis at a willingness-to-pay value of \$100,000 in two of the examined trials. Further, when the INB value was close to 0 at the examined willingness-to-pay threshold, as in BR.10 at a willingness-to-pay value of \$100,000, the sample size needed to reject the null hypothesis increased considerably, compared to the other examined trials.

Furthermore, in order to examine the relationship between power and sample size of the examined trials, as a secondary analysis we also examined two-sided hypothesis testing in Additional file 1: Appendix Figs. 2 and 3. Specifically, the null and alternative hypothesis are modelled as  $H_0 : b(\lambda) = 0$ , and  $H_1 : b(\lambda) \neq 0$ . In Fig. 2 the correlation coefficient was modelled as  $\rho = 0$ ,  $\rho = 1$ , and  $\rho = \text{observed}$ , and the willingness-to-pay threshold was examined as \$100,000. In Fig. 3, the correlation coefficient was examined as  $\rho = \text{observed}$  and the willingness-to-pay threshold was examined at \$50,000, \$100,000, and \$150,000. In general, we observed similar trends to those



**Fig. 3** Impact of the willingness-to-pay threshold on the sample size power function curve, at the correlation coefficient observed in the respective trial

Legend: Panels **a**), **b**), and **c**) were examined based on the hypothesis test:  $H_0 : b(\lambda) \leq 0$ , versus  $H_1 : b(\lambda) > 0$ . Panels **d**) and **e**) were examined based on the hypothesis test  $H_0 : b(\lambda) \geq 0$ , versus  $H_1 : b(\lambda) < 0$

examined in the primary analysis. Predictably, the sample sizes observed in the two-sided testing analysis were larger compared to the primary analysis.

**Discussion**

In this paper we examine the application of the INB method for calculating sample size and power in economic evaluations, using original trial data from the CCTG. In general, the present analysis reported that the correlation coefficient between  $\Delta E$  and  $\Delta C$ , had a potentially meaningful impact on the power and sample size calculations in the examined economic analyses. Specifically, when the correlation coefficient increased from  $\rho = 0$  to  $\rho = observed$ , the variance of the INB decreased. Additionally, when the correlation coefficient increased from  $\rho = 0$  to  $\rho = observed$  the sample size needed to detect the smallest important differences in value also decreased. The present analysis examined the INB; however, by varying the willingness-to-pay value, inferences with respect to the ICER are also possible.

Because of the historical absence of a priori cost-effectiveness analysis power calculations in pivotal

cancer trials [5, 11], we have also developed an easy-to-use online program ([http://statapps.tk/icer\\_samplesize](http://statapps.tk/icer_samplesize)), for analysts to calculate power and sample size at the design stage of cost-effectiveness analyses. Additionally, as an illustrative example of the program, possible input parameters for the BR.21 trial is presented in Fig. 5 of the Additional file 1: Appendix. Importantly, the application of our proposed program may be informed by the empirical observations presented in the current study. As examined by Willan (2001), the proposed methods may be used to identify a subset of the trial population for cost-effectiveness analysis, in contrast to historical methods, which required larger sample sizes compared to the corresponding effectiveness analysis, to detect changes in cost-effectiveness [4, 11]. This program will also determine the standard deviation of the effectiveness outcome based on the design of the original clinical trial, including type I error rate, power, survival probability of the control arm at a given time point and the hazard ratio.

In the present analysis, we observed a low to modest correlation between  $\Delta E$  and  $\Delta C$  in all the examined trials. Based on the sample size analysis, assuming the correlation coefficient  $\rho = 0$  may be a conservative estimate to ensure sufficient power of cost effectiveness analyses.



Additionally, when sufficient evidence of indication-specific correlation coefficients exists in literature, power calculations in cost-effectiveness analyses may consider correlation coefficients, in order to design more efficient economic evaluations when survival and costs are positively correlated. Further, as it is uncommon for cost and survival to be negatively correlated, negative correlation coefficients may only be applicable in unique clinical scenarios where a novel intervention would substantially reduce the utilization of a downstream expensive intervention while increasing survival. Assuming a negative correlation coefficient without sufficient biological justification will lead to an overly conservative large sample size, just as assuming a positive correlation coefficient without sufficient biological justification will lead to an insufficiently small sample size.

