Open Access Medical Statistics

Dovepress
open access to scientific and medical research

Open Access Full Text Article

METHODOLOGY

Setting sample size using cost efficiency in fMRI studies

This article was published in the following Dove Press journal:
Open Access Medical Statistics
19 May 2012
Number of times this article has been viewed

Qing Guo[1,3]
Geoffrey Hall[2]
Margaret McKinnon[2,4,7]
Lehana Thabane[1,3,5]
Ron Goeree[1,5,6]
Eleanor Pullenayegum[1,3,5]

[1]Department of Epidemiology and Biostatistics, [2]Department of Psychiatry and Behavioral Neurosciences, McMaster University, Hamilton, ON, Canada; [3]Biostatistics Unit, [4]Mood Disorders Program, [5]Center for Evaluation of Medicine, [6]Programs for Assesment of Technology in Health (PATH) Research Institute, St, Joseph's Healthcare, Hamilton, ON, Canada; [7]Kunin-Lunenfeld Applied Research Unit, Baycrest, Toronto, ON, Canada

Background: Sample size calculations are rarely performed for functional magnetic resonance imaging studies involving clinical populations. This may be due to uncertainty as to the size of expected effect and the variance of the blood oxygenation level dependent response. Moreover, existing sample size methods ignore the costs associated with performing the proposed study. The current paper describes how cost efficiency, a recently proposed method, can be used in conjunction with existing methods to address these issues.

Methods: Cost efficiency is the ratio of a study's value to its cost, and sample size is chosen to maximize cost efficiency (ie, to maximize return on investment). It is suggested that sample size calculations begin by calculating the sample sizes required to achieve a given power, through varying the input parameters to the calculation over their plausible ranges. Cost efficiency can then help narrow the resulting range of sample sizes and help choose one sample size. The approach is illustrated through a recent functional magnetic resonance imaging study of autobiographical memory retrieval in patients with major depressive disorder.

An example: Setting power to 80% and type 1 error rate to 5%, the method of Mumford and Nichols was used to calculate sample size. There were no reported effect sizes for similar studies in the literature; consequently, this parameter was varied over its plausible range (Cohen's *d* varying from 0.2 to 0.8). This yielded sample sizes ranging from 50 to 800. Within these, cost efficiency gave a sample size of 88.

Conclusion: Poor reporting of the input parameters to power-based methods of sample size determination results in a wide range of candidate sample sizes. The cost efficiency approach supplies a way of narrowing this range and choosing a sample size from that.

Keywords: cost efficiency, sample size, power, fMRI studies

Introduction

Functional magnetic resonance imaging (fMRI) has been widely used to examine patterns of neural activation at rest and while performing motor and cognitive tasks. Despite the widespread use of fMRI technology, sample size calculations for fMRI studies have proved challenging. fMRI studies commonly focus on the effect of a stimulus or effects of different stimuli on the blood-oxygenation-level-dependent (BOLD) response of neural regions of interest, often contrasting stimulus-specific (eg, high versus low working memory load) and group effects (eg, patients versus controls). Data are collected sequentially during a scanning session, and at each time point the BOLD response is measured for each voxel in the brain. Given that the change in BOLD in response to a stimulus will vary across brain regions and time, analyzing fMRI data is more complicated than the simple comparisons of means that are required

Correspondence: Qing Guo
Department of Epidemiology and Biostatistics, Faculty of Health Sciences, McMaster University, Hamilton, ON, Canada
Email guoq@mcmaster.ca

for many clinical studies. Here, the correlation in responses over time and space and the multiplicity inherent in fMRI data make sample size determination difficult.[1]

Despite these challenges, there has been progress in sample size estimation in fMRI studies.[2–5] However, these calculations are subject to the same limitations as conventional sample size calculations. The input parameters (eg, minimal clinically important difference, standard deviation) are often unknown. The common approach to uncertainty in input parameters is a sensitivity analysis in which the parameter values are varied over a plausible range, thus producing a range of candidate sample sizes. The difficulty with this approach in fMRI studies is that the innovative nature of the field, coupled with limited reporting, often leads to a high degree of uncertainty as to the values of the input parameters, resulting in a very large range of sample sizes. Thus, sensitivity analysis can be of limited use when planning studies in which a budget needs to be determined a priori.

