Regarding the factors that affect the local control rate, we investigated the age (less than 65 years of age *vs* 65 years of age and over), the T stage (T2, T3 *vs* T4), the site (tongue *vs* oral cavity except tongue), the presence of systemic chemotherapy (Groups 1, 3 *vs* Groups 2, 4), the difference between selective intra-arterial infusion and non-selective intra-arterial infusion (external carotid artery), and the difference between agents for intra-arterial infusion (carboplatin *vs* cisplatin). Regarding the factors that affect the survival rate, we investigated N stage (N0, 1 *vs* N2, 3), clinical stage (III *vs* IV), and PS (0, 1 *vs* 2, 3) in addition to the six factors noted above.

We used the Kaplan–Meier method for survival and local recurrence-free analyses and the log-rank test to determine whether any significant differences existed between different patients in terms of end points. Survival and local recurrence-free rates were calculated (as of April 1, 2007 or the date of the last medical examination) from the start of treatment to the date of the event.

The Cox regression model was used to perform a multivariate analysis.

RESULTS

Patient population

The subjects consisted of 136 patients with locally advanced oral cavity cancer who underwent intra-arterial chemotherapy combined with radiation therapy between January 1993 and December 2006 (Table 1). Because the amount of agent for intra-arterial infusion in two cases out of 136 was less than 50% of the scheduled amount, we performed an analysis with 134 cases infused with 50% or more of the prescribed amount.

Table 2 shows the TNM staging, age, and PS among the four groups. The median age of Groups 1 and 3 was 17 years older than that of the Groups 2 and 4. The percentage of good PS patients was higher in the Groups 2 and 4. Written informed consent was obtained from all the patients.

Follow-up studies were sufficiently performed in the 131 patients except in three patients as of April 2007. The median follow-up duration for patients who were alive was 45.4 months (range: 5–168 months).

Treatment delivery

The selected arteries consisted of the lingual artery in 52 patients, the bilateral lingual arteries in 12 patients, the facial artery in 12 patients, the faciolingual trunk in 4 patients, the maxillary artery in 1 patient, the external carotid artery in 48 patients, the external carotid artery and contralateral lingual artery in 3 patients, and the external carotid artery and contralateral facial artery in 1 patient. During the treatment course, the route was changed from the lingual artery to the external carotid artery in two patients. The total dose of carboplatin ranged from 240 to 800 mg, with a median of 430 mg. The total dose of cisplatin ranged from 40 to 390 mg, with a median of 120 mg. In the arterial injections of carboplatin, 87% of the cases were administered the scheduled quantities of carboplatin, whereas 5% of the cases were administered 50% or less of the scheduled quantities of carboplatin. In the arterial injections of cisplatin, 75% of the cases were administered the scheduled quantities of cisplatin, whereas 5% were administered 50% or less of the scheduled quantities of cisplatin.

Of the patients, 74 (55.2%) patients received systemic chemotherapy. The number of chemotherapy courses was one in 9 patients, and two in 65 (87.8%) patients.

The radiation dose ranged from 27 to 78 Gy, with a median of 63 Gy. Brachytherapy was performed in 41 (30.6%) patients; interstitial irradiation using a Cs needle was performed on 14 patients, and interstitial irradiation using Au grain was performed on 27 patients.

Table 1 Characteristics of 134 patients with squamous cell carcinoma of the oral cavity

Age (years)	
Median	67
Range	25–89
Gender	
Male	89
Female	45
Performance status (ECOG)	
0	26
1	93
2	11
3	4
TNM (2002) T stage	
T1	0
T2	16
T3	67
T4a	49
T4b	2
TNM (2002) N stage	
N0	61
N1	34
N2a	1
N2b	28
N2c	6
N3	4
Stage	
III	63
IV A	67
IV B	4
Primary tumour site	
Tongue	88
Lower gingiva	16
Floor of the mouth	14
Buccal mucosa	12
Upper gingiva	3
Hard palate	1
Reasons not performing surgery	
Refusal	78
Old age	27
Poor performance status	10
Poor cardio-pulmonary function	10
Inoperable advanced lesion	9

Treatment results

A complete response was achieved in 109 patients, and a partial response in 25 patients. A relapse was detected in 65 patients: primary site, 36 patients; cervical lymph node, 19 patients; primary site and cervical lymph node, 2 patients; primary site and distant metastasis, 1 patient; cervical lymph node and distant metastasis, 2 patients; distant metastasis, 5 patients. The 3-year local (primary site) recurrence-free rate of all patients was 68.6% (95% confidence interval (CI): 60.6–77.7%) (Figure 2A). Cumulative local recurrence-free rate of T2-3 and T4 patients at 3 years were 77.9% (95% CI: 69.1–87.9%) and 51.3% (95% CI: 37.5–70.2%), respectively (Figure 2B).

