Journal of Evaluation in Clinical Practice
International Journal of Public Health Policy and Health Services Research

Journal of Evaluation in Clinical Practice ISSN 1356-1294

Analytical decision model for sample size and effectiveness projections for use in planning a population-based randomized controlled trial of colorectal cancer screening

Sherry Y-H Chiu PhD,[1] Nea Malila PhD,[2] Amy M-F Yen PhD,[3] Ahti Anttila PhD,[4] Matti Hakama PhD[5] and H-H Chen PhD[6]

[1]Research Fellow, Tampere School of Public Health, University of Tampere, Tampere, Research Fellow, Finnish Cancer Registry and Mass Screening Registry, Helsinki, Finland and Assistant Professor, Department and Graduate Institute of Health Care Management, Chang Gung University, Tao-Yuan county, Taiwan
[2]Professor, Tampere School of Public Health, University of Tampere, Tampere and Director, Finnish Cancer Registry and Mass Screening Registry, Helsinki, Finland
[3]Research Fellow, Division of Biostatistics, Graduate Institute of Epidemiology and Research Fellow, Centre of Biostatistics Consultation, College of Public Health, National Taiwan University, Taipei, Taiwan
[4]Research Director, Finnish Cancer Registry and Mass Screening Registry, Helsinki, Finland
[5]Professor, Tampere School of Public Health, University of Tampere, Tampere and Professor, Finnish Cancer Registry and Mass Screening Registry, Helsinki, Finland
[6]Professor, Tampere School of Public Health, University of Tampere, Tampere, Finland, Professor, Division of Biostatistics, Graduate Institute of Epidemiology and Chief, Centre of Biostatistics Consultation, College of Public Health, National Taiwan University, Taipei, Taiwan

Keywords
colorectal cancer, mortality reduction, natural history, randomized controlled trial, sample size, surrogate end point

Correspondence
Professor Hsiu-Hsi Chen
Division of Biostatistics
Graduate Institute of Epidemiology
Centre of Biostatistics Consultation
College of Public Health
National Taiwan University
Room 533, No. 17, Hsuchow Road
Taipei 100
Taiwan
E-mail: chenlin@ntu.edu.tw

Funding
This research work was supported by the FiDiPro Research Project of Tampere School of Public Health Granted from Academy of Finland and also supported by the National Science Council in Taiwan (NSC 96-2628-B-002-096-Mr3).

Accepted for publication: 1 September 2009

doi:10.1111/j.1365-2753.2010.01378.x

Abstract

Rationale, aims and objectives Population-based randomized controlled trials (RCTs) often involve enormous costs and long-term follow-up to evaluate primary end points. Analytical decision-simulated model for sample size and effectiveness projections based on primary and surrogate end points are necessary before planning a population-based RCT.
Method Based on the study design similar to two previous RCTs, transition rates were estimated using a five-state natural history model [normal, preclinical detection phase (PCDP) Dukes' A/B, PCDP Dukes' C/D, Clinical Dukes' A/B and Clinical Dukes' C/D]. The Markov cycle tree was assigned transition parameters, variables related to screening and survival rate that simulated results of 10-year follow-up in the absence of screening for a hypothetical cohort aged 45–74 years. The corresponding screened arm was to simulate the results after the introduction of population-based screening for colorectal cancer with fecal occult blood test with stop screen design.
Results The natural course of mean sojourn time for five-state Markov model were estimated as 2.75 years for preclinical Dukes' A/B and 1.38 years for preclinical Dukes' C/D. The expected reductions in mortality and Dukes' C/D were 13% (95% confidence intervals: 7–19%) and 26% (95% confidence intervals: 20–32%), respectively, given a 70% acceptance rate and a 90% colonoscopy referral rate. Sample sizes required were 86 150 and 65 592 subjects for the primary end point and the surrogate end point, respectively, given an incidence rate up to 0.0020 per year.
Conclusions The sample sizes required for primary and surrogate end points and the projection of effectiveness of fecal occult blood test for colorectal cancer screening were developed. Both are very important to plan a population-based RCT.

