Bayesian Methods for Pilot Studies

Andrew R Willan

Ontario Child Health Support Unit

Sickkids Research Institute, Toronto Ontario

Dalla Lana School of Public Health

University of Toronto, Toronto, Ontario, Canada

Lehana Thabane

Department of Health Research Methods, Evidence, and Impact

McMaster University, Hamilton Ontario

Biostatistics Unit

St Joseph's Healthcare, Hamilton, Ontario, Canada

Corresponding Author:

Andrew R Willan

555 University Avenue, PGCRL-11.9708, Toronto, ON M5G 1X8, Canada

Email: andy@andywillan.com

Abstract

Background/Aims

The use of pilot studies to help inform the design of randomized controlled trials (*RCT*s) has increased significantly over the last couple of decades. A pilot study can provide estimates of feasibility parameters, such as the recruitment, compliance and follow-up probabilities. The use of frequentists confidence intervals of these estimates fails to provide a meaningful measure of the uncertainty as it pertains to the design of the associated *RCT*. The objective of this paper is to introduce Bayesian methods for the analysis of pilot studies for determining the feasibility of an associated *RCT*.

Methods

An example from the literature is used to illustrate the advantages of a Bayesian approach for accounting for the uncertainty in pilot study results when assessing the feasibility of an associated *RCT*. Vague Beta distributed priors for the feasibility parameters are used. Based on the results from a feasibility study, simulation methods are used to determine the expected power of specified recruitment strategies for an associated *RCT*.

Results

The vague priors used for the feasibility parameters are demonstrated to be considerably robust. Beta distributed posteriors for the feasibility parameters lead to Beta-binomial predictive distributions for an associated *RCT* regarding the number of patients randomized, the number of patients that are compliant and the number of patients that complete follow-up. Ignoring the uncertainty in pilot study results can lead to inadequate power for an associated *RCT*.

Conclusions

Applying Bayesian methods to pilot studies results provides direct inference about the feasibility parameters and quantifies the uncertainty regarding the feasibility of an associated *RCT* in an intuitive and meaningful way. Furthermore, Bayesian methods can identify recruitment strategies that yield the desired power for an associated *RCT*.

Key words

Pilot studies, Bayesian methods, Recruitment feasibility, Compliance feasibility, Follow-up feasibility

Background

The use of pilot studies to inform the design of randomized controlled trials (*RCT*s) has grown substantially over the past couple of decades. In 1991 there were less than 3000 publications indexed in PubMed with the term "feasibility" or "pilot" in the title; in 2018 there were over 26,000. An extension to the CONSORT statement for randomized pilot and feasibility trials was published in 2016[1], and a new journal, entitled *Pilot and Feasibility Studies*, began publishing in 2015. In spite of this growing interest, few, if any, innovative methods for assessing the evidence provided by a pilot studies have been developed, and an examination of publications for the first four years of *Pilot and Feasibility Studies* reveal none. It is not clear how frequentists point estimates and associated confidence intervals from pilot studies inform the design of the associated *RCT*s. In this paper a Bayesian approach is proposed that uses the evidence from pilot studies to make direct inference regarding the design of associated *RCT*s in an intuitive and meaningful manner.

A pilot randomized controlled trial (*pRCT*), whether internal or external, can be employed to examine the feasibility of performing the associated "main" trial (*mRCT*)[2]. The following performance parameters, among others, might be of interest: (*i*) the probability that an eligible patient consents and is randomized (randomization probability, denoted P_r), (*ii*) the probability that a randomized patient receives the assigned intervention (compliance probability, denoted P_c), and (*iii*) the probability that a randomized patient is evaluated for the primary outcome (evaluated probability, denoted P_e). The reports of *pRCT*s often provide the observed proportions for these feasibility performance parameters which are sometimes accompanied by associated

confidence intervals.[1,3,4] Since proportions are point estimates, they don't convey the associated uncertainty, and although the confidence intervals provide an expression of the uncertainty in the frequentist framework, they are difficult to interpret and usually include values that imply the associated *mRCT* is not feasible.

A working example is provided in the next section, followed by the introduction of a Bayesian model as applied to *pRCT*s. The model is then illustrated using the working example, followed by an examination of the feasibility of the associated *mRCT* with regard to randomization and follow-up.