Of the patients demonstrating a relapse, salvage surgery was performed in 15 patients, intra-arterial chemoradiation in 9 patients, intra-arterial chemotherapy in 3 patients, chemoradiation

therapy in 3 patients, radiation therapy in 5 patients, and chemotherapy in 1 patient. Of these 36 patients, 9 patients had successful salvage (surgery in 5 patients, intra-arterial chemoradiation in 3 patients, chemoradiation therapy in 1 patient), becoming disease-free after the procedure.

Table 2 TNM stage, age, and PS among the groups

	Group 1 ($n=39$)	Group 2 ($n=26$)	Group 3 ($n=21$)	Group 4 ($n=48$)
Stage				
III	26	14	5	18
IVA	13	12	14	28
IVB	0	0	2	2
Age (years)				
Median	73	59	77	59
Range	51–90	25–73	62–87	25–73
PS				
0	2	8	3	13
1	29	17	14	33
2	6	1	3	1
3	2	0	1	1

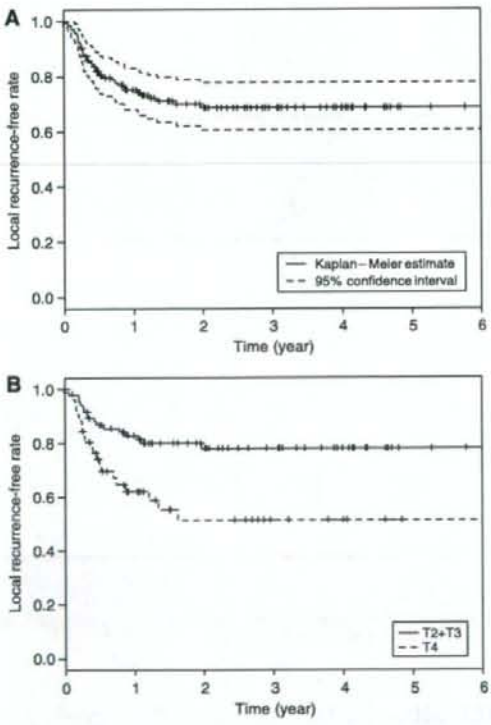


Figure 2 Actuarial local (primary site) recurrence-free rate of patients with advanced squamous cell carcinoma of the oral cavity by Kaplan–Meier method. (**A**) Actuarial local (primary site) recurrence-free rate of all patients. A solid line: local control rate curve. A broken line: 95% CI. (**B**) Actuarial local (primary site) recurrence-free rate according to the T stage. A solid line: T2+T3 cases ($n=83$). A broken line: T4 cases ($n=51$). $P=0.00531$.

At the time of analysis, 65 patients had died, 66 patients were still alive, and 3 patients had been lost to the follow-up. In the 65 patients who had died, the cause of death was oral cavity cancer in 45, other diseases in 18, and treatment-related complication in 2. The 3-year overall survival of all patients was 53.9% (95% CI: 45.4–64.0%) (Figure 3A). Cumulative survival rates of stages III and IV patients at 3 years were 62.9% (95% CI: 51.4–77.0%) and 45.3% (95% CI: 33.9–60.5%), respectively (Figure 3B).

Factors of survival and local recurrence

In a univariate analysis, T factor, the selected artery, and the site were found to have a significant impact on local recurrence, whereas systemic chemotherapy and difference of IA chemotherapy had only a marginal significance. In a multivariate analysis, T factor and difference of IA chemotherapy were of borderline significance (Table 3).

In a univariate analysis, age, systemic chemotherapy, and difference of IA chemotherapy were found to have a significant impact on survival. In a multivariate analysis, age, difference of IA chemotherapy, and selected artery were found to have a significant impact on survival, whereas systemic chemotherapy was not a significant factor (Table 4).