A further limitation of existing methods is that cost, which investigators cannot ignore in practice, has no role in conventional sample size approaches. Two methods consider cost in sample size estimation, yet neither is currently used in fMRI studies. One is value of information and the other is cost efficiency. Value of information chooses the sample size to maximize the expected value of information gained through the trial minus the expected cost incurred. Value of information methods are most widely used with randomized controlled trials that are conducted to inform a decision on whether to adopt a new intervention (eg, a treatment or diagnostic tool) into routine practice. Here, study value is often measured in terms of the quality-adjusted life years gained by society as a result of the information gathered in the study. For example, the study might show that there are serious side effects associated with the new intervention, leading to a decision to retain the standard treatment, thus saving quality-adjusted life years by avoiding the introduction of a potentially harmful intervention. In fMRI studies, quantifying the expected study value (ie, quality-adjusted life years saved) is difficult since fMRI is often used at an early stage of discovery when it is unclear how the information will be used in improving patient care. Cost efficiency, recently proposed by Bacchetti et al,[6] aims to maximize the value-per-unit cost and can be implemented without quantifying the projected study value. While the concept of cost efficiency is not widely used in health research, maximizing expected return on investment is a criterion most people consider when investing their own money.

This paper discusses the potential use of cost efficiency for setting sample size in fMRI studies. The mathematical basis for the cost efficiency method is explained, and its use is illustrated through an fMRI study. The potential role of cost efficiency in fMRI studies is discussed and some conclusions are offered.

Methods

The cost efficiency approach suggested by Bacchetti et al focuses on the ratio of the value of information to the cost, thereby aiming to maximize return on investment.[6] Let v_n be the expected scientific, clinical, or practical value of the study if the sample size is n, and let c_n be the corresponding cost of the study. Cost efficiency chooses n to maximize the ratio of v_n to c_n (ie, to maximize v_n/c_n).

Cost in this context is measured from the perspective of the investigator and thus includes all financial expenditures: the fixed cost to set up and administer the study, perform data analysis, and disseminate the findings, as well as the cost per patient to cover scanning and travel costs. The study value v_n is measured in terms of the information gained from the study results. As discussed above, study value is most easily quantified in studies that will be used to inform a decision as to whether to adopt a new intervention. In contrast, fMRI studies tend to be used at an earlier stage of discovery, and the information gained could ultimately benefit patients in many possible ways. Although this information is certainly valuable to society, it is difficult to quantify its value. The advantage of the cost efficiency approach is that study value does not need to be quantified.

The key concept in cost efficiency is that the study value v_n can often be replaced by a simple stand-in function of sample size. It has been shown that if a function $f(n)$ can be found such that $v_n/f(n)$ does not increase as n increases, then choosing the sample size n to minimize $c_n/f(n)$ provides more cost efficiency (ie, a higher ratio of v_n/c_n) than any larger choice of n.

Two widely applicable choices of $f(n)$ have been proposed. The first option is $f(n) = n$ with a resulting sample size $n_{\min}$, which minimizes total study cost divided by sample size (average cost per subject). Using $f(n) = n$ requires that v_n/n be nonincreasing in n, and this condition is denoted as $C_{\min}$. $C_{\min}$ has been shown to hold under a wide range of definitions of study value, such as value proportional to study power,[6,7] inversely proportional to confidence interval width, proportional to reduction in Bayesian credible interval width from its prior width,[8–10] proportional to the reduction in squared error loss versus using the prior mean,

and proportional to gain in Shannon information.[11] Thus, Bacchetti et al conclude that it is reasonable to use $n_{\min}$ without verifying condition $C_{\min}$ for each specific study.[6] The second choice is to take $f(n) = \sqrt{n}$ so that n_{root} minimizes total study cost divided by the square root of the sample size. When $f(n) = \sqrt{n}$, it is required that $v_n / \sqrt{n}$ be nonincreasing in n, and this condition is denoted as C_{root}. The condition C_{root} is more stringent than $C_{\min}$ and generally holds for all $n > 4$ when there is low prior information (in the sense that the Bayesian priors have equal prior means for the two groups, and at least one prior having a standard deviation at least as large as its mean). Bacchetti et al also suggest using n_{root}, with no need to verify C_{root} for each specific study provided that there is low prior information.[6]