Methods

Working Example

As an example, consider the report of the *pRCT* by Morone *et al*.[5] The authors report that of the 77 eligible patients approached, 37 (48.1%) were recruited and randomized. A specific goal was not given. They further report that 30 of the 37 randomized patients (81.1%) completed the eight-week follow-up period and were evaluated for the primary outcome. Again, no specific goal was given. In addition, there were no measures of the uncertainty associated with the observed proportions provided. Using Wilson's score methods,[6] the 95% confidence interval for P_r, based on the data from this *pRCT*, is 37.3% to 59.0%. A similar confidence interval for P_e is 65.8% to 90.5%. The interpretation of confidence interval for P_r is as follows: if the *pRCT* was repeated in an attempt to recruit and randomize 77 eligible patients many, many, many times

then in the limit, the proportion of such confidence intervals from the repetitions that include P_r is 95%. The interpretation of the confidence interval for the associated *mRCT* is unclear. In this paper we argue that when using the data from a *pRCT*, a Bayesian approach can make direct inference regarding the feasibility parameters for the associated *mRCT* and can quantify the uncertainty in a simple, intuitive and relevant manor.

The Bayesian Approach

A Bayesian approach provides direct inference because it allows the uncertainty regarding a parameter of interest, such as P_r, to be represented by a probability distribution by treating the parameter as a random variable. Typically, a beta distribution is used if the parameter of interest is a probability. The beta distribution is defined on the interval [0,1] and is parameterized by two positive shape parameters. More importantly, the beta distribution is conjugate with binomial sampling, typically used to investigate probabilities. Meaning that, if the prior distribution for a probability is beta then following binomial sampling, the posterior distribution is beta. Specifically, if the prior distribution for the probability of a success is Beta(a, b), and if r success are observed in n attempts, then the posterior distribution is Beta(α, β), where $\alpha = a + r$, and $\beta = b + n - r$. The posterior distribution can then be used to characterize the uncertainty when making direct inference on the probability of success with statements such as the probability that the probability of success is greater or equal to x is y. Furthermore, a predictive distribution, in this case a beta-binomial, can be used to determine the probability of observing a given number of successes in a future trial of a fixed number of attempts.

Using the example from Morone *et al*.[5] the posterior distribution for P_r is Beta(a + 37, b + 40) , where the prior distribution is Beta(a, b). If we assume a vague flat distribution for the prior (*i.e.* Beta(1, 1)) then the posterior distribution is Beta(38, 41). For the sake of argument, assume that the investigators believed that the associated *mRCT* is feasible if $P_r \geq 0.5$. Based on the evidence from the *pRCT* and assuming a flat prior, the probability that $P_r \geq 0.5$ is 36.7%, and the probability that it is greater than or equal to 0.4 is 92.6%.

If $X \sim$ Beta(1, 1) then Prob($\alpha \leq X \leq \beta$) = $\beta - \alpha$. As a consequence, using the prior Beta(1, 1) implies a 20% prior belief that P_r is greater than 0.8, a value those involved in patient recruitment might consider a little optimistic. On-the-other-hand, it implies a 20% prior belief that P_r is less than 0.2, perhaps overly pessimistic. As an alternative, investigators might consider other sufficiently vague priors with less "fat" tails. Possibilities are Beta(a, a), where a is relatively small, say a = 4. An assumed prior of Beta(4, 4) has mean of 0.5, as does Beta(1, 1), and probability of being greater than 0.8 (or less than 0.2) equal to 3.3%, numbers that are perhaps more in-line with prior expectations. The posterior for P_r for our example then becomes Beta(41, 44). The probability that $P_r \geq 0.5$ is now 37.2%, and the probability that it is ≥ 0.4 is 93.7%. By comparing these values to those for the flat Beta(1, 1) prior illustrates, at least in this example, that the posterior is fairly robust against choices for sufficiently vague priors. Further illustration of the robustness can be seen in Figure 1 in which the probability that $P_r \geq 0.4$, 0.45 and 0.5 are given for various priors defined by its expected value and the sum of $a + b$. The probability that $P_r \geq 0.4$ ranges from 92.0%, (mean = 0.4, $a + b$ = 1) to 94.0% (mean = 0.5, $a + b$ = 10), demonstrating a reasonable level of robustness. The insensitivity of the posterior

distribution might be expected given that the sample size of the pilot is 77 and the effective sample size of the prior distribution ranges from 2 to 10.