Acute toxicity

Acute toxicity is summarised in Table 5. Grade 3 or higher toxic changes included granulocytopaenia in 60 (45%) patients, thrombopaenia in 31 (23%) patients, anaemia in 26 (19%) patients, and mucositis in 15 (11%) patients. There was no significant difference in the degree of acute toxicity among the four groups.

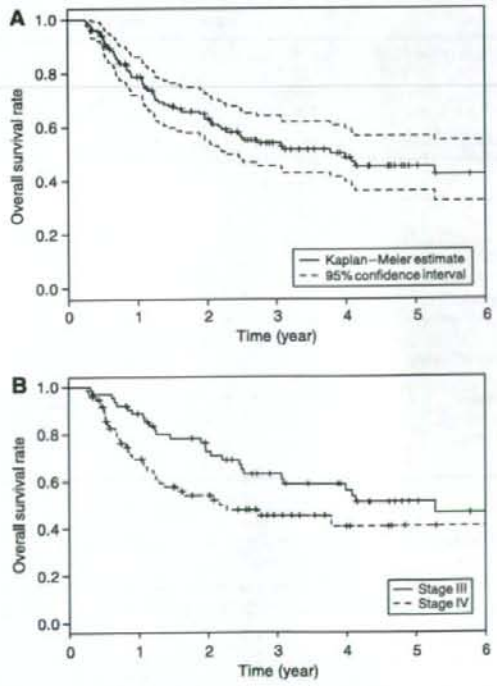


Figure 3 Actuarial survival rates of patients with advanced squamous cell carcinoma of the oral cavity by Kaplan–Meier method. (**A**) Actuarial survival rates in all 134 patients. A solid line: overall survival curve. A broken line: 95% CI. (**B**) Actuarial survival rates according to the stage. A solid line: stage III cases ($n=63$). A broken line: stage IV cases ($n=71$). $P=0.117$.

Table 3 Results of the multivariate analysis of prognostic factor on local recurrence-free time based on Cox proportional-hazards model

Selected factor	Level	Adjusted *P*-value[a]	Adjusted HR (95% confidence interval)
T classification	T2 or T3	0.0501	1.000 (referent)
	T4		2.11 (0.999, 4.45)
Site	Tongue	0.684	1.000 (referent)
	Oral cavity except tongue		1.16 (0.558, 2.42)
Systemic chemotherapy	NO (Group 1 or 3)	0.265	1.000 (referent)
	YES (Group 2 or 4)		0.682 (0.347, 1.33)
IA chemotherapy	CBDCA	0.0884	1.000 (referent)
	CDDP		0.552 (0.278, 1.09)
Artery	Selected artery	0.106	1.000 (referent)
	External carotid artery		1.76 (0.885, 3.53)

[a]The *P*-value for the log-rank test.

Table 4 Results of the multivariate analysis of prognostic factors on overall survival based on Cox proportional-hazards model

Selected factor	Level	Adjusted *P*-value)[a]	Adjusted HR (95% confidence interval)
Age (year)	<65	0.0185*	1.000 (referent)
	>=65		2.05 (1.12, 3.75)
Systemic chemotherapy	NO (Group 1 or 3)	0.104	1.000 (referent)
	YES (Group 2 or 4)		0.610 (0.336, 1.10)
IA chemotherapy	CBDCA	0.0141*	1.000 (referent)
	CDDP		0.477 (0.265, 0.861)
Artery	Selected artery	0.0290*	1.000 (referent)
	External carotid artery		1.74 (1.05, 2.86)

[a]The *P*-value for the Wald's test. *P-value <0.05.