Introduction

The effectiveness of population-based screening for colorectal cancer (CRC) with the fecal occult blood (FOB) test has been demonstrated in several randomized controlled trials (RCTs), with a 15–18% reduction in CRC mortality [1–4]. However, the absolute benefit of such population-based screening is still subject to several determinants, including the incidence rate of CRC in the underlying population, the attendance rate of the invited population, compliance with any confirmatory procedures, the disease's

natural history in terms of progression of colorectal neoplasms, and the survival rate of screen-detected and clinically detected CRC. Additionally, the implementation of a population-based randomized trial may require tremendous costs and long-term follow-up to achieve sufficient primary end point (mortality) cases.

At the inception of study design, the determination of a sample size large enough to achieve sufficient statistical power is essential, given an expected mortality reduction; the latter, in turn is a function of attendance rate, compliance with colonoscopy, the progression of colorectal neoplasms and the improvement in survival as a result of screening. Of these variables, parameters related to the disease's natural history in regard to the progression of colorectal neoplasms play an important role in the prediction of the effectiveness of CRC screening. Examining the disease's natural history also enables the prediction of the effectiveness of CRC screening based on surrogate end points, such as Dukes' stages, that can substantially reduce the enormous time and costs involved in long-term follow-up. The parameters of the disease's natural history, in conjunction with basic elements of screening (such as attendance rate and compliance rate for referral) and prognosis of screen-detected and clinically detected cases, provide the ability to estimate the required sample size or to perform a power calculation, given a capacity for enrolling participants, and to predict the likely effectiveness of a screening programme before implementation of any RCT.

To integrate these parameters into an analytical framework for estimating the required sample size and projecting the mortality reduction, given a proposed population-based screening for CRC with the FOB test, requires the development of a complex analytical decision model. The aim of this study was to illustrate how to develop a Markov decision model by gleaning information, particularly the parameters governing the progression of CRC disease, from the existing literature and to combine these elements to project effectiveness, such as mortality reduction, if a population-based RCT for CRC screening with the FOB test is to be conducted. Specifically, we first estimated the parameters related to the disease's natural history, calling upon two existing RCTs implemented in Europe that used a five-state Markov model, based on the classification of Dukes' stages [1,2]. An analytical decision model was developed, based on a framework of the Markov process, screening procedure, surveillance for colorectal adenomas, and the prognosis of screen-detected and clinically detected CRC. A computer-simulation approach was applied to calculating the number of predicted deaths from CRC for the screened group and the control group. This involved simulating a hypothetical cohort, aged 45–74 years, similar to the target populations in previous population-based RCTs. Sample size determination and the effectiveness in reducing mortality were projected, based on a biennial FOB test screening programme.

Materials and methods

Study design for colorectal cancer screening

Our proposed method can be adapted to any study design for population-based CRC screening. In the current study, we adopted a stop-screen design [5,6] to simulate a hypothetical cohort, aged 45–74 years, similar to those in two established population-based RCTs [1,2], both of which have already long-term follow-up results, and one ongoing RCT in Finland with the preliminary results published elsewhere [7,8]. This cohort was randomly allocated to the two arms, biennial screening with the FOB test and the control group. The study period was 6 years, with four rounds of screening in the study arm; members of the control group were invited for screening over the same period, but without offering the FOB test. We followed CRC deaths in the two groups over time until 10 years after randomization. Figure 1 shows a flow chart of the study design.