Bacchetti et al argue that the sample sizes $n_{\min}$ and n_{root} are more cost-efficient than any larger sample size calculated by another sample size determination method, and therefore cannot be considered inadequate, regardless of the power they achieve.[6] In contrast, the current paper suggests using cost efficiency alongside conventional power calculations. Studies with low power may not detect the effect of interest. Since studies without statistically significant results are less likely to be published than studies with significant findings, investigators may feel driven to produce significant results. When power on the primary comparison is low, this encourages data dredging with a danger of spuriously significant findings.[12–15] Thus, it is suggested that in fMRI studies, cost efficiency considerations be used in conjunction with power-based methods.

Specifically, it is proposed that investigators begin with sample size calculations based on achieving a given statistical power. Due to the uncertainty of input parameters, a range of candidate sample sizes are calculated through a sensitivity analysis. Cost efficiency can then be used to choose a sample size from this range. This approach is now illustrated through an example.

An example

The example is a 2-year neuroimaging study of autobiographical memory retrieval in patients with major depressive disorder. This study examined patterns of neural activation during autobiographical memory retrieval of negative events compared to positive and neutral events in patients with recurrent major depressive disorder and matched healthy controls. It was hypothesized that the difference in activation of the hippocampus for negative events compared to positive and neutral events would be greater for patients than controls. In a prescan interview, participants identified six events from the past 2 years: two highly positive events (eg, a birthday party), two highly negative events (eg, receiving bad news), and two comparatively neutral events (eg, swimming). During the scanning session, over a 20-second period, participants were presented with event period titles describing one of the events identified in the interview and asked to recall the event, or an incomplete sentence to be completed. This was followed by 10 seconds during which patients rated their degree of autobiographical reexperiencing and a 5-second fixation. The stimuli was presented in a six-block format in a random order, with two runs per event type and each run containing five stimuli corresponding to one of two memories generated in the prescan interview. The total scan time, including set up and localization, was about 1 hour.

How sample size can be determined will now be illustrated. To begin with, the sample size method of Mumford and Nichols was used,[5] ie, using sensitivity analysis to deal with input parameters whose values are uncertain. Cost efficiency was then applied to help narrow the range of sample sizes.

Mumford and Nichols' sample size calculation requires estimates of within- and between- subject variances, and the size of the effect to be detected.[5] No information about these estimates was reported in prior studies in this population, and so for the purposes of illustration the values presented by Mumford and Nichols were used: a first-order autoregressive correlation of 0.73, a first-order autoregressive total variance of 0.98, white noise variance of 1.313, and between-subject variance of 0.421.

There was substantial uncertainty about the likely size of the difference between patients and controls in the relative activation of the hippocampus for positive versus neutral events. Values of Cohen's *d* effect size ranging from 0.2 (small) to 0.8 (large) were plausible, and this parameter was varied in sensitivity analyses. Using values of 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, and 0.8 yielded sample sizes of 786, 350, 198, 126, 88, 66, and 51, respectively, to achieve 80% power at a significance level of 0.05. The type 1 error rate was not adjusted for multiple comparisons because the primary region of interest was restricted a priori to the hippocampus. The relationships between power and sample size for different effect sizes are displayed in Figure 1.