Table 5 Toxicity according to the group

	Group 1 ($n=39$)			Group 2 ($n=26$)			Group 3 ($n=21$)			Group 4 ($n=48$)		
Toxicities	0–2	3	4	0–2	3	4	0–2	3	4	0–2	3	4
Haematologic												
White blood cell	21	15	3	15	5	6	9	8	4	24	18	6
Granulocyte	22	13	4	15	5	6	11	8	2	26	14	8
Platelet	32	4	3	14	6	6	18	2	1	39	4	5
Haemoglobin	34	4	1	20	6	0	18	3	0	36	10	2
Non-haematologic												
Liver	39	0	0	25	1	0	21	0	0	47	1	0
Kidney	39	0	0	26	0	0	21	0	0	48	0	0
Vomiting	39	0	0	26	0	0	21	0	0	46	2	0
Mucocitis	37	2	0	23	3	0	19	2	0	40	8	0
Fever	37	2	0	25	1	0	20	1	0	48	0	0

In addition, no transient or persistent central nervous complications were observed. Treatment-related death was confirmed in two patients. Although tumours in both patients disappeared as a result of therapy, both patients died of gastrointestinal bleeding.

Chronic toxicity

We studied chronic toxicity in 97 patients who survived more than 12 months after the treatment. Although these 97 patients did not develop severe problems in their phonation or deglutition function and were able to eat almost normally, continuous glossalgia was recognised in one of the patients and analgesic was sometimes necessary for this patient, and two patients developed osteoradionecrosis, which needed surgery.

DISCUSSION

The results of radiotherapy for advanced oral cavity cancer alone were poor (Decroix and Ghossein, 1981; Horiuchi *et al*, 1982). Currently, surgery is the standard treatment (Poulsen *et al*, 1996). Several studies have reported arterial injection therapy for oral cavity cancer; however, the number of patients in such studies tended to be small, and its usefulness has not yet been clearly

demonstrated (Hirai *et al*, 1999; Damascelli *et al*, 2003; Kovacs, 2004). No randomised controlled trials have yet been performed to compare the effectiveness of surgery with chemoradiation, and the usefulness of chemoradiation therapy involving systemic chemotherapy thus remains to be clarified.

After surgery, the 5-year survival rate ranged from 27 to 60% of stage III patients, while it ranged from approximately 12 to 40% of stage IV patients (Chen *et al*, 1999; Sessions *et al*, 2002; Greenberg *et al*, 2003; Lo *et al*, 2003; Gorsky *et al*, 2004; Liao *et al*, 2006; Fan *et al*, 2007); our results were therefore similar to the results after surgery. In particular, the results of arterial injection therapy by cisplatin with sodium thiosulphate were excellent.

Currently, there are two procedures for performing arterial infusion therapy for head and neck cancer: namely, a procedure in which a catheter is inserted into the target artery through the superficial temporal artery, as presented in this study, and a procedure in which a catheter is inserted into the target artery through the femoral artery by Seldginger's procedure (Robbins *et al*, 1994, 2000; Balm *et al*, 2004; Alkureishi *et al*, 2006). The latter procedure is simpler than the former procedure, and it facilitates the administration of anticancer agents into several arteries; however, drug administration over a long duration is impossible, and catheter operation-related cranial nerve disorders may sometimes occur. Although Robbins *et al* reported the incidence of cranial nerve disorders to be 2–4%, no patient showed a cranial nerve disorder in our study. While, in his series, the mean age was 56 years, the median age was 67 years in our study. Therefore, arterial injection therapy in which a catheter is inserted into the target artery through the superficial temporal artery may therefore be appropriate for elderly patients, and the drug can thus be administered over a long duration.

The dose of cisplatin in our study was approximately 1/5 of that described by Robbins *et al*. However, the duration of administration was 60 times longer. It has been reported that the antitumour effects of cisplatin are correlated with the concentration and duration of administration (DeConti *et al*, 1973; Drewinko *et al*, 1973); low-dose cisplatin may exhibit potent antitumour effects.

It remains unclear whether a neutralizer sodium thiosulphate needs to be added to such administered quantities, but we believed that the addition of sodium thiosulphate was necessary, because our study involved elderly patients and patients who have malfunctioning kidneys, and also because the combined usage of systemic chemotherapy was one of the assumptions of the study. Regarding the radiation dosage when arterial injection chemotherapy and radiotherapy were used concomitantly, we administered the same doses as those that we used for cases of combined use of systemic chemotherapy. One of our future tasks is to examine long-term adverse effects, but no severe late-onset effects have been observed thus far.