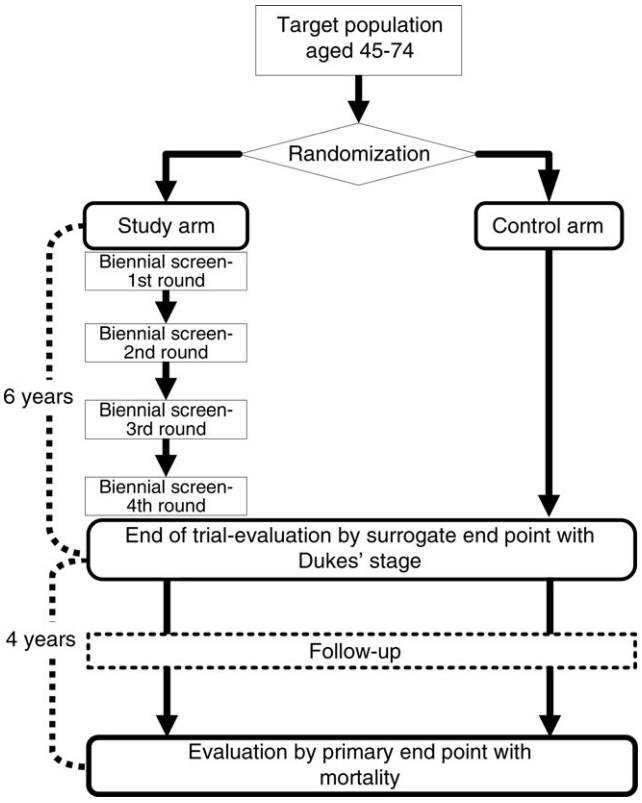

Figure 1 Flow chart for colorectal cancer screening with a stop-screen design.

Five-state model for the disease's natural history by Dukes' stages

To simulate the results of the control group, who were not given the FOB test, we estimated the transition parameters responsible for the natural course of CRC progression. We built a five-state model to delineate the natural history of CRC by Dukes' stage: free of cancer (normal), the preclinical detection phase (PCDP) Dukes' A + B (localized CRC), the PCDP Dukes' C + D (non-localized CRC), the clinical Dukes' A + B phase and the clinical Dukes' C + D phase. The progression of CRC by Dukes' stages is illustrated in Fig. 2. We used findings from both the Nottingham [1] and Denmark [2] RCTs for CRC screening with the FOB test to estimate the parameters in relation to the natural history of CRC. Information on each arm for CRC cases, the distribution of Dukes' stages, interval cancers, screening rounds, number of attendees

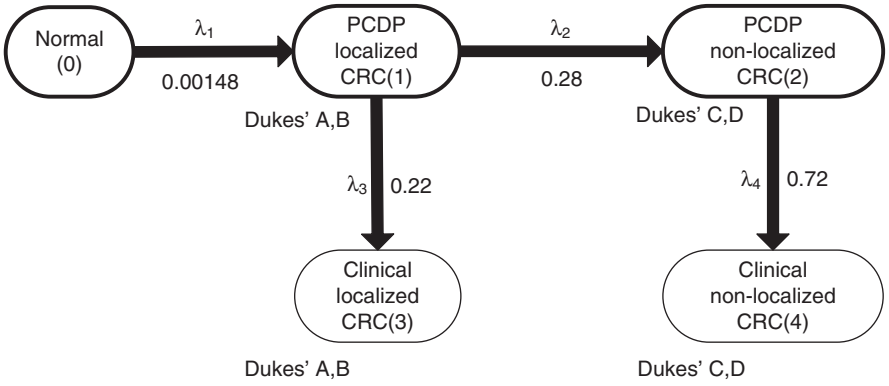

Figure 2 The five-stage natural history with transition rate of colorectal cancer. λ_1: Normal → PCDP Dukes' stage A or B. λ_2: PCDP Dukes' stage A or B → PCDP Dukes' stage C or D. λ_3: PCDP Dukes' stage A or B → Clinical Dukes' stage A or B. λ_4: PCDP Dukes' stage C or D → Clinical Dukes' stage C or D. PCDP, preclinical detection phase.

Table 1 Parameter estimation and control validation

		Nottingham, UK		Funen, Denmark	
Items	Mode	Observed	Expected	Observed	Expected
Prevalent screening	Normal	46 622	46 617.94	20 635	20 625.85
	PCDP A + B	72	79.35	26	30.71
	PCDP C + D	32	28.71	11	15.43
Re-screening	Normal	88 789	88 721.52	66 025	65 964.82
	PCDP A + B	99	100.74	62	68.88
	PCDP C + D	33	21.48	21	17.52
Interval cancer	Clinical A + B	116	123.9	82	73.28
	Clinical C + D	133	116.77	66	60.46
Non-responder	Normal	30 015	30 039.82	9895	9931.99
	Clinical A + B	187	163.99	93	73.08
	Clinical C + D	213	211.2	102	84.93
External validation	Mode	Observed	Expected	Observed	Expected
(control group)	Normal	74 142	73 847.98	30 483	30 361.10
	Dukes' stage A + B	395	371.14	245	224.52
	Dukes' stage C + D	461	478.09	238	260.90
Goodness-of-fit		$\chi^2_{(2)} = 3.32$		$\chi^2_{(2)} = 4.37$	
		P value = 0.1905		P value = 0.1126	

PCDP, preclinical detection phase.

and follow-up person years were obtained from the literature [1–3,9]. Table 1 shows descriptive data from the two reported RCTs.