Since these calculations yielded sample sizes ranging from 50 to 800, additional criteria were needed to select an appropriate sample size. Cost efficiency can help with this. In this study, the cost components included fixed costs such as part-time research assistant's salary, which covered the responsibilities of scheduling, subject recruitment, data entry,

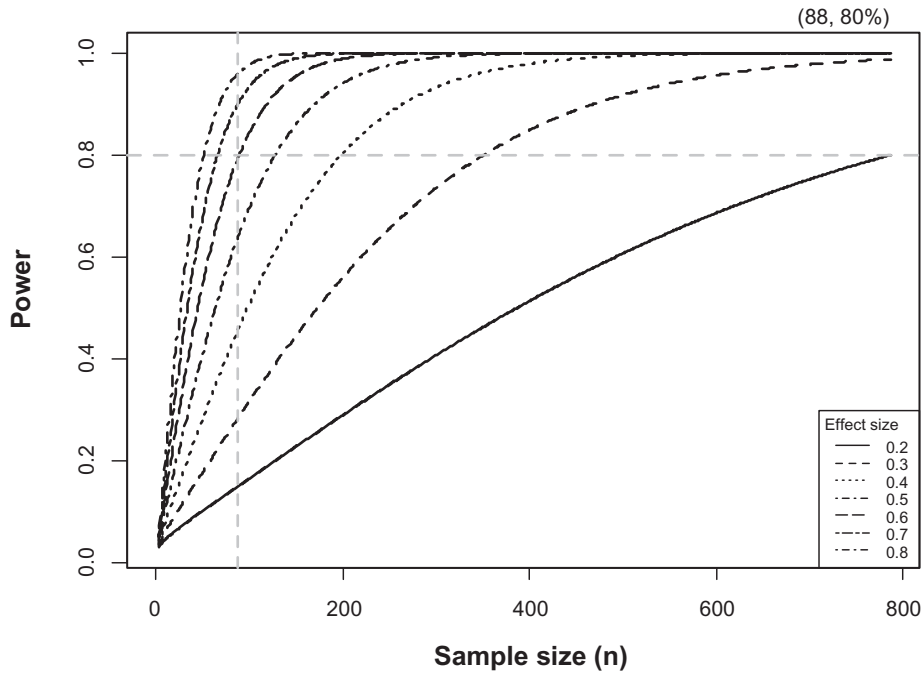

Figure 1 Power estimates for a block study for different sample sizes.
Notes: Each curve is for different group effect sizes (Cohen's *d*). The horizontal grey dotted line indicates 80% power and the vertical grey dotted line represents most cost-efficient sample size (n_{root} = 88).

data analyses, conference travel fees, and test administration (in total, $39,671). The cost per patient included participant reimbursement fees to accommodate study participants' parking, traveling and testing time ($50 per person), and scanner time ($400 per person). Therefore the total cost was $39,671 plus $450 per patient. Since this was the first neuroimaging study of autobiographical memory retrieval in patients with major depressive disorder, n_{root} was a reasonable choice. It is easy to show (Appendix 1) that n_{root} is equal to fixed costs divided by variable costs (ie, $39,671 is divided by $450), which in the current study came to 88 patients (Figure 2).

Here it is demonstrated step-by-step how n_{root} helps to choose a cost-efficient sample size. As studies with low