A randomised controlled trial was performed in the Netherlands, comparing chemoradiation that used cisplatin as a drug agent for arterial injection in a method similar to that of Robbins *et al* with chemoradiation that used cisplatin as systemic chemotherapy. Unfortunately, the initial outcome showed no difference between the survival rates of the therapies. In the Netherlands study, there were few oral cavity cancer patients (Balm and Rasch, 2006). Although the effectiveness of chemoradiation that concomitantly uses systemic chemotherapy has been clarified for cases of pharyngeal and laryngeal cancer, it has not yet been clarified for cases of oral cavity cancer. Thus, we suggest that randomised controlled trials limit their target cases to, for example, patients with oral cavity cancer.

The results of our study were similar to those of surgery, thus suggesting the usefulness of arterial injection therapy combined with radiation therapy. In particular, the results of arterial injection therapy by cisplatin with sodium thiosulphate were excellent, so we believe that it will be a new therapy for locally advanced oral cavity cancer.

Conflict of interest

The authors state no conflict of interest.

REFERENCES

Alkureishi LW, de Bree R, Ross GL (2006) RADPLAT: an alternative to surgery? *Oncologist* **11:** 469–480

Balm AJ, Rasch CR (2006) Results of the Dutch trial IA *vs* IV chemoirradiation in locally advanced head and neck cancer. *Abstract of the International Workshop on Intra-Arterial Chemotherapy for Head and Neck Cancer*; Chicago, IL

Balm AJ, Rasch CR, Schornagel JH, Hilgers FJ, Keus RB, Schultze-Kool L, Ackerstaff AH, Busschers W, Tan IB (2004) High-dose superselective intra-arterial cisplatin and concomitant radiation (RADPLAT) for advanced head and neck cancer. *Head Neck* **26:** 485–493

Chen YK, Huang HC, Lin LM, Lin CC (1999) Primary oral squamous cell carcinoma: an analysis of 703 cases in southern Taiwan. *Oral Oncol* **35:** 173–179

Damascelli B, Patelli GL, Lanocita R, Di Tolla G, Frigerio LF, Marchiano A, Garbagnati F, Spreafico C, Ticha V, Gladin CR, Palazzi M, Crippa F, Oldini C, Calo S, Bonaccorsi A, Mattavelli F, Costa L, Mariani L, Cantu G (2003) A novel intraarterial chemotherapy using paclitaxel in albumin nanoparticles to treat advanced squamous cell carcinoma of the tongue: preliminary findings. *Am J Roentgenol* **181:** 253–260

DeConti RC, Toftness BR, Lange RC, Creasey WA (1973) Clinical and pharmacological studies with *cis*-diamminedichloroplatinum (II). *Cancer Res* **33:** 1310–1315

Decroix Y, Ghossein NA (1981) Experience of the Curie Institute in treatment of cancer of the mobile tongue: I. Treatment policies and result. *Cancer* **47:** 496–502

Drewinko B, Brown BW, Gottlieb JA (1973) The effect of *cis*-diamminedichloroplatinum (II) on cultured human lymphoma cells and its therapeutic implications. *Cancer Res* **33:** 3091–3095

Fan KH, Lin CY, Kang CJ, Huang SF, Wang HM, Chen EY, Chen IH, Liao CT, Cheng AJ, Chang JT (2007) Combined-modality treatment for advanced oral tongue squamous cell carcinoma. *Int J Radiat Oncol Biol Phys* **67:** 453–461

Fuwa N, Ito Y, Matsumoto A, Kamata M, Kodaira T, Furutani K, Sasaoka M, Kimura Y, Morita K (2000) A combination therapy of continuous superselective intraarterial carboplatin infusion and radiation therapy for locally advanced head and neck carcinoma. Phase I study. *Cancer* **89:** 2099–2105

Fuwa N, Kodaira T, Furutani K, Tachibana H, Nakamura T, Daimon T (2007) Chemoradiation therapy using radiotherapy, systemic chemotherapy with 5-fluorouracil and nedaplatin, and intra-arterial infusion using carboplatin for locally advanced head and neck cancer – Phase II study. *Oral Oncol* **43:** 1014–1020

Gorsky M, Epstein JB, Oakley C, Le ND, Hay J, Stevenson-Moore P (2004) Carcinoma of the tongue: a case series analysis of clinical presentation, risk factors, staging, and outcome. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod* **98:** 546–552