Other parameters

In addition to parameters pertaining to the natural history of the disease, Table 2 shows other parameters used for the Markov decision analysis, including attendance rate for the FOB test, compliance rate with colonoscopy, sensitivity and specificity of the FOB test, survival rate by Dukes' stage, and proportions of Dukes' stages, in the light of the two RCTs. Because the sensitivity reported from the RCT programme was a programme sensitivity, whereas we needed a test sensitivity to assign to the Markov cycle tree, we converted the 51% programme sensitivity into a 64% test sensitivity using the Day method [10], based on a mean sojourn time (MST) equivalent to around 3 years and a 2-year interscreening interval. Although the survival rate by Dukes' stage can be extrapolated by using literatures, 10-year survival rate by Dukes' stage was directly computed using CRC cases between 1990 and 2003, 14 years before screening, from Finnish cancer registry [11] as individual data are available and our study design is also similar to the Finnish RCT mentioned above. Note that although our study design is also similar to the Finnish randomized trial we did not use data from this trial to compute the disease natural history because the project has been still going but not finished yet. However, as the average age and annual incidence rates in both randomized trials were identical to the Finnish RCT, we believe the application of 10-year survival from Finnish cancer registry is reasonable.

Statistical analyses

Parameter estimation for the natural history of colorectal cancer

Transition rates related to the five-state model by Dukes' stage (Fig. 2) were estimated by the likelihood function in light of the transition probabilities derived from the forward Kolmogorov equation, as developed in previous studies [12], and applied to the high-risk CRC group in the Wong *et al.* study [13]. A meta-analysis was also performed to obtain pooled estimates by

Table 2 The transition probability for five-state Markov model of colorectal cancer

Transition	Base-case value Annual transition probability	Resource
Normal → Normal	0.998981	Meta-analysis from two RCTs [2,3] and generated from transition rates to probabilities
Normal → PCDP Dukes' A or B	0.000800	
Normal → PCDP Dukes' C or D	0.000097	
Normal → Clinical A or B	0.000095	
Normal → Clinical C or D	0.000025	
Normal → CRC death	0.000002	
PCDP Dukes' A or B → PCDP Dukes' A or B	0.603301	
PCDP Dukes' A or B → PCDP Dukes' C or D	0.153101	
PCDP Dukes' A or B → Clinical Dukes' A or B	0.171587	
PCDP Dukes' A or B → Clinical Dukes' C or D	0.064767	
PCDP Dukes' A or B → CRC death	0.007244	
PCDP Dukes' C or D → PCDP Dukes' C or D	0.484339	
PCDP Dukes' C or D → Clinical Dukes' C or D	0.470499	
PCDP Dukes' C or D → CRC death	0.045162	
Screening tool: FOB test		
Sensitivity	63%	Derived from Day method [10]
Specificity	97%	
Screening characteristics		
FOB test acceptance rate	70% (50–100%)	Sensitivity
Colonoscopy compliance rate	90% (80–100%)	Analysis
Survival rate		
Dukes' A or B	0.76	Finnish Cancer Registry
Dukes' C or D	0.22	

CRC, colorectal cancer; FOB, fecal occult blood; RCT, randomized controlled trial; PCDP, preclinical detection phase.

weighting each estimate with the inverse of the variance. These transition rates between states, the MSTs between stages, and the average dwell time in a specific stage were estimated in light of the likelihood function [13]. Parameter estimation was based on non-linear methods.

Model validation

To assess the accuracy of the parameters estimated from data from the two RCTs, we applied parameters estimated from the study arm in those trials and calculated the expected values using data from the control arms. The comparison between the expected and the observed parameters was further assessed using the Pearson chi-squared test.

