Contemporary Clinical Trials 33 (2012) 213–217

Contents lists available at SciVerse ScienceDirect

Contemporary Clinical Trials

journal homepage: www.elsevier.com/locate/conclintrial

Strategy for power calculation for interactions: Application to a trial of interventions to improve uptake of bowel cancer screening

A.R. Brentnall [a,*], S.W. Duffy [a], G. Baio [b], R. Raine [c]

[a] Queen Mary University of London, Centre for Cancer Prevention, Wolfson Institute of Preventive Medicine, Barts and The London, Charterhouse Square, London EC1M 6BQ, UK

[b] University College London, Department of Statistical Science, Gower Street, London WC1E 6BT, UK

[c] University College London, Department of Epidemiology and Public Health, 1–19 Torrington Place, London, WC1E 7HB, UK

ARTICLE INFO

Article history:
Received 30 April 2011
Received in revised form 21 September 2011
Accepted 27 September 2011
Available online 8 October 2011

Keywords:
Deprivation
Logistic regression
Ordinal
Sample size
Simulation

ABSTRACT

Poorer postcodes within 5 regions in England have a lower response to bowel-cancer screening invitations than do richer postcodes. An extension of the sample-size formula for two proportions is used to determine that needed to detect an increase in response rate that varies by deprivation quintile. The proportions plugged into the formula are weighted averages based on the relationship between response and deprivation; the response rate is adjusted to be constant across deprivation quintiles. From a baseline period between October 2006 and January 2009, detection of an absolute or relative increase of at least 1,2,3,4 and 5% in response rate is required for the richest to poorest quintiles respectively because the interventions were chosen as those most likely to have an effect in the lower socioeconomic groups. A computer simulation experiment shows that the approach is more conservative than a likelihood-ratio calculation, and it appears sensible when compared with repeated application of a two-sample calculation at each quintile.

1. Introduction

The rate at which individuals respond to invitations for bowel-cancer screening, π, is currently linked to a measure of deprivation x. Fig. 1 shows the response rates by region in England between October 2006 and January 2009. The observed differences between regions are statistically significant because the sample size is more than 2 millions, with a minimum of 30 thousand invitations in each deprivation and centre combination. Possible mechanisms underlying the social inequalities in cancer screening generally are explored in the review by [1]. Policy makers would like to change the pattern by trialling new approaches to invitation.

The application considered here is a trial of an experimental health-literacy intervention, targetted at people with low health literacy. Since there is an association between literacy and socio-economic status [2,3], the intervention is expected to improve take up in higher deprivation groups; [4] and [5] show a relationship between literacy and motivation and ability to engage with information about bowel-cancer screening. One feature of the intervention is that it will use language and a story that is relevant to these groups. Although there is no direct evidence to suggest that socio-economic status is associated with preference for narrative information, [6] provide a general review on the possible link.

This article is motivated by the design of a trial to compare two policies $k=0, 1$, in this case the *status quo* or a new intervention, where the statistical problem is to plan the sample size n to use when $n/2$ is apportioned into both groups. That is, to perform a sample-size calculation when a response probability π is linked to a covariate x. In the next section we outline methods for this, and then we apply some of them to determine the sample size of the bowel-cancer screening study. The main contribution of the article is the development of a simple method to estimate sample size when π might depend on deprivation x through an interaction with invitation k, by adjusting a two-sample calculation.

* Corresponding author. Phone: +44 20 7882 3531 and Fax: +44 20 7882 3890.

E-mail addresses: a.brentnall@qmul.ac.uk (A.R. Brentnall), s.duffy@qmul.ac.uk (S.W. Duffy), g.baio@ucl.ac.uk (G. Baio), r.raine@ucl.ac.uk (R. Raine).