Examining the Randomization Feasibility of the Main RCT

The real question is: what does the posterior distribution for P_r from the *pRCT* tell the investigators about the feasibility of randomizing the required number of patients in the associated *mRCT*? Assuming the uncertainty regarding P_r, based on the data from the *pRCT*, follows a beta distribution then the number of patients randomized (k) in the *mRCT*, given a fixed number of eligible patients (m), follows a beta-binomial distribution. That is, if the posterior from the *pRCT* for the randomization probability is Beta(α, β) then $k \sim$ Beta-binomial(m, α, β). This is the predictive distribution of the observed number of patients randomized in the future *mRCT*. The implication of this is illustrated in Figure 2. Continuing with the Morone *et al*.[5] example and using the same prior distributions given in Figure 1, the number of eligible patients that must be approached to have a 90% probability of randomizing 100, 500 or 1000 patients is provided. For example, suppose the investigators wanted to randomize a total of 500 patients in the *mRCT*. Choosing the flat prior Beta(1, 1), they would need to approach 1228 eligible patients to have a 90% probability of achieving their goal. The robustness of the choice of priors is further illustrated in Figure 2; with Beta(4, 6) (*i.e.* mean = 0.4, $a + b = 10$) as the prior, they would need to approach 1246. To have a 95% (99%) probability of achieving the recruitment goal of 500, the respective numbers are 1293 (1431) for the prior Beta(1, 1) and 1311 (1455) for the prior Beta(4, 6).

In their *pRCT* Morone *et al*.[5] managed to randomize 48.1% of eligible patients. This could lead them to believe that for the associated *mRCT* they need only approach 1040 (*i.e*. 500/0.481) eligible patient to meet a randomization goal of 500 (herein after referred to the naïve method). Such thinking, however, ignores the uncertainty in the estimate of P_r from the *pRCT* and the uncertainty regarding the observed proportion of eligible patients randomized in the *mRCT*, given P_r. If the investigators merely approached 1040 eligible patients, the probability of achieving their goal is only 50.4%, assuming the prior Beta(1, 1), and only a 43.4%, assuming the prior Beta(4, 6). This may explain why many trials fail to meet their recruitment goals. From the analysis above, to be almost certain (*i.e*. 99%) to meet the goal of 500, over 1400 eligible patient need to be approached.

Examining the Follow-up Feasibility of the Main RCT

A similar Bayesian approach can be taken to assess the feasibility with respect to other performance measures, such as P_c and P_e. Recall that Morone *et al*.[5] reported that of the 37 patients randomized, 30 (81.1%) were evaluated for the primary outcome. Choosing the flat Beta(1, 1) prior would be too pessimistic, since it assumes that the probability that $P_e \geq 0.75$ is only 25%. For illustration consider Beta(2.2, 1.1) as the prior. It has a mean of 0.667 and the probability that $P_e \geq 0.75$ is 50.7% and that it is ≥ 0.9 is 33.2%. This seems not overly-optimistic based on experiences with many other trials, while still reasonable vague. The associated posterior becomes Beta(32.2, 8.1). In Figure 3, one minus the cumulative distributions (1 – CDF) for two posteriors are shown: Beta(32.2, 8.1) and Beta(31, 8) (*i.e.* based on a prior of Beta(1, 1)).

The robustness of the choice of priors is again illustrated; the probability that $P_e \geq 0.75$ (0.8) is 79.1% (52.5%) for Beta(32.1, 8.1) and 76.9% (50.0%) for Beta(31, 8).

Suppose for the *mRCT* the investigators want 500 patients evaluated for the primary outcome in total (*i.e.* followed for the complete eight-week follow-up period). Suppose further they feel for the sake of credibility they must limit lost-to-follow-up to 20%. As a consequence they recruit 625 (=500/0.8) patients. Based on the information regarding P_e from the *pRCT*, and using the predictive Beta-binomial distribution as before for the number of randomized patients that are evaluated, the probability that 500 of the 650 patients have complete follow-up is 70.3%, for the prior Beta(2.2, 1.1), and 67.8%, for the prior Beta(1, 1).