Markov decision model

The decision in our study consisted of two arms, the biennial screening and the control. Each arm followed the infrastructure of the Markov decision model detailed which are given with available Supporting Information at website (http://homepage.ntu.edu.tw/~ntucbsc/tony_e.htm). The decision analysis scenario emulated a stop-screen design, as mentioned above. The study duration was approximately 6 years, and the follow-up period lasted until 10 years after randomization. To address the uncertainty in some parameters, sensitivity analysis was performed for influential parameters, including the FOB test acceptance rate (50%, 60%, 70%, 80%, 90%, 100%) and the colonoscopy compliance rate (80%, 90%, 100%). Effectiveness in the reduction of both advanced cancer and mortality was projected by a series of combinations of the attendance and compliance rates.

Sample size and power calculations

Based on the power calculation of Chen *et al*. [14], the sample sizes required for the primary (mortality) end point and the surrogate end point (Dukes' stage transition) were computed according to a stop-screen design. To demonstrate the flexibility of our proposed method, we calculated the sample sizes required for primary end point (mortality) and the surrogate end point (Dukes' stage transition) for three scenarios: low (1 per 1000 years), intermediate (1.5 per 1000 years) and high (2 per 1000 years) incidences of PCDP cancer.

Results

Parameter estimation for natural history of colorectal cancer

The RCTs of CRC screening, the Nottingham and Funen studies, involved 76 466 versus 76 384 and 30 967 versus 30 966 patients, respectively, for the study and control arms. Annual transition rates (per year) for the five-state natural history model, normal → PCDP Dukes' stage A or B, PCDP Dukes' stage A or B → PCDP Dukes'

stage C or D, PCDP Dukes' stage A or B → clinical Dukes' stage A or B, and PCDP Dukes' stage C or D → clinical Dukes' stage C or D, were 0.00146, 0.27540, 0.21420 and 0.76270, respectively, for the Nottingham study, and 0.00158, 0.32470, 0.28010 and 0.64780, respectively, for the Funen study. The results of the meta-analysis using the inverse-variance method are listed in Fig. 2. The estimated annual incidence of entering PCDP Dukes' was 1.48 per 1000. The annual transition rate (in person years) was 0.28 for progression from PCDP Dukes' A + B to PCDP Dukes' C + D and 0.22 for the progression from PCDP Dukes' A + B to clinical-phase Dukes' A + B. The annual transition was up to 0.72 for progression from PCDP Dukes' C + D to clinical Dukes' C + D. In terms of the transition rates between any two states, the average dwell time of staying in PCDP, as determined by meta-analysis, was 2.75 years and 1.38 years for PCDP Dukes' A or B and PCDP Dukes' C or D, respectively. The annual transition probabilities from each state, converted from the annual transition rates, are also listed in Table 2. Regarding the validation of parameters using the control groups, as there is lacking of statistical significance for Nottingham ($\chi^2_{(2)} = 3.32$, $P = 0.19$) and Funen trials ($\chi^2_{(2)} = 4.37$, $P = 0.11$) this suggest an adequate fit of model.

Base-case parameters used for the simulation are listed in Table 2. The annual transition probabilities listed in Table 2 were assigned to the Markov cycle tree transition probabilities. The 10-year survival rates by Dukes' stage from the literature were 76.40% and 22.42% for Dukes' stages A and B and Dukes' stages C and D, respectively.

Predicted effectiveness, sample size and power calculations

Table 3 shows results in regard to the effectiveness as measured by mortality (primary end point) reduction and reduction in Dukes' stage C + D (surrogate end point), by a series of combinations of attendance rates and the rates of compliance with colonoscopy. In the case of a 70% FOB test acceptance rate and a 90% compliance rate with colonoscopy, the mortality reduction was predicted to be 13% (relative rate = 0.87, 95% confidence intervals: 0.81–0.93). The expected reduction in Dukes' C + D was 26% (relative rate = 0.74, 95% confidence intervals: 068–0.80).