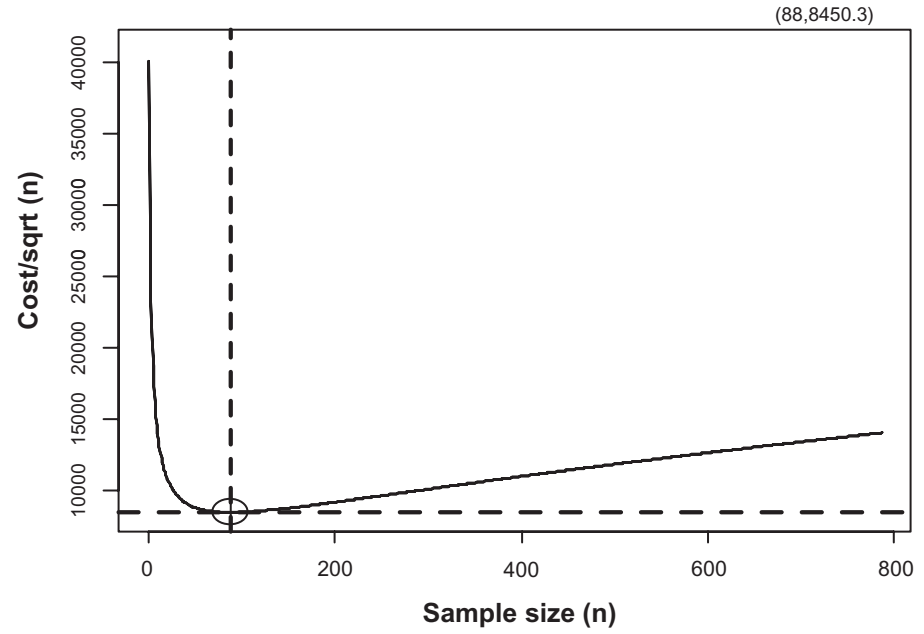

Figure 2 The relationship between the cost divided by the square root of the sample size and sample size for all assumed group effect sizes.

power will have small sample sizes, yielding results with a wide confidence interval (ie, low precision), using the reduction from the width of confidence interval is a measure of study value. Therefore, the study value was defined as inversely proportional to confidence interval width in this demonstration. The parameter of interest was the difference between patients and controls in the mean activation of the hippocampus for negative events compared to positive and neutral events (ie, the emotional valence by group interaction). It can be observed in Figure 3 that study value divided by the square root of the sample size was monotone, decreasing as sample size increased. Figure 4 illustrates that cost divided by the square root of the sample size was minimized at the sample size of 88, and as can be seen from Figure 5, cost efficiency at the sample size of 88 was bigger than any larger choice among the range of sample sizes.

Study value was chosen to be inversely proportional to confidence interval width to illustrate how the cost efficiency method works in practice. Bacchetti et al have shown that similar results hold when using other definitions of study value.[6]

Discussion

Conventional methods of sample size calculation in fMRI studies involving clinical populations are limited by substantial uncertainty in the values of input parameters. Varying these parameters over their plausible ranges can result in a very large range of candidate sample sizes (in the current example, the range was 50 to 800). It is argued that cost efficiency can help choose a sample size from this range. In the current example, a sample size of 88 was more cost-efficient than any larger choice. Thus, cost efficiency provides an upper bound to the sample size that the investigator should consider.

The current approach to using cost efficiency differs from Bacchetti et al's approach. Arguing that the standard choice of 80% for statistical power is arbitrary, Bacchetti et al proposes cost efficiency as a stand-alone method for choosing sample size.[6] The current authors agree that cost efficiency is a reasonable criterion: although an unfamiliar concept in health research, most individuals aim to maximize their cost efficiency (return on investment) when investing their own money. This difference of opinion with Bacchetti et al comes from the belief that the need for significant results to achieve publication is likely stronger in the fMRI literature than in other areas of medicine. This, coupled with the enormous scope for multiple testing in fMRI studies, makes the danger of false positive findings in underpowered fMRI studies particularly acute. It can thus be argued that studies with very low power may in fact have negative value.