1551-7144/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.cct.2011.09.021

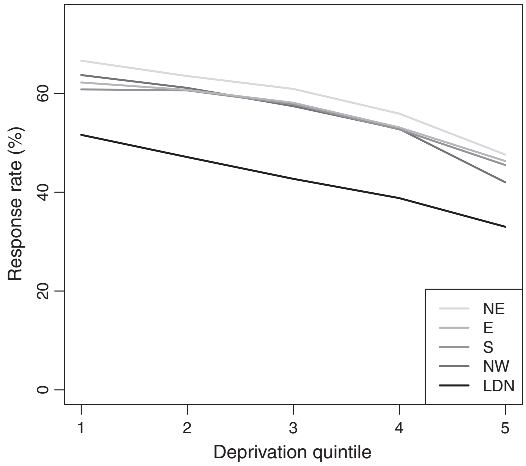

Fig. 1. Bowel screening response rates between October 2006 and January 2009 for regions in England (NE – North East; E – East; S – South; NW – North West; LDN – London) and deprivation quintile from richest (1) to poorest (5). A description of the data is in Section 3.1.

2. Sample size

One way to describe the present relationship between response to invitation to bowel screening π_x and deprivation quintile $x = 0, 1, 2, 3, 4$ in a given geographical area is

$$\mathrm{logit}(\pi_x) = \theta_0 + \theta_1 x \tag{1}$$

where $\theta = (\theta_0, \theta_1)$ are parameters and logit(.) denotes the logistic function. If this model is taken then the criteria used to choose the sample size might be that a change in θ_1 of at least $\theta_1^* - \theta_1$ (where θ_1^* is the new policy) would be detected with probability $1 - \beta$ when a two-sided hypothesis test at level α is used for $H_0 : \theta_1 = \theta_1^*$. A reason for taking the logistic regression model is that it ensures that the model response is between 0 and 1. However, different models and methods might also be applicable, depending on the data. We next consider some sample-size methods for logistic regression models, and then present two approaches to help plan the bowel-screening trial.

2.1. Logistic regression sample size

When there is a single covariate in the logistic regression that is a dichotomous factor or a normal random variable, then Hsieh et al. [7] noted that one or two-sample power calculations may be applied. The situation becomes more complicated for multiple regression. Whittemore [8] proposed a formula for the approximate sample size of rare event logistic regression based on the information matrix. The approach may be used when the covariates arise from a general exponential family distribution. Hsieh [9] gave a modification of Whittemore's formula in the multivariate normal case, to provide approximate maximum sample sizes for multivariate logistic regression, for rare and non-rare events. However, Agresti [10] cautioned that the formula provides only a rough indication of sample size. Further developments followed in [7], where sample size formulae were derived for multiple regression by extending two-sample tests. In particular, the case of a binary variable and normal random variable were considered. This idea was also used in [11], who devised an approach to sample size calculation for generalized linear models and Cox regression. Demidenko [12] gave formulae based on asymptotic information matrix derivations for logistic regression with a binary exposure and one covariate; [13] extended the formulae for their interaction, and also reviewedsome other work on interaction effects in logistic regression models.

The set-up considered here is slightly different to some previous work. For example, the focus in [14] was for a test on whether an interaction was present after allowing for a main effect. The proposed experimental change to the invitation to screening leads us to specify that the difference between interventions is in the interaction term only, *e.g.* for a model such as (1), θ_1 changes, not θ_0.

2.2. General approach based on two-sample test

When the hypothesis involves a single parameter it is likely to be equivalent to a test on the overall proportion π. In model (1) this translates a test of $\theta_1 = \theta_1^*$ to one of $\pi_x = \pi_x^*$ for $x = 0, 1 \ldots, m$ ordinal values, where taking $\mathbf{x} = (1, x)$ and $\theta^* = (\theta_0,\ \theta_1^*)$ we have $\mathrm{logit}(\pi_x) = \theta'\mathbf{x}$ and $logit(\pi_x^*) = \theta^{*\prime}\mathbf{x}$. Given the null model's parameter θ, it is possible to adjust π_x through factors $c_1, \ldots, c_m$ so that $\pi_0 = c_x\pi_x$ for $x = 1, \ldots, m$ when

$$c_x = \frac{1 + \exp(\theta_0 + \theta_1 x)}{\exp(\theta_1 x)\{1 + \exp(\theta_0)\}}. \tag{2}$$