Greenberg JS, El Naggar AK, Mo V, Roberts D, Myers JN (2003) Disparity in pathologic and clinical lymph node staging in oral tongue carcinoma. Implication for therapeutic decision making. *Cancer* **98:** 508–515

Hirai T, Korogi Y, Hamatake S, Nishimura R, Baba Y, Takahashi M, Uji Y, Taen A (1999) Stages III and IV squamous cell carcinoma of the mouth: three-year experience with superselective intraarterial chemotherapy using cisplatin prior to definitive treatment. *Cardiovasc Intervent Radiol* **22:** 201–205

Horiuchi J, Okuyama T, Shibuya H, Takeda M (1982) Results of brachytherapy for cancer of the tongue with special emphasis on local prognosis. *Int J Radiat Oncol Biol Phys* **8:** 829–835

Kovacs AF (2004) Intra-arterial induction high-dose chemotherapy with cisplatin for oral and oropharyngeal cancer: long-term results. *Br J Cancer* **90(7):** 1323–1328

Liao CT, Chang JT, Wang HM, Ng SH, Hsueh C, Lee LY, Lin CH, Chen IH, Kang CJ, Huang SF, Tsai MF, Yen TC (2006) Surgical outcome of T4a and resected T4b oral cavity cancer. *Cancer* **107:** 337–344

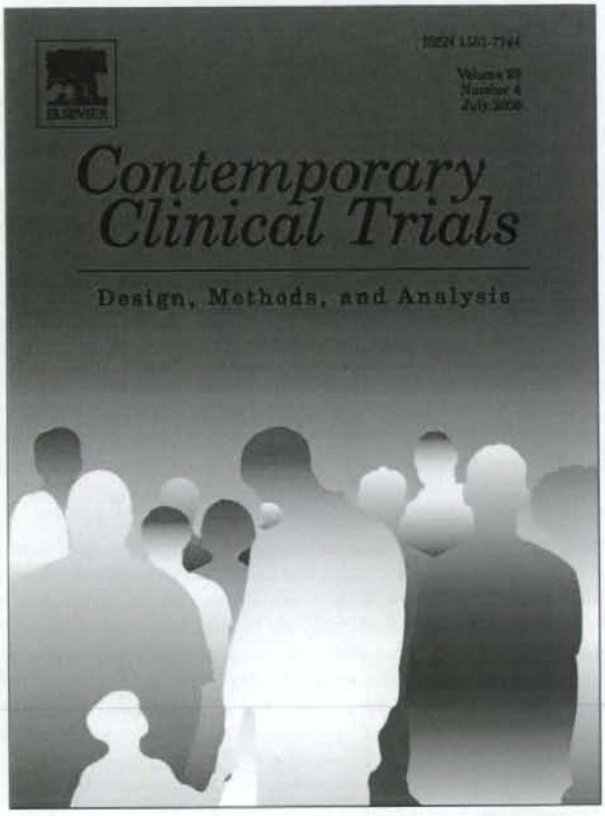
Lo WL, Kao SY, Chi LY, Wong YK, Chang RC (2003) Outcomes of oral squamous cell carcinoma in Taiwan after surgical therapy: factors affecting survival. *J Oral Maxillofac Surg* **61:** 751–758

Poulsen M, Aldren C, Tripcony L, Walker Q (1996) Is surgery necessary in stage III and stage IV cancer of head and neck that responds to induction chemotherapy? *Arch Otolaryngol Head Neck Surg* **122:** 467–471

Robbins KT, Kumar P, Wong FS, Hartsell WF, Flick P, Palmer R, Weir III AB, Neill HB, Murry T, Ferguson R, Hanchett C, Vieira F, Bush A, Howell SB (2000) Targeted chemoradiation for advanced head and neck cancer: analysis of 213 patients. *Head Neck* **22:** 687–693

Robbins KT, Storniolo AM, Kerber C, Vicario D, Seagren S, Shea M, Hanchett C, Los G, Howell SB (1994) Phase I study of highly selective supradose cisplatin infusions for advanced head and neck cancer. *J Clin Oncol* **12:** 2113–2120

Sessions DG, Spector GJ, Lenox J, Haughey B, Chao C, Marks J (2002) Analysis of treatment results for oral tongue cancer. *Laryngoscope* **112:** 616–625