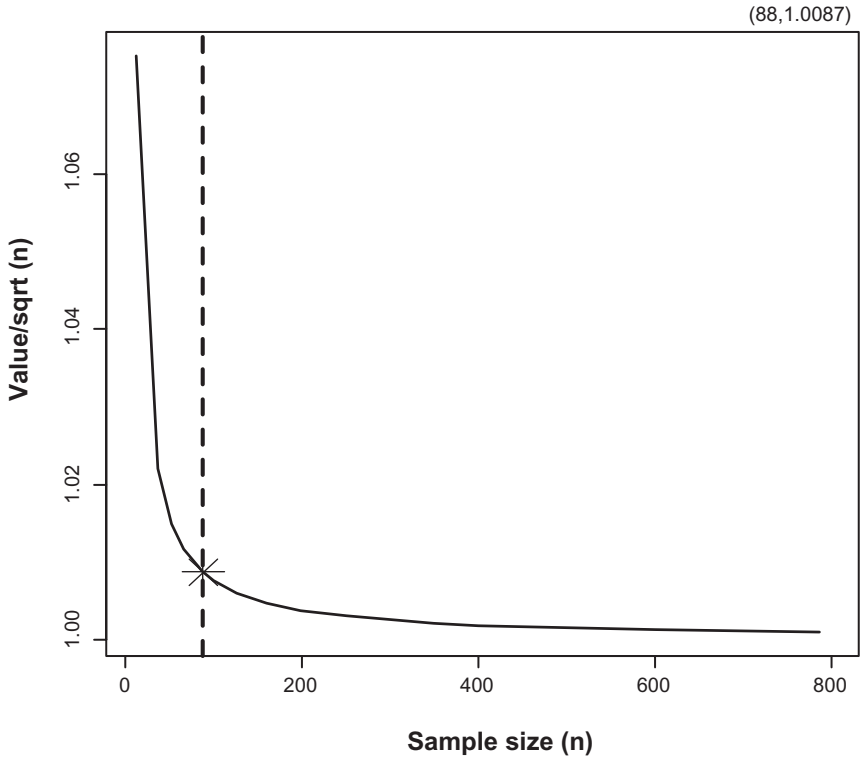

Figure 3 The relationship of the study value (which is inversely proportional to confidence interval width) divided by the square root of the sample size versus sample size.

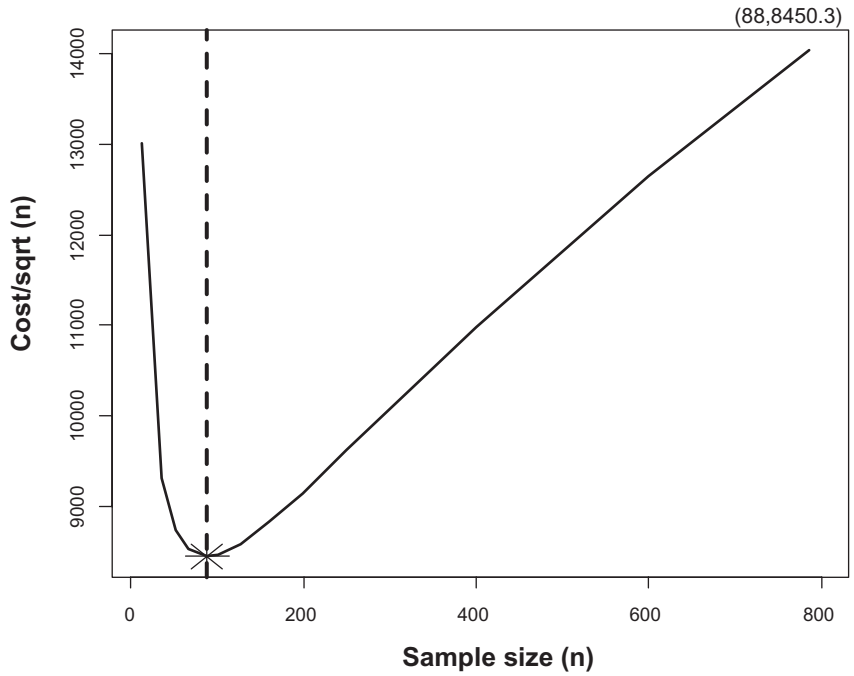

Figure 4 The relationship of cost divided by the square root of the sample size versus sample size.

Thus, it is suggested that cost efficiency considerations be used in conjunction with power-based methods.

The suggested approach of using cost efficiency as a supplement to traditional power-based methods can overcome the potential conflict between a sample size calculated from the traditional method and the cost efficiency sample size. While larger sample sizes will always have greater power, they cost more as well. Consequently, there is a trade-off between the study power and cost. Cost efficiency provides a way of compromising between statistical power and cost.