Results

Using the Results of a pRCT to Inform the Design of the Associated mRCT

Using the Bayesian methods given above a procedure for employing the results of a *pRCT* to inform the design of the associated *mRCT* is provided in this section. By incorporating the uncertainty associated with the results of the *pRCT*, the procedure estimates the *power* of potential recruitment strategies for the *mRCT*. This allows the investigators to choose the strategy that most efficiently achieves the desired power. A recruitment strategy consists of N_a, the maximum number of eligible patients to be approached, and $N_r < N_a$, the maximum number of patients randomized. Consider the situation faced by the investigators who wish to plan a *mRCT* informed by the pilot data from Morone *et al*.[5] Suppose they chose Beta(1, 1) for the prior for P_r,

and Beta(2.2, 1.1) for the prior for P_e, leading to the respective posteriors Beta(38, 41) and Beta(32.2, 8.1). Recognizing that the actually number of randomized and evaluated patients are random variables with Beta-binomial predictive distributions, the procedure uses simulations to determine the expected *power* of a given strategy defined by N_a and N_r. Suppose the type I error probability for the *mRCT* is set to α, using a two-sided test of the null hypothesis. Let δ equal the smallest clinically important difference, expressed in standard deviations. The simulations are performed using the following steps:

Determine n_r = min(Beta-binomial(N_a, 38, 41), N_r) (the number randomized)

Generate $m_e \sim$ Beta-binomial(n_r, 32.2, 8.1) (the number evaluated)

Determine *power* = $\Phi\left(\delta\sqrt{m_e}/2 - z_{(1-\alpha/2)}\right)$, where Φ is the CDF and $z_{(1-\alpha/2)}$ is the $1-\alpha/2$ cut-point for the standard normal random variable.

The average of *power* over the simulations is the expected power of approaching at most N_a eligible patients and randomizing at most N_r of them.

Returning to the example, suppose $\alpha = 0.05$, $\delta = 0.15$ and the desired *power* is 80%, then the total sample size requirement is $\left\lceil 2(z_{(0.975)} + z_{(0.8)})/\delta \right\rceil^2 = 1395$. The naïve approach would be to randomize 1396/0.811 = 1720 and approach 1722/0.481 = 3576. (Recall that 0.481 was the proportion of eligible patients randomized, and 0.811 was the proportion of randomized patients evaluated.) The resulting expected *power* is 0.771. To ensure sufficient *power* the investigators would need to increase the maximum number of eligible patients approached and randomized. For example, if the maximum number of eligible patients approached is set to 4000 and the maximum number of patients randomized is set to 1800, the expected *power* is 0.8.

Alternatively, if the numbers are set to 3800 and 1900, respectively, the expected *power* is also 0.8. Since there are many pairs of N_a and N_r that provide the required expected *power*, the investigators choice would need to consider budgetary and patient availability constraints.

Figure 4 allows examination of the loss of expected *power* due to applying the naïve method. The results are shown as a function of the number of eligible patients approached in the pilot study, the proportion of those that are randomized and the proportion of randomized patients that are evaluated. The expected *power* ranges from 72.1% (for 25 eligible patients approached, with 40% randomized and 95% evaluated) to 78.1% (for 100 eligible patients approached, with 60% randomized and 75% evaluated). In general, since the precision of the parameter estimators increases with the number approached and the proportion randomized, so does the expected *power*. The ranges in Figure 4 of the number approached and the proportion randomized and evaluated span those seen in most pilot studies, and this analysis reveals a small to moderate lost in expected *power* from using the naïve method. Nonetheless, the loss of expected *power* can be avoided by choosing a recruitment strategy that provides, in simulations, the required *power*.

Conclusions

Pilot studies can be an integral part of the planning *RCT*s and are becoming increasingly popular. They can be used for estimating, among other parameters, the randomization, compliance and evaluation probabilities. However, ignoring the uncertainty regarding these estimates can lead to underpowered *RCT*s. To address this issue we propose Bayesian methods for the analysis of pilot studies. While Bayesian methods have become mainstream in many areas of health research,[7,8]

their use in pilot studies is not common. The proposed methods provide direct inference regarding the feasibility of the associated main *RCT* in a simple, intuitive and relevant manner. Data from a published pilot study are used to demonstrate the methods, with particular emphasis on assessing randomization and evaluation feasibility.

For this example at least, the robustness of the choice of priors is demonstrated. Using beta priors and binomial sampling for probability parameters in the pilot study, beta-binomial predictive distributions can be used to determine the probability of randomizing and evaluating the required number of patients in the main *RCT*. Further the Bayesian approach allows the investigators, using the data from a pilot study, to determine the expected power of their patient recruitment strategy for the main *RCT*.