If the null model holds then applying the same factors to the alternative will mean that $\pi_0 = \pi_0^* = \ldots = c_m\pi_0^*$. Thus a possible approach for power calculation is to use a test for the difference in two proportions. The first proportion is

$$\hat{\pi} = (n_1)^{-1}\left(n_{10}\hat{\pi}_0 + \sum_{x=1}^{m} n_{1x}c_x\hat{\pi}_x\right) \tag{3}$$

where for y_x equal to the number of cases that responded out of n_{1x} in each deprivation category x we have $\hat{\pi}_x = y_x/n_{1x}$, and $n_1 = \sum_{x=0}^{m} n_{1x}$ is the total number of individuals in the control group. The alternative proportion is

$$\hat{\pi}^* = (n_2)^{-1}\left(n_{20}\hat{\pi}_0^* + \sum_{x=1}^{m} n_{2x}c_x\hat{\pi}_x^*\right)$$

where similar definitions apply: n_2 is the number of individuals sampled in the new policy group ($n = n_1 + n_2$) and n_{2x} are the number in each deprivation category x ($n_2 = \sum_{x=0}^{m} n_{2x}$). Then

$$\begin{aligned}\mathrm{var}(\hat{\pi}^* - \hat{\pi}) &= (n_1)^{-1}\{\pi(1-\pi)\} + (n_2)^{-1}\{\pi^*(1-\pi^*)\}\\ &= \sigma^2,\end{aligned}$$

and from the central limit theorem

$$z = (\sigma^2)^{-1}\{(\hat{\pi}^* - \hat{\pi}) - (\pi^* - \pi)\}$$

is approximately normally distributed with mean 0 and variance 1. The power for a one-sided test

$$1-\beta = pr\{(\sigma^2)^{-1}(\hat{\pi}^*-\hat{\pi})\geq z_\alpha\}$$

may thereby be obtained given α, n_{10}, …, n_{1m}, n_{20}, …, n_{2m}, π^* and π; power for two-sided tests may also be calculated.

We have constructed $\pi=\pi_0$ using the c-factors, it only remains to determine π^* under the alternative hypothesis. If the difference between the treatments is related via $\theta_1^*=\theta_1\Delta$, where Δ is to be specified, then setting a difference of Δ in the original model would correspond to

$$\pi^* = (n_2)^{-1}\sum_{x=0}^{m} n_{2x}c_x\pi_0^*$$

and the power calculation is obtained by using this in the difference $\pi^*-\pi_0$.

The approach is presented above for logistic regression models because this is a common approach taken for binary data. However, the c-factor approach may be applied to any model for π through $c_x=\pi_0/\pi_x$, where π_x is determined by the model.

2.3. Simulation

Another general-purpose approach to determining sample size is to use computer simulation for whatever statistical model is chosen. In the next section we consider two models. In the first,

$$\pi_x = \theta_x \tag{4}$$

and

$$\pi_x^* = \theta_x(1+\delta x) \tag{5}$$

for $x=1$, …, 5 where θ_1, …, θ_5, δ are parameters. That is, the basic model allows response rates to vary by deprivation quintiles x ($m=5$), but the new policy changes response rates at each of them by an amount linked to δ, which is equivalent to setting $\theta_0=0$ in Eq. (1). The second model replaces Eq. (5) by

$$\pi_x^* = \theta_x + \delta x. \tag{6}$$

The two models correspond to relative and absolute percentage increases in response rate, where both are assumed to occur following a linear trend.

A likelihood-ratio test may be constructed to compare the models against a null model where (4) is used for both treatment and control arms, and a p-value may be obtained from the asymptotic χ^2 lookup distribution with 1 degree of freedom. Parameters may be estimated by a quasi-Newton or another optimisation algorithm, with $0\leq\theta_x\leq1$ and the likelihood set to infinity for infeasible δ. More precisely, the pseudo-algorithm that is used next to estimate power from a likelihood ratio approach is as follows.

- Set sample size n and significance level α. The expected sample size n_x in each deprivation quintile ($x=1$, …, 5) is obtained using Table 1.
- For $j=1$, …, R simulation replicates:

Table 1
Percentage in each deprivation quintile between October 2006 and January 2009, by region.