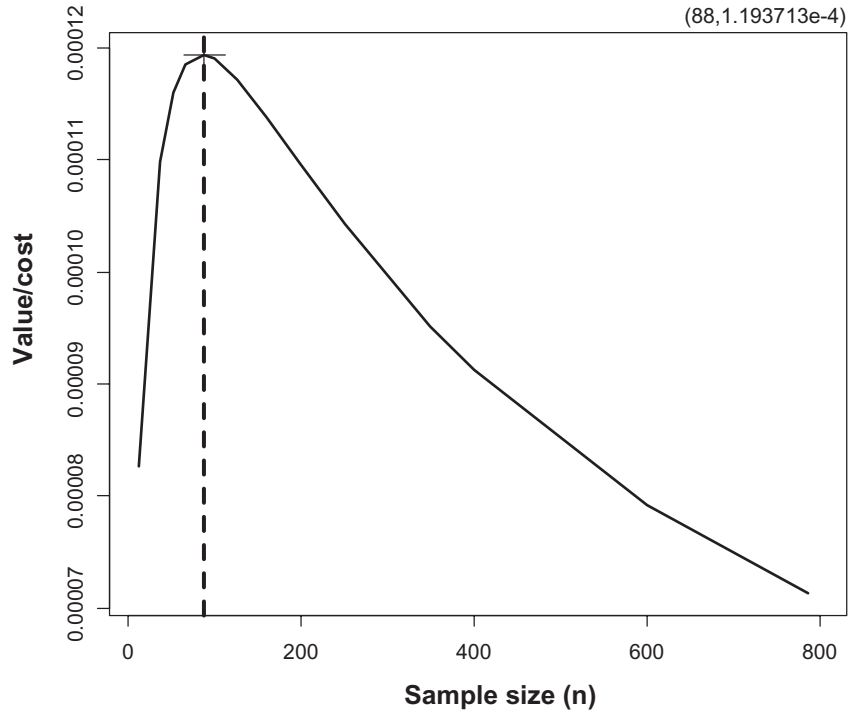

Figure 5 The relationship between cost efficiency and sample size.

Since the authors propose choosing the initial range of sample sizes by using traditional power calculation, and applying the cost efficiency criterion in order to choose a sample size among this range, the current approach will have at least 80% power to detect at least one of the proposed effect sizes.

There are some practical limitations to cost efficiency. First, only financial cost is considered. Societal costs such as inconvenience and risks to study participants are not included. Moreover, in the face of limited resources, funding one study means that another study is not funded; the costs used in cost efficiency calculations do not include this opportunity cost. Second, financial costs must be estimated accurately. This is a particular concern when writing grant applications for agencies that routinely cut budgets, since it is common for investigators to pad their budget against the cut. This will distort cost efficiency calculations. Third, it is possible that the study's projected value is less than the cost incurred. When this happens, the study does not add additional value[16] and should not be undertaken. The cost efficiency approach does not consider this. Fourth, whilst n and $\sqrt{n}$ are chosen as two widely applicable choices of $f(n)$, the authors do not claim that they are optimal, but rather that they are useful because their properties have been evaluated extensively through theory and simulations. It is therefore possible that some other forms of $f(n)$ may exist and do a better job in terms of identifying the sample size with the optimal cost efficiency. Future research may be helpful. Finally, cost efficiency may produce a sample size that is beyond the investigator's budget. This is a problem shared by other methods of sample size determination, in particular methods based on achieving a given power. In the current approach, budgetary constraints could be incorporated by narrowing the range of sample sizes produced by sensitivity analysis down to those that are feasible.

In the current example, sample size calculations were based on hypothetical values for the variance components of the BOLD response. Thus, the results are for the purposes of illustration only. An alternative approach to this might be a Bayesian approach in which a prior is placed on these unknown parameters. Then the probability of rejecting the null hypothesis is calculated as the classical power averaged over the prior distribution.[17] In the current example, it was decided not to adopt the Bayesian approach; the results of such a power calculation are sensitive to the choice of prior, and the lack of reported values of the variance components in the literature makes it extremely difficult to place a realistic prior on them. However, if these variance components were to be more widely reported in the future, Bayesian sample size determination methods would become more viable.

This lack of reported variance components in the literature is a serious concern. If studies do not include these estimates in their results sections, there will remain substantial uncertainty as to their value, leaving investigators unable to power their studies accurately. Moreover, the uncertainty in variance parameters and effect sizes is not because the data to quantify them does not exist, but rather because it is simply not reported. There is a pressing need for a reporting guideline for fMRI studies outlining which values investigators should report to facilitate adequately powered studies in the future.