Table 4 shows sample size determinations by primary and surrogate end points. In the case of a 70% FOB test acceptance rate and 90% compliance rate with colonoscopy, 172 685, 114 764 and 86 150 subjects are required for the low, medium and high incidences using the primary end point (mortality), whereas 131 524, 87 402 and 65 592 subjects are required for low, medium and high incidences using the surrogate end point, given 80% statistical power and a 5% level of type I error.

The sample size required for the surrogate end point is about three-quarters of that required for the primary end point (mortality), regardless of the FOB test acceptance and colonoscopy compliance rates.

Discussion

In this study, we developed a useful and feasible method to predict the effectiveness of screening, sample size determinations and power calculations, using mortality as a primary end point and Dukes' stage progression as a surrogate end point, before a large-scale population-based RCT. The proposed five-state natural history model and Markov decision model are illustrated, using a stop-screen design with a hypothetical cohort, aged 45–74 years, similar to those seen in two actual RCTs [1,2]. The results show a 13% mortality reduction and a 26% reduction in Dukes' C + D at 10 years after randomization, given a 70%

Table 3 Predictive effectiveness by primary mortality and surrogate end point

FOB test Attendance rate (%)	Colonoscopy Compliance rate (%)	Mortality end point RR*	Surrogate end point RR†
50	80	0.910	0.830
50	90	0.900	0.811
50	100	0.890	0.792
60	80	0.894	0.800
60	90	0.882	0.777
60	100	0.870	0.755
70	80	0.878	0.770
70	90	0.865	0.744
70	100	0.852	0.719
80	80	0.863	0.741
80	90	0.848	0.712
80	100	0.834	0.685
90	80	0.848	0.712
90	90	0.832	0.682
90	100	0.817	0.652
100	80	0.834	0.685
100	90	0.817	0.652
100	100	0.800	0.620

*Relative rate of death from CRC.
†Relative rate of advanced Dukes' stage C or D (non-localized).
Low, 0.0010; medium, 0.0015; high, 0.0020 CRC incidence rates.
CRC, colorectal cancer; FOB, fecal occult blood.

Table 4 Sample size calculation by primary mortality and surrogate end point

FOB test Attendance rate (%)	Colonoscopy Compliance rate (%)	Mortality end point			Surrogate end point		
		Low	Medium	High	Low	Medium	High
50	80	409 991	272 491	204 500	309 605	205 750	154 368
50	90	327 084	217 534	163 059	247 496	164 587	123 336
50	100	267 558	177 720	133 373	202 844	134 724	101 077
60	80	289 249	192 302	144 228	219 121	145 666	109 219
60	90	231 083	153 564	115 230	175 454	116 584	87 457
60	100	189 167	125 794	94 432	143 933	95 704	71 825
70	80	215 550	143 444	107 604	163 777	108 980	81 728
70	90	172 685	114 764	86 150	131 524	87 402	65 592
70	100	141 383	94 266	70 717	107 925	71 951	53 963
80	80	167 677	111 430	83 598	127 752	84 891	63 670
80	90	134 386	89 424	67 069	102 646	68 298	51 211
80	100	110 537	73 542	55 168	84 623	56 296	42 221
90	80	134 386	89 424	67 069	102 646	68 298	51 211
90	90	108 046	71 863	53 909	82 739	55 026	41 268
90	100	89 021	59 222	44 424	68 324	45 450	34 085
100	80	110 537	73 542	55 168	84 623	56 296	42 221
100	90	89 021	59 222	44 424	68 324	45 450	34 085
100	100	73 516	48 894	36 691	56 547	37 605	28 213

Sample size required given power $(1 - \beta) = 80$ and α level = 5.
Low, 0.0010; medium, 0.0015; high, 0.0020 CRC incidence rates.
CRC, colorectal cancer; FOB, fecal occult blood.

attendance rate and a 90% compliance rate with colonoscopy. Given these expected effectiveness estimates and the high incidence, the required sample sizes are around 86 150 using the primary end point and 65 592 using the surrogate end point. Both results are informative for any location where such a population-based RCT applied to evaluating the effectiveness of any new or modified screening modality is planned. Information provided in this way may also be helpful for assessing the effects of the attendance rate for the FOB test and compliance rate with confirmatory procedure such as colonoscopy. It can be seen that the expected reductions in mortality and surrogate end point decline to approximately 9% if the attendance rate is lowered to 50% and the compliance rate to 80%. On the other hand, the expected reductions in mortality and surrogate end point are enhanced, to 20% and 38%, respectively, if both the attendance and compliance rates are 100%.