	Deprivation quintile				
	1	2	3	4	5
S	37	25	19	14	5
E	28	26	22	15	9
LDN	9	14	21	30	26
NW	11	17	17	20	35
NE	10	17	19	24	29

1. Simulate the number of events y_{jkx} in the baseline group ($k=0$) and the new initiative ($k=1$) from a binomial distribution with $n_x/2$ individuals in each, and parameter π_{kx}^* (where $\pi_{0x}^*=\pi_x$ and $\pi_{1x}^*=\pi_x^*$ above).
2. Fit the model Eq. (5) or Eq. (6) by maximising a log-likelihood

$$l_j^* \propto \sum_{x=1}^{5}\sum_{k=0}^{1} y_{jkx}log(\pi_{kx}^*) + (n_x/2-y_{jkx})log(1-\pi_{kx}^*)$$

   using a quasi-Newton algorithm. Also fit the null model to obtain l_j.
3. Determine a p-value for the deviance $2(l_j^*-l_j)$ test statistic by using a χ_1^2 asymptotic lookup distribution.
4. Record an indicator variable z_j equal to one if the p-value is less than significance level α, 0 otherwise.

- Finally, estimate the power $(1-\hat{\beta})$ as $1/R\sum_{j=1}^{R} z_j$.

3. Application

The results of using the techniques in Sections 2.2 and 2.3 are reported in this section.

3.1. Data

Between October 2006 and January 2009 approximately 2.7 millions invitations were sent and 1.4 million kits were

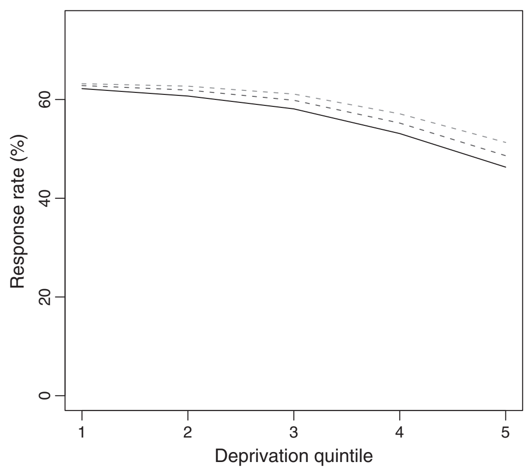

Fig. 2. Alternative hypothesis used in sample size calculation for Eastern region. The black solid line (—) is the current situation; the grey dashed (— — —) lines are the alternatives going from a 5% relative increase for the most deprived quintile, to an absolute 5% increase.

Table 2
Number of individuals (thousands) in each arm ($n/2$) needed to detect two alternatives (most deprived has 5% absolute or relative difference) using the approach in Section 2.2. Two-sided test with significance level $\alpha=5\%$, and power $(1-\beta)=90\%$. The overall number is based on different models being fitted to each region. The approach is conservative in relation to the estimated power under a likelihood-ratio (LR) test (10,000 replicates).

Region	Absolute increase		Relative increase	
	$n/2$	LR power	$n/2$	LR power
	(thousands)	(%)	(thousands)	(%)
S	15.7	99	50.1	96
E	11.8	98	38.7	94
LDN	5.1	95	35.1	90
NW	5.5	96	23.7	92
NE	5.6	95	19.6	91
Overall	43.7		167.2	

returned. The invitation location was recorded at the postcode sector level. To obtain the deprivation quintiles, each postcode sector was mapped to an estimated value of the index of multiple deprivation (2004) score, by using census data at the lower super output area level. This involved computing a weighted average of the deprivation scores associated with the output areas related to each postcode sector, taking into account that a single postcode sector could be part of more than one output area, but also that a single output area could comprise many postcode sectors. The deprivation quintiles were chosen based on the national distribution of the continuous deprivation scores. The proportion of individuals in each deprivation quintile is shown in Table 1.

3.2. Results

Two alternative hypotheses are considered, and shown in Fig. 2 for the Eastern region. In the first, the new policy is taken to produce a 1–2–3–4–5% relative increase in response rate across the deprivation quintiles (richest to poorest), *i.e.* the new policy reponse rate for the highest deprivation quintile in the current percentage multiplied by 1.05. The second alternative adds 1–2–3–4–5% to the response rate.