Issues remain, however. Vague priors were used in the example presented here and were illustrated to be reasonably robust, but in any specific application, the robustness of the choice of priors needs to be examined carefully. Although prior elicitation is beyond the scope of this paper, it is important to note that some sensitivity analysis in the hyper-parameters in Figure 1 might be useful in providing a numerical context, and quantitative feedback, to facilitate eliciting beta priors reflecting an investigator's beliefs about Prob(Pr > 0.50) and similar feasibility probabilities. Furthermore, future research regarding the use of more informative priors based on previous similar, or even unrelated, pilot studies needs to be examined. Where appropriate, the use of such priors would lead to a more precise assessment of the feasibility of future *RCT*s. While we acknowledge that determining the sample size for a feasibility study is beyond the scope of this paper, we refer the reader to Cocks and Torgerson[9] for further reading.

References

1. Eldridge SM, Chan CL, Campbell MJ *et al*. CONSORT 2010 statement: extension to randomised pilot and feasibility trials. *Pilot and Feasibility Studies* 2016; 2(1): 64.
2. Eldridge SM, Lancaster GA, Campbell MJ *et al*. Defining feasibility and pilot studies in preparation for randomised controlled trials: Development of a conceptual framework. *PLoS ONE* 2016; 11(3): e0150205.
3. Lancaster GA, Dodd S, Williamson PR. Design and analysis of pilot studies: Recommendations for good practice. *J Eval Clin Pract* 2004; 10(2): 307-312.
4. Thabane L, Ma J, Chu R *et al*. A tutorial on pilot studies: the what, why and how. *BMC Med Res Methodol* 2010 10: 1.
5. Morone NE, Greco CM, Weiner DK. Mindfulness meditation for the treatment of chronic low back pain in older adults: A randomized controlled pilot study. *Pain* 2008; 134(3): 310-319.
6. Newcombe RG. Two-sided confidence intervals for the single proportion: Comparison of seven methods. *Stat Med* 1998; 17(8): 857-872.
7. Berry DA. Bayesian clinical trials. *Nat Rev Drug Discov* 2006; 5(1):27-36.
8. Spiegelhalter DJ, Abrams KR, Myles JP. *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. Chichester: Wiley, 2004.
9. Cocks K, Torgerson BJ. Sample size calculations for pilot randomized trials: a confidence interval approach. *J Clin Epid* 2013; 66(2):197-201.

Figure 1. The probability (%) that the posterior for the randomization probability equals or exceeds the respective column heading, as a function of the prior, defined by the sum of *a* and *b* and the expected value.

	Expected value for prior distribution											
	0.5				0.45				0.4			
a + *b*	*a*	*b*	0.4	0.5	*a*	*b*	0.4	0.5	*a*	*b*	0.4	0.5
2	1	1	**92.6**	**36.7**	0.9	1.1	**92.3**	**35.9**	0.8	1.2	**92.0**	**35.0**
4	2	2	**93.0**	**36.9**	1.8	2.2	**92.4**	**35.2**	1.6	2.4	**91.7**	**33.6**
6	3	3	**93.4**	**37.0**	2.7	3.3	**92.5**	**34.6**	2.4	3.6	**91.5**	**32.2**
8	4	4	**93.7**	**37.2**	3.6	4.4	**92.5**	**33.9**	3.2	4.8	**91.2**	**30.8**
10	5	5	**94.0**	**37.3**	4.5	5.5	**92.6**	**33.3**	4	6	**91.0**	**29.5**

Figure 2. Each number in boldface is the number of eligible patients that need to be approached to have a 90% probability of randomizing the number of patients indicated in the column heading, as a function of the prior, defined by the sum of *a* and *b* and the expected value.

	Expected value for prior distribution								
	0.5			0.45			0.4		
a + *b*	100	500	1000	100	500	1000	100	500	1000
2	**250**	**1228**	**2450**	**252**	**1230**	**2452**	**252**	**1234**	**2464**
4	**249**	**1224**	**2443**	**252**	**1230**	**2452**	**252**	**1238**	**2468**
6	**248**	**1220**	**2435**	**252**	**1230**	**2452**	**254**	**1242**	**2474**
8	**247**	**1217**	**2427**	**250**	**1230**	**2452**	**254**	**1244**	**2480**
10	**247**	**1213**	**2420**	**250**	**1230**	**2452**	**254**	**1246**	**2486**