Available online at www.sciencedirect.com
ScienceDirect
Contemporary Clinical Trials 29 (2008) 507–516
Contemporary Clinical Trials
www.elsevier.com/locate/conclintrial

Bayesian sample size calculations for a non-inferiority test of two proportions in clinical trials ☆

Takashi Daimon *

Department of Drug Evaluation & Informatics, Graduate School of Pharmaceutical Sciences, University of Shizuoka, 52-1, Yada, Suruga-ku, Shizuoka 422-8526, Japan

Received 18 July 2007; accepted 7 December 2007

Abstract

In the process of clinical trials and health-care evaluation, Bayesian approaches have increasingly become the center of attention. In this article, sample size calculations for a non-inferiority test of two independent binomial proportions in a clinical trial are considered in a Bayesian framework. The hybrid Neyman–Pearson–Bayesian (hNPB) probability, the conditionally Bayesian (cB) probability and the unconditionally Bayesian (uB) probability are formulated through a conjugate normal analysis. The sample sizes are calculated based on formulas where normal prior distributions are assumed, and are compared with the Neyman–Pearson (NP) sample size. Our results show that the sample size based on the hNPB probability allows us to critically evaluate the appropriateness of the NP sample size. It is suggested that the sample size calculated based on the cB probability formula is smaller than the NP sample size.

Keywords: Prior distribution; Predictive probability; Conjugate normal analysis; Futility

1. Introduction

In clinical research, clinical trials are usually conducted for evaluation of the efficacy and safety of a test drug as compared to a placebo control or an active control agent (e.g., a standard treatment) in terms of discrete variables, for example, clinical response (e.g., complete response, partial response, and stable response), survival in cancer trials, and the presence of adverse events. The objectives of the intended clinical trials usually include the evaluation of the effect, the demonstration of therapeutic equivalence/non-inferiority and the establishment of superiority, and for evaluation of these objectives based on discrete clinical endpoints, the proportions of events that have occurred in different treatment groups are often compared.

For calculation of a sample size for comparing proportions between two independent groups, especially in a confirmatory clinical trial, an approach based on Neyman–Pearson (NP) hypothesis testing is usually utilized. In this approach, a value for a true difference between two proportions to be detected that has clinical meaning, together with type I and type II errors have to be specified in advance of conducting a trial. However,

☆ This research was supported partly by a grant from the Japanese Ministry of Education, Culture, Sports, Science and Technology (Grant number 18700272).
* Tel./fax: +81 54 264 5770.
E-mail address: daimon@u-shizuoka-ken.ac.jp.

1551-7144/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.cct.2007.12.001

some authors have recommended assuming some prior distribution for the difference, rather than pre-specifying such a value in a Bayesian framework [1–7]. Most of their calculated sample sizes are based on the posterior distribution.

For calculation of a sample size for demonstrating therapeutic non-inferiority of a test treatment against an active control, in addition to specifying the value for the difference, a margin, or a delta (Δ) for non-inferiority under a given trial design needs to be considered [8]. For sample size calculation based on NP hypothesis testing, a specific value for the margin needs to be set. The value for the margin is pre-specified by taking into account the size of the effect of the control treatment and factors such as the efficacy, safety, cost and applicability of each treatment, although the value for the difference is pre-specified based on knowledge from previous research.

Investigators planning a trial may have a variety of opinions about the pre-specified values for the margin and the difference, on the basis of their own experience, knowledge, or impressions of a test treatment or a control. Thus, considerable discussion is often necessary to arrive at a consensus on specific values for the margin and the difference. This process requires considerable effort and is often extremely difficult [9–11].

Therefore, since investigators may have differing opinions on appropriate values for the margin and the difference, it may be necessary to calculate varying sample sizes that reflect their various opinions, to compare them with the sample size calculated for the agreed values, and to recognize the differences among the calculated sample sizes. In addition, investigators may be interested not only in interpretations of the results based on the frequentist approach, such as a *p*-value, an estimate, and a confidence interval, but also, often in an intuitive Bayesian answer to the question 'what is the probability that a test treatment is not inferior to a control?' Therefore, it is important to provide an approach for the design or sample size calculation to allow for such a Bayesian answer, analysis, interpretation, or report.