Some reasons for the alternative hypotheses chosen are as follows. Although it might be easier to reach higher socioeconomic groups with health messages, the interventions have been chosen as those most likely to have an effect in the lower socioeconomic groups, for which there is a clear need. Of course we do not know if either of the alternative hypotheses will be borne out (the reason for the trial), but it makes most sense to power the study to find an effect in the general direction to that which we have designed the intervention to achieve.

Table 2 shows a summary table of results for the sample size in each region. The general pattern is that the poorer regions (*c.f.* Table 1) require a smaller sample size than the richer regions because most of the hypothesized increase is in the poorer postcodes: the sample size is not balanced across deprivation quintiles. Table 2 also shows that the approach is conservative in relation to the simulated power

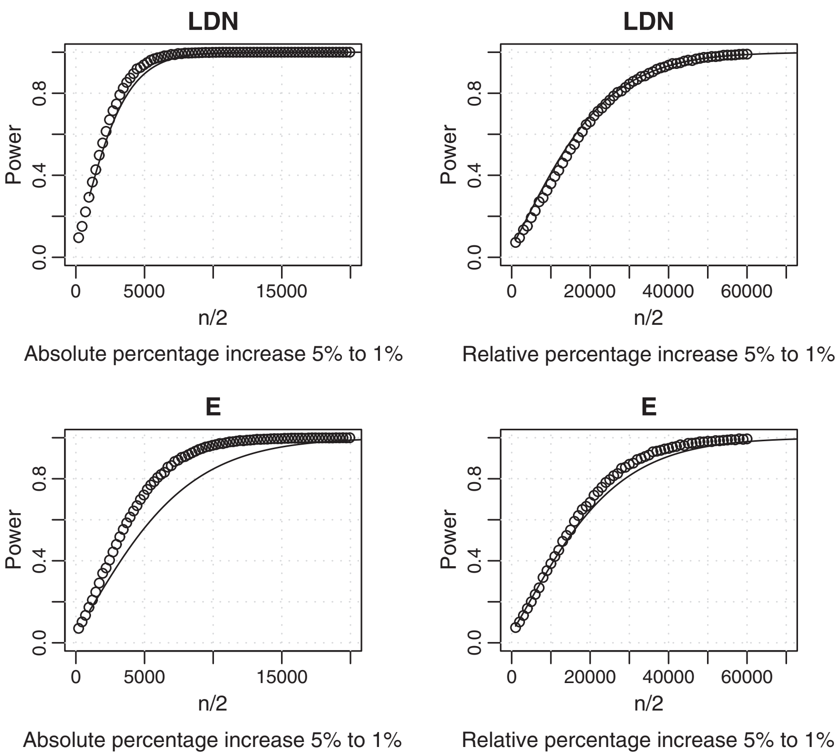

Fig. 3. Sample size for each arm in Eastern and London regions, for a relative increase in response rate of 5% of the highest deprivation quintile to 1% in the lowest deprivation quintile. The line (—) is from the approach in Section 2.2, the dots (o) are from a simulation study using likelihood-ratio tests with the correct model form from Section 2.3. 10,000 simulation repeats were used to obtain the power at each point.

Table 3
Sample size from single calculation at each quintile (thousands).

	Deprivation quintile				
	1	2	3	4	5
Absolute					
LDN	52.4	13.1	5.8	3.2	1.9
S	49.8	12.4	5.6	3.2	2.1
E	49.1	12.4	5.6	3.2	2.1
NW	48.3	12.4	5.6	3.2	2.1
NE	46.4	12.0	5.5	3.2	2.1
Relative					
LDN	197.0	59.1	31.4	20.9	17.3
S	135.1	34.0	16.9	11.8	10.1
E	127.3	33.8	16.7	11.6	9.8
NW	119.3	33.2	17.2	11.7	11.7
NE	104.9	30.0	14.9	10.3	9.3

from a likelihood ratio test, and that it is more conservative for the absolute alternative hypothesis, and smaller sample size. The power difference is shown in more detail in Fig. 3 for the London and Eastern regions. A sense check in Table 3 applies the usual two-sample sample-size calculation with significance level $\alpha = 0.05$ and power $1-\beta = 0.9$ to each deprivation quintile, following [10] p. 242 where

$$n/2 = \left(z_{\alpha/2} + z_{\beta}\right)^2 \{\pi(1-\pi) + \pi^*(1-\pi^*)\}(\pi-\pi^*)^{-2}.$$

Besides showing that the sample sizes appear to be of a reasonable magnitude in Table 2, in conjunction the unbalanced deprivation quintiles in Table 1, it helps to explain the difference between the sample size by centre.