Our study results on sample size determination with a range from the low to high incidences are also informative for policymakers contemplating designing such a population-based RCT. The cost incurred in planning any RCT can be estimated by projecting the sample sizes. It can be seen that the use of the surrogate end point can reduce sample sizes by about 25%, compared with the primary end point.

We can apply our model and parameters to predicting the results of the Finnish first-round RCT [8] that has been undertaken and still underway. Given the FOB test acceptance rate and colonoscopy compliance rates obtained from the preliminary results of the first round, 70% and 90%, respectively [8], the sample size required is around 172 000 in each arm if the primary end point of mortality is used. This is reassuring, given the originally planned sample of 250 000 in each arm; indeed, the study therefore should have sufficient statistical power (up to 90.7%) and will be even more powerful if the surrogate end point is used (power = 96.3%).

Credibility of parameters

In our study, the built-in five-state disease natural history model of CRC plays a key role in the subsequent Markov decision analysis. The credibility of our estimated results is subject to the adequacy of the model. Annual transition rates related to the disease natural history were estimated using data from two randomized trials in Europe. Transition parameters in the Nottingham trial were not substantially different from those of the Funen trial, although the sample size in the former was considerably larger. External validation using the control arm suggested an adequate fit of the five-state model. Using the weighted estimates from both studies, our model predicted 59.2% and 40.8% of Dukes' A + B and Dukes' C + D, respectively, which are close to the corresponding figures of 53.4% and 46.6% in the Nottingham trial and of 58.8% and 41.1% in the Funen trial. These findings add credibility to our disease natural history model. Regarding the external validity of our parameters, by adjusting the annual preclinical incidence rate to 0.00102, similar to the clinical incidence rate of CRC in Finland, we applied other transition parameters in Fig. 2 to predict the occurrence of CRC by Dukes' stage in the year 2003, just before the implementation of nationwide screening for CRC with the FOB test. The number of observed Dukes' A + B ($n = 234$) and C + D ($n = 282$) [11] cases based on population 500 000 aged 60–69 is close to the corresponding expected number of Dukes' A + B ($n = 227$) and Dukes C + D ($n = 289$). Indeed, these numbers are not statistically significantly different ($\chi^2_{(2)} = 4.36$, $P = 0.11$). This again points to the external validity of the transition rates after the onset of disease.

Compared with sample size calculation for population-based screening for CRC with FOB test such as Nottingham trial [15] that is only tailored for evaluation of primary end point, that is, mortality, our proposed method for sample size determination and power calculation not only considers primary end point but also surrogate end point. In addition, factors affecting the process are also explicitly modelled by using analytical decision Markov model. These include attendance rate of FOB test, preclinical incidence rate of CRC, natural history of Dukes', compliance with colonoscopy and survival rate. Our model is therefore flexible for investigating the influential parameters with a series of sensitivity analysis as seen in a range of preclinical incidence rate in our study.

In conclusion, our study not only demonstrated a feasible method of predicting the effectiveness of a population-based RCT, but also showed the sample sizes required with different combinations of FOB test acceptance rate and colonoscopy compliance rate. The built-in disease natural history model, in conjunction with a Markov decision analysis, are useful for health policy makers to estimate the required sample sizes while planning a new trial and to project the effectiveness of the planned RCT in the interim period and at the end of the trial.