In the BR.21 trial at a willingness-to-pay threshold of \$100,000, the sample size needed to reject the null hypothesis was considerably larger, compared to the sample size of the actual cost-effectiveness analysis. The interpretation of failing to reject the null hypothesis is that the cost-effectiveness of the treatment, when compared to the control, is not different from the willingness-to-pay threshold. Therefore, we cannot make a conclusion with respect to if a therapy is cost-effective at the examined willingness-to-pay threshold. Additionally, this sample size ballooning occurred when the INB estimate was near zero at the examined willingness-to-pay threshold. This observation may be relevant because novel first-in-class anticancer agents with no a priori pricing information conventionally enter the market with their prices based on the willingness-to-pay threshold [10]. In these scenarios, cost-effectiveness analysis may be under powered to reject the null hypothesis. Therefore, in the absence of a priori sample size or power calculations, cost-effectiveness analyses may not be able to identify if novel treatments are more or less cost-effective compared to the control treatments.

In two of the examined trials (and one subgroup population), the sample size required to reject the null hypothesis, was smaller compared to the corresponding efficacy analysis, at  $\rho = \textit{observed}$  and a willingness-to-pay threshold of \$100,000 (i.e. the cost-effectiveness analyses were over-powered). In practice, it may not be feasible for analysts to embed cost-effectiveness analyses within clinical trials that require sample sizes larger than the corresponding efficacy analyses. However, a priori power calculations in cost-effectiveness analyses embedded within pivotal clinical trials may prioritize the development of cost-reduction therapies, and agents that preserve durable long-term response and survival. Additionally, when economic evaluations are over-powered, a

priori power calculations may facilitate the identification of a population subset for sub-group cost-effectiveness analysis, in order to minimize wasting resources. Further, within a net benefit regression framework, power may also be derived based on a t-test or bootstrapping methods [27, 28]. The net benefit regression framework may also benefit similarly from the observations and proposed considerations presented in the current study.

Importantly, because the empirical parameter values observed in the present study could not be systematically compared to external studies, the validity of these estimates based on trial characteristics is not well established (e.g., the impact of sample size, dropout, or type of outcomes). This limitation highlights the need for future CEAs to report detailed cost and efficacy variance and covariance data, in order to iteratively refine a range of possible values and formal guidelines for CEA analysts. Additionally, changes in price during cost-negotiations may result in changes to the associated sample size and power estimates. Therefore, analysts may consider examining a range of possible parameter estimates.

The proposed a priori cost-effectiveness power and sample size formula utilizes the predicted variance of the cost, effectiveness, and the correlation between cost and effectiveness as well as the expected event rate. The present study addresses the gap in literature with respect to original trial data of cost and effectiveness variances, as well as the correlation between cost and effectiveness.

### Considerations

Based on the empirical exploration of original-trial data, we present six potential considerations for future individual patient-based economic evaluations embedded in clinical trials:

1. At the design stage of clinical trials, embedded CEAs may calculate  $\Delta E$  and standard error of  $\Delta E$  based on the efficacy assumptions for the primary efficacy endpoint statistical design.
2. At the design stage of clinical trials, embedded health care perspective CEAs may calculate  $\Delta C$  based on the expected difference in the duration and costs of the experimental drug/regimen versus the control drug/regimen. If the price of the drug is not yet known at the time of the design, reviewers could identify a marketed drug from a comparable class (e.g., “me-too” drugs may base costs on the associated first-in-class drug) or a range of possible estimates, before adjusting for cost add-ons or offsets.
3. At the design stage of future comparable cancer CEAs embedded in clinical trials, standard error of the control and experimental costs could be based on

the range of values observed in the trials from CCTG as reported in the present study (\$500 to \$4,000).

4. At the design stage of future comparable CEAs embedded in clinical trials, correlation coefficients between  $\Delta C$  and  $\Delta E$ , could be based on the range of values we observed in the present study ( $\rho = 0$  to 0.5), where  $\rho = 0$  will lead to a conservative estimate of sample size when the true  $\rho > 0$ . These estimates may be examined as a range of possible values and should be iteratively refined as additional estimates become available in literature.
5. When the sample size of an examined trial is defined based on the efficacy analysis, the corresponding a priori cost-effectiveness analysis power calculations and potential trial population subset may be completed using the INB approach such as using our online calculator application.
6. Future cost-effectiveness analysis based on individual patient data should consistently report their observed variances and correlation coefficients of  $\Delta E$  and  $\Delta C$  in order to provide additional information to facilitate future power calculations in future studies in related cancer settings.