4. Conclusion

This article has presented a method to estimate sample size for a bowel-cancer screening trial. The approach may be used when there is relationship between a response probability π and a covariate x that interacts with the experimental arm. We used the method to calculate approximate sample sizes based on existing bowel-screening uptake data. The results were investigated through comparison with repeated application of two-sample calculations, and computer-simulation estimates of power from a likelihood-ratio test. The approach provided similar sample-size calculations to the likelihood-ratio, but it was more conservative. It performed better for smaller percent differences from the null model.

Some extensions of the method might be useful in other situations. One straightforward development is for continuous x, rather than ordinal x. In this case the c-factors may be obtained at any particular x using equations such as Eq. (2), and integrals will replace summations in terms such as Eq. (3). Another extension is when the response is not binary. Here a different two-sample test could be applied, such as a z-test when a normal distribution is reasonable assumption for the response.

Acknowledgements

Part of Dr Baio's work was undertaken at UCLH/UCL who received a proportion of funding from the Department of Health's NIHR Biomedical Research Centres funding scheme. We thank Dr Christian Von Wagner for his guidance on health literacy and understanding inequalities in bowel-cancer screening.

References

[1] Von Wagner C, Good A, Whitaker KL, Wardle, J. Psychosocial Determinants of Socioeconomic Inequalities in Cancer Screening Participation: A Conceptual Framework. Epidemiol Rev Forthcoming, doi:10.1093/epirev/mxq018 First.
[2] Kutner M, Greenberg E, Jin Y, Paulsen C, White S. Literacy in everyday life: Results from the 2003 National Assessment of Adult Literacy. Washington, DC: U.S.; Department of Education; 2007.
[3] Von Wagner C, Knight K, Steptoe A, Wardle J. Functional health literacy and health promoting behaviour in a national sample of British adults. J Epidemiol Community Health 2007;61:1086–90.
[4] Von Wagner C, Semmler C, Good A, Wardle J. Health literacy and self-efficacy for participating in colorectal cancer screening: the role of information processing. Patient Educ Couns 2009;75:352–7.
[5] Smith SK, Trevena L, Nutbeam D, Barratt A, McCaffery KJ. Information needs and preferences of low and high literacy consumers for decisions about colorectal cancer screenings: utilizing a linguistic model. Health Expect 2008;11:123–36.
[6] Kreuter MW, Green M, Cappella J, Slater M, Wise M, Storey D, et al. Narrative communication in cancer prevention and control: a framework to guide research and application. Ann Behav Med 2007;33:221–35.
[7] Hsieh FY, Bloch DA, Larsen MD. A simple method of sample size calculation for linear and logistic regression. Stat Med 1998;17:1623–34.
[8] Whittemore AS. Sample size for logistic regression with small response probability. J Am Stat Assoc 1981;76:27–32.
[9] Hsieh FY. Sample size tables for logistic regression. Stat Med 1989;8: 795–802.
[10] Agresti A. Categorical Data Analysis. 2nd ed. New Jersey: Wiley-Interscience; 2002.
[11] Væth M, Skovlund E. A simple approach to power and sample size calculations in logistic regression and Cox regression models. Stat Med 2004;23:1781–92.
[12] Demidenko E. Sample size determination for logistic regression revisited. Stat Med 2007;26:3385–97.
[13] Demidenko E. Sample size and optimal design for logistic regression with binary interaction. Stat Med 2008;27:36–46.
[14] Gardiner J, Pathak D, Indurkhya A. Power calculations for detecting interaction in stratified 2×2 tables. Stat Probab Lett 1999;31:267–75.