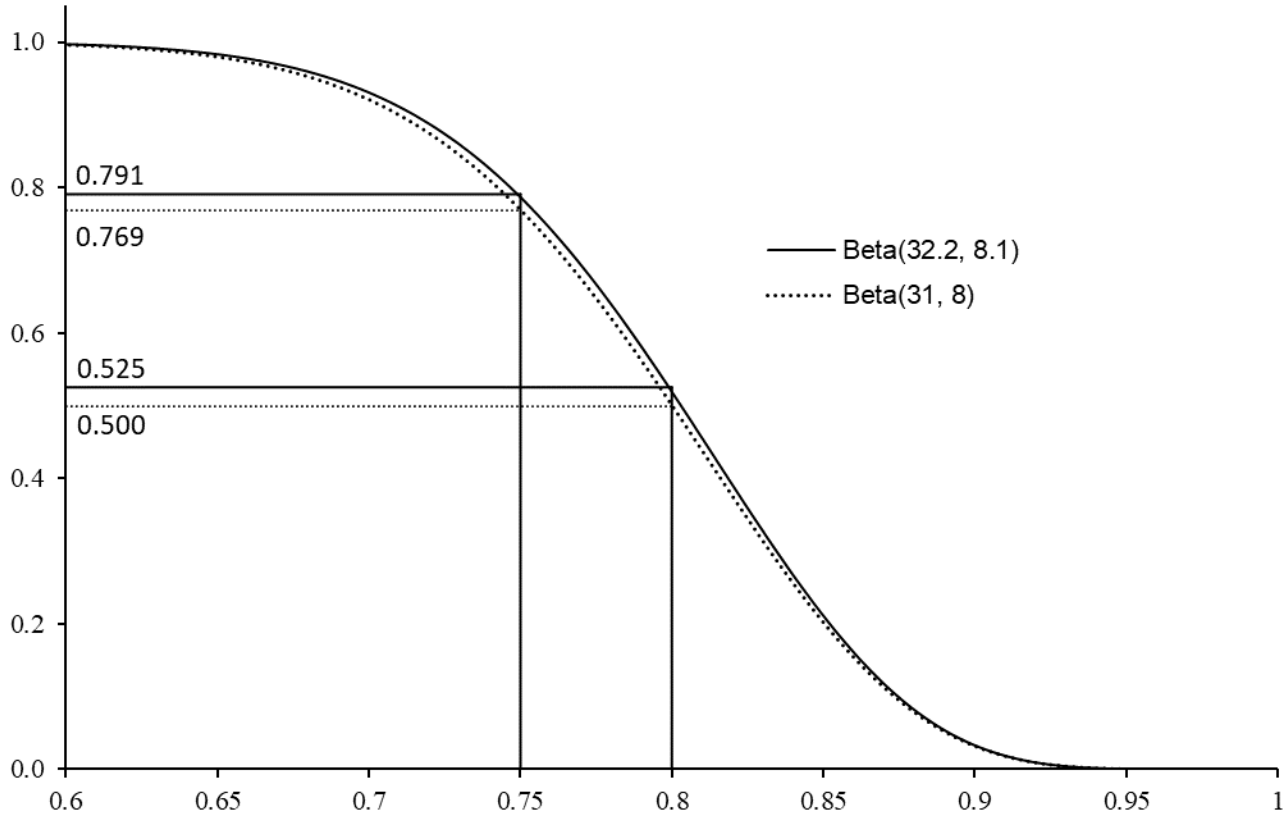

Figure 3. One – CDF of the posterior distribution for the follow-up probability based on two different priors.

Figure 4. Power (%) of applying the naïve method, broken down by the number of eligible patients approached in the pilot study, the proportion of those that were randomized and the proportion of randomized patients that were evaluated. The goal of 80% power was based on a two-sided, 5%-level test of the null hypothesis and a smallest clinically important difference of 0.15 standard deviations.

Number of eligible patients approached	Proportion of eligible patients randomized	Proportion of randomized patients evaluated		
		0.75	0.85	0.95
25	0.4	**73.5**	**72.7**	**72.1**
	0.5	**74.7**	**74.2**	**73.8**
	0.6	**75.5**	**74.7**	**74.4**
50	0.4	**76.0**	**75.5**	**75.2**
	0.5	**76.5**	**76.3**	**75.9**
	0.6	**77.0**	**76.7**	**76.5**
75	0.4	**76.8**	**76.5**	**76.2**
	0.5	**77.3**	**77.2**	**77.0**
	0.6	**77.7**	**77.5**	**77.4**
100	0.4	**77.3**	**77.1**	**76.9**
	0.5	**77.8**	**77.6**	**77.4**
	0.6	**78.1**	**78.0**	**77.8**