## Conclusion

Based on our empirical observations of original-trial data, we present six potential considerations for future economic evaluations of clinical trials. Additionally, the online program presented in the paper may facilitate a priori calculations of power and sample size in future cost-effectiveness analyses.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12874-023-01956-y>.

**Additional file 1.**

## Acknowledgements

K.K.W.C. and M.C.C. contributed to the manuscript equally. We affirm that all individuals who contributed significantly to this work are listed as authors.

## Authors' contributions

Substantial contributions to the conception: MCC, KKWC. Substantial contributions to the design: BEC, MCC, KKWC. Substantial contributions to the acquisition and analysis: BEC, LE, AH, MCC, KKWC. Substantial contributions to the interpretation of data: LE, BEC, AEH, MCC, KKWC. Substantial contributions to the drafting and revising: LE, BEC, AEH, MCC, KKWC. All authors read and approved the final manuscript.

## Funding

The Canadian Centre for Applied Research in Cancer Control (ARCC) is funded by the Canadian Cancer Society Research Institute, grant 2015–703549. The funding agreement ensured the authors' independence in designing the study, interpreting the data, writing, and publishing the report.

## Availability of data and materials

The data that support the findings of this study are available from CCTG but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available upon reasonable request to the corresponding author (kelvin.chan@sunnybrook.ca) and with permission of CCTG as outlined in CTG-POL-0043 Data Sharing and Access Policy.

## Declarations

### Ethics approval and consent to participate

Because the present study utilized exclusively secondary non-identifiable information, ethical approval and informed consent was waived by Sunnybrook Research Institute Research Ethics Board. The study was carried out in accordance with the ethical guidelines of Sunnybrook Research Institute Research Ethics Board.

### Consent for publication

Not applicable.

### Competing interests

The authors declare no competing interests.

### Author details

<sup>1</sup>Odette Cancer Centre, Sunnybrook Health Sciences Centre, 2075 Bayview Ave, Toronto, ON M4N 3M5, Canada. <sup>2</sup>Department of Public Health Sciences, Canadian Cancer Trials Group, Queen's, University, Kingston, ON, Canada. <sup>3</sup>University of Toronto, Toronto, ON, Canada. <sup>4</sup>Canadian Centre for Applied Research in Cancer Control, Toronto, ON, Canada. <sup>5</sup>Cancer Care Ontario, Toronto, ON, Canada.

Received: 3 October 2022 Accepted: 20 May 2023

Published online: 03 August 2023

## References

1. Cheung M, Chan KKW. Measuring value and benefit—a matter of perspective. *The Lancet Oncology*. 2017;18:839–40.
2. Saluja R, et al. Examining trends in cost and clinical benefit of novel anticancer drugs over time. *J Oncol Pract*. 2018;14(5):e280–94.
3. Lachin JM. Introduction to sample size determination and power analysis for clinical trials. *Control Clin Trials*. 1981;2(2):93–113.
4. Willan AR. Analysis, sample size, and power for estimating incremental net health benefit from clinical trial data. *Control Clin Trials*. 2001;22(3):228–37.
5. Ramsey SD, Wilke RJ, Glick H, Reed SD, Augustovski F, Jonsson B, Briggs A, Sullivan SD. Cost-effectiveness analysis alongside clinical trials II—an ISPOR Good Research Practices Task Force report. *Value in Health*. 2015;18(2):161–72.
6. Willan AR. Sample size determination for cost-effectiveness trials. *Pharmacoeconomics*. 2011;11:933–49.
7. O'Hagan A, Stevens JW. Bayesian assessment of sample size for clinical trials of cost-effectiveness. *Med Decis Making*. 2001;21(3):219–30.
8. Cheung MC, et al. Minimization of resource utilization data collected within cost-effectiveness analyses conducted alongside Canadian cancer trials group phase III trials. *Clin Trials*. 2021;18(4):500–4.
9. Willan AR. Incremental net benefit in the analysis of economic data from clinical trials, with application to the CADET-hp trial. *Eur J Gastroenterol Hepatol*. 2004;16(6):543–9.
10. Briggs A. *Statistical methods for cost-effectiveness research: a guide to current issues and future developments*. London: OHE; 2003.
11. Willan AR, Lin D. Incremental net benefit in randomized clinical trials. *Stat Med*. 2001;20(11):1563–74.
12. Winton T, et al. Vinorelbine plus cisplatin vs. Observation in resected non-small-cell lung cancer. *N Engl J Med*. 2005;352(25):2589–97.
13. Shepherd FA, et al. Erlotinib in previously treated non-small-cell lung cancer. *N Engl J Med*. 2005;353(2):123–32.
14. Jonker DJ, et al. Cetuximab for the treatment of colorectal cancer. *N Engl J Med*. 2007;357(20):2040–8.