Conclusion

There is often substantial uncertainty in the effect sizes and variance components that form the input parameters to conventional sample size calculations. For any given study, this can lead to a wide range of sample size estimates. Authors should be encouraged to report effect sizes and estimates of within- and between-subject variances in their manuscripts to facilitate sample size calculations for future studies. Until this practice is widespread, the authors have argued that cost efficiency can supplement conventional sample size methods by narrowing the range of sample sizes under consideration on the basis of maximizing the expected return on investment.

Acknowledgments

Qing Guo was partly supported by funding from the Canadian Institute of Health Research (CIHR) training award and Ontario Graduate Scholarship (OGS).

Disclosure

The authors report no conflicts of interest in this work.

References

1. Lazar N. *The Statistical Analysis of Functional MRI Data*. New York, NY: Springer; 2008.
2. Desmond JE, Glover GH. Estimating sample size in functional MRI (fMRI) neuroimaging studies: statistical power analyses. *J Neurosci Methods*. 2002;118(2):115–128.
3. Murphy K, Garavan H. An empirical investigation into the number of subjects required for an event-related fMRI study. *Neuroimage*. 2004; 22(2):879–885.
4. Hayasaka S, Peiffer AM, Hugenschmidt CE, Laurienti PJ. Power and sample size calculation for neuroimaging studies by non-central random field theory. *Neuroimage*. 2007;37(3):721–730.
5. Mumford JA, Nichols TE. Power calculation for group fMRI studies accounting for arbitrary design and temporal autocorrelation. *Neuroimage*. 2008;39(1):261–268.
6. Bacchetti P, McCulloch CE, Segal MR. Simple, defensible sample sizes based on cost efficiency. *Biometrics*. 2008;64(2):577–594.

7. Bacchetti P, Wolf LE, Segal MR, McCulloch CE. Ethics and sample size. *Am J Epidemiol*. 2005;161(2):105–110.
8. Joseph L, Belisle P. Bayesian sample size determination for normal means and differences between normal means. *Statistician*. 1997;46(2):209–226.
9. Lindley DV. The choice of sample size. *Statistician*. 1997;46(2):129–138.
10. Pham-Gia T. On Bayesian analysis, Bayesian decision theory and the sample size problem. *Statistician*. 1997;46(2):139–144.
11. Bernardo JM. Statistical inference as a decision problem: the choice of sample size. *Statistician*. 1997;46(2):151–153.
12. Dickersin K, Chan S, Chalmers TC, Sacks HS, Smith H Jr. Publication bias and clinical trials. *Control Clin Trials*. 1987;8(4):343–353.
13. Dickersin K. The existence of publication bias and risk factors for its occurrence. *JAMA*. 1990;263(10):1385–1389.
14. Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR. Publication bias in clinical research. *Lancet*. 1991;337(8746):867–872.
15. Ioannidis JP. Why most published research findings are false. *PLoS Med*. 2005;2(8):e124.
16. Briggs A, Sculpher M, Claxton K. *Decision Modelling for Health Economic Evaluation*. Oxford, UK: Oxford University Press; 2006.
17. Spiegelhalter DJ, Abrams KR, Myles JP. *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. Chichester, UK: John Wiley and Sons Ltd; 2004.

Appendix 1

If the study is composed of a linear structure of a fixed cost plus a per-patient cost, that is, $C_n = C_f + C_s \times n$, where C_f – the fixed cost independent of n – is greater than zero and C_s – the cost per subject – is greater than zero, then n_{root} is the ratio of C_f to C_s, that is, $n_{root} = C_f / C_s$.

Proof: If $\frac{\partial\left(\frac{C_n}{\sqrt{n}}\right)}{\partial n} = \frac{C_s \times \sqrt{n} - (C_s \times n + C_f)}{n} \times \frac{1}{2} \times n^{\frac{1}{2}} = 0,$

then $n = C_f / C_s$; and $\frac{\partial^2\left(\frac{C_n}{\sqrt{n}}\right)}{\partial n^2} = \frac{3 \times C_f - C_s \times n}{4 \times n^{\frac{5}{2}}} > 0$ at $n = C_f / C_s$. Therefore, $C_n/\sqrt{n}$ reaches the local minimum at $n = C_f / C_s$. Thus, $n_{root} = C_f / C_s$.