References

1. Hardcastle, J. D., Chamberlain, J. O., Robinson, M. H., Moss, S., Amar, S. S., Balfour, T. W., James, P. D. & Mangham, C. M. (1996) Randomised controlled trial of faecal-occult-blood screening for colorectal cancer. *Lancet*, 348, 1427–1477.
2. Kronborg, O., Fenger, C., Olsen, J., Jorgensen, O. D. & Sondergaard, O. (1996) Randomised study of screening for colorectal cancer with faecal-occult-blood test. *Lancet*, 348, 1467–1471.
3. Jorgensen, O. D., Kronborg, O. & Fenger, C. (2002) A randomized study of screening for colorectal cancer using faecal occult blood testing: results after 13 years and seven biennial screening rounds. *Gut*, 50, 29–32.
4. Scholefield, J. H., Moss, S., Sufi, F., Mangham, C. M. & Hardcastle, J. D. (2002) Effect of faecal occult blood screening on mortality from colorectal cancer: results from a randomized controlled trial. *Gut*, 50, 840–844.
5. Breslow, N. E. & Day, N. E. (1980) Statistical Methods in Cancer Research. Volume II: The Design and Analysis of Cohort Studies. Lyon: International Agency for Research on Cancer.
6. Etzioni, R. D., Connor, R. J., Prorok, P. C. & Self, S. G. (1995) Design and analysis of cancer screening trials. *Statistical Methods in Medical Research*, 4, 3–17.
7. Malila, N., Anttila, A. & Hakama, M. (2005) Colorectal cancer screening in Finland: details of the national screening programme implemented in autumn 2004. *Journal of Medical Screening*, 12, 28–32.
8. Malila, N., Oivanen, T., Malminiemi, O. & Hakama, M. (2008) Test, episode, and programme sensitivities of screening for colorectal cancer as a public health policy in Finland: experimental design. *British Medical Journal*, 337, a2261.
9. Whynes, D. K. on behalf of the Nottingham FOB Screening Trial (2004) Cost-effectiveness of screening for colorectal cancer: evidence from the Nottingham faecal occult blood trial. *Journal of Medical Screening*, 11, 11–15.
10. Day, N. E. (1985) Estimating the sensitivity of a screening test. *Journal of Epidemiology and Community Health*, 39, 364–366.
11. Finnish Cancer Registry (2005) *Institute for statistical and epidemiological cancer research*. Available at: http://www.cancerregistry.fi (last accessed September 2007).
12. Chen, T. H. H., Kuo, H. S., Yen, M. F., Lai, M. S., Tabar, L. & Duffy, S. W. (2000) Estimation of sojourn time in chronic disease screening without data on interval cases. *Biometrics*, 56, 167–172.
13. Wong, J. M., Yen, M. F., Lai, M. S., Duffy, S. W., Smith, R. A. & Chen, T. H. H. (2004) Progression rates of colorectal cancer by Dukes' stage in a high-risk group: analysis of selective colorectal cancer screening. *Cancer Journal*, 10, 160–169.
14. Chen, H. H., Prevost, T. C. & Duffy, S. W. (1999) Evaluation of screening for nasopharyngeal carcinoma: trial design using Markov chain models. *British Journal of Cancer*, 79, 1894–1900.
15. Moss, S., Draper, G. J., Hardcastle, J. D. & Chamberlain, J. (1987) Calculation of sample size in trials of screening for early diagnosis of disease. *International Journal of Epidemiology*, 16, 104–110.

Appendix: Markov decision model

To simulate the scenario of the FOB test screening project with a stop-screen design (Fig. 1), a Markov decision model underpinned by a five-state disease natural history model was constructed. In the control arm, the decision tree structure follows the five-state disease natural history (Fig. 2): normal, PCDP Dukes' stage A + B, PCDP Dukes' stage C + D, clinical Dukes' stage A + B and clinical Dukes' stage C + D. Each state has a corresponding prognostic effect on death from CRC and death from a competing cause. In the screening arm or the control arm, offered a one-shot screen at the close of the trial, the 'treated PCDP Dukes' A + B' and 'treated Dukes' C + D' are added into the Markov cycle as a result of early detection by the screen.

Supporting information

Additional Supporting Information may be found in the online version of this article:

Figure S1 Markov cycle tree for colorectal cancer screening.
Figure S2 Normal.
Figure S3 PCDP Dukes' A, B.
Figure S4 PCDP Dukes' C, D.
Figure S5 Clinical Dukes' A, B.
Figure S6 Clinical Dukes' C, D.
Figure S7 Treated PCDP Dukes' A, B.
Figure S8 Treated PCDP Dukes' C, D.
Figure S9 Surveillance for colonoscopy negative.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.