15. Crump M, et al. Randomized comparison of gemcitabine, dexamethasone, and cisplatin versus dexamethasone, cytarabine, and cisplatin chemotherapy before autologous stem-cell transplantation for relapsed and refractory aggressive lymphomas: NCIC-CTG LY. 12. *J Clin Oncol*. 2014;32(31):3490–6.
16. Ng R, Hasan B, Mittmann N, Florescu M, Shepherd FA, Ding K, Butts CA, Cormier Y, Darling G, Goss GD, Inculter R. Economic Analysis of NCIC CTG JBR. 10: A Randomized Trial of Adjuvant Vinorelbine Plus Cisplatin Compared With Observation in Early Stage Non–Small-Cell Lung Cancer—A Report of the Working Group on Economic Analysis, and the Lung Disease Site Group, National Cancer Institute of Canada Clinical Trials Group. *J Clin Oncol*. 2007;25(16):2256–61.
17. Bradbury PA, Tu D, Seymour L, Isogai PK, Zhu L, Ng R, Mittmann N, Tsao MS, Evans WK, Shepherd FA, Leighl NB. Economic analysis: randomized placebo-controlled clinical trial of erlotinib in advanced non-small cell lung cancer. *J Nat Cancer Inst*. 2010;102(5):298–306.
18. Cheung MC, Hay AE, Crump M, Imrie KR, Song Y, Hassan S, Risebrough N, Sussman J, Couban S, MacDonald D, Kukreti V. Gemcitabine/dexamethasone/cisplatin vs cytarabine/dexamethasone/cisplatin for relapsed or refractory aggressive-histology lymphoma: cost-utility analysis of NCIC CTG LY. 12. *JNCI: Journal of the National Cancer Institute*. 1;107(7). 2015.
19. Mittmann N, Au HJ, Tu D, O'Callaghan CJ, Isogai PK, Karapetis CS, Zalcberg JR, Evans WK, Moore MJ, Siddiqui J, Findlay B. Prospective cost-effectiveness analysis of cetuximab in metastatic colorectal cancer: evaluation of National Cancer Institute of Canada Clinical Trials Group CO. 17 trial. *J Nat Cancer Inst*. 2009;101(17):1182–92.
20. Glick HA. Sample size and power for cost-effectiveness analysis (part 1). *Pharmacoeconomics*. 2011;29:189–98.
21. Briggs AH, Gray AM. Power and sample size calculations for stochastic cost-effectiveness analysis. *Med Decis Making*. 1998;18(2\_Suppl):S81–92.
22. Lee KM, McCarron CE, Bryan S, Coyle D, Krahn M, McCabe C. Guidelines for the Economic Evaluation of Health Technologies: Canada—4th Edition. Ottawa: CADTH; 2019.
23. Zethraeus N, Löthgren M. On the Equivalence of the Net Benefit and the Fieller's Methods for Statistical Inference in Cost-Effectiveness Analysis. No 379, Working Paper Series in Economics and Finance from Stockholm School of Economics. 2000.
24. Willan AR, Briggs AH. Statistical analysis of cost-effectiveness data. Chichester: Wiley; 2006.
25. Laska EM, Meisner M, Siegel C. Power and sample size in cost-effectiveness analysis. *Med Decis Making*. 1999;19(3):339–43.
26. Lachin JM. Introduction to sample size determination and power analysis for clinical trials. *Controlled clinical trials*. 1981;2(2):93–113.
27. Hoch JS, Rockx MA, Krahn AD. Using the net benefit regression framework to construct cost-effectiveness acceptability curves: an example using data from a trial of external loop recorders versus Holter monitoring for ambulatory monitoring of "community acquired" syncope. *BMC Health Serv Res*. 2006;6(1):1–8.
28. Hoch JS, Hay A, Isaranuwachai W, Thavorn K, Leighl NB, Tu D, Trenaman L, Dewa CS, O'Callaghan C, Pater J, Jonker D. Advantages of the net benefit regression framework for trial-based economic evaluations of cancer treatments: an example from the Canadian Cancer Trials Group CO. 17 trial. *BMC cancer*. 2019;19(1):1–9.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