In this paper, we provide a simplified Bayesian approach to calculate the sample size in a non-inferiority clinical trial with two proportions, by assuming a normal prior distribution for the sum of the difference and the margin. In Section 2, formulas for the power and sample size based on NP hypothesis testing are derived. In Section 3, several probabilities are formulated through a conjugate normal analysis. In Section 4, sample sizes are calculated based on the formulated probability, as compared with the Neyman–Pearson sample size. In Section 5, we conclude with several remarks on applicability of the proposed approach.

2. The Neyman–Pearson power

Let X_{ij} be a binary response from the jth subject in the ith treatment group, $j=1,\ldots,n_i$, $i=0,1$. For example, X_{ij} could be an indicator for an event of interest, i.e., $X_{ij}=1$ for the presence of the event and $X_{ij}=0$ for the absence. For a fixed i, it is assumed that X_{ij}'s are independently and identically distributed with $Pr\ (X_{ij}=1)=\pi_i$. In practice, π_i is usually estimated by the observed proportion in the ith treatment group:

$$\hat{\pi}_i = \frac{1}{n_i}\sum\nolimits_{j=1}^{n_i} X_{ij}.$$

Let $\delta=\pi_1-\pi_0$ be the difference between the true mean proportion of a test treatment (π_1) and a control (π_0). Without loss of generality, we may consider $\delta>0$ ($\delta<0$) as an indication of improvement (worsening) due to the test treatment as compared to the control value. The problem of testing non-inferiority can be given by the following null and alternative hypotheses:

$$\mathrm{H}_0 : \delta + \delta^*_{\mathrm{margin}} \le 0 \text{ versus } \mathrm{H}_A : \delta + \delta^*_{\mathrm{margin}} > 0, \qquad (1)$$

where $\delta^*_{\mathrm{margin}}$ is known as the non-inferiority margin and is fixed at a pre-specified value in advance of conducting a trial. When $\delta^*_{\mathrm{margin}}>0$, the rejection of the null hypothesis indicates the non-inferiority of the test treatment against the control. On the other hand, when $\delta^*_{\mathrm{margin}}<0$, the rejection of the null hypothesis indicates the superiority of the test treatment over the control.

For large n_i, under the null hypothesis, the statistic $T=\hat{\pi}_1-\hat{\pi}_0+\delta^*_{\mathrm{margin}}$ is assumed to be normally distributed with mean $\delta+\delta^*_{\mathrm{margin}}$ and variance τ^2: $T\sim N[\delta+\delta^*_{\mathrm{margin}}, \tau^2]$, where $\tau^2=\pi_1(1-\pi_1)/n_1+\pi_0(1-\pi_0)/n_0$. The null hypothesis will be rejected at the α level of significance if T satisfies

$$T > -z_\alpha\tau, \qquad (2)$$

where z_α is the lower αth percentile of the standard normal distribution and $\Phi(z_\alpha)=Pr(Z\le z_\alpha)=\int_{-\infty}^{z_\alpha}\frac{1}{\sqrt{2\pi}}\exp(-z^2/2)\mathrm{d}z=\alpha$, where Z is the standard normally distributed random variable. If this event ($T>-z_\alpha\tau$) occurring is denoted as C^{NP}_α, then under the alternative hypothesis when specifying δ as $\delta^*=\pi_1^*-\pi_0^*$ in (1), H_A^*: $\delta^*+\delta^*_{\mathrm{margin}}>0$, the power for the test of hypotheses (1) is approximately

$$Pr\left(C^{\mathrm{NP}}_\alpha|\delta^*+\delta^*_{\mathrm{margin}}\right)=\Phi\left[z_\alpha+\frac{\delta^*+\delta^*_{\mathrm{margin}}}{\tau}\right]=1-\beta. \qquad (3)$$

where β is the type II error. The above probability is often referred to as 'classical power', but in this paper we call it the 'Neyman–Pearson (NP) power'. For $n_1=n_0$, the sample size per group needed for achieving power $1-\beta$, n^{NP} (NP sample size), is given by the following [12]:

$$n^{NP}=\frac{(z_\alpha+z_\beta)^2}{\left(\delta^*+\delta^*_{margin}\right)^2}\left[\pi_1^*\left(1-\pi_1^*\right)+\pi_0^*\left(1-\pi_0^*\right)\right]. \tag{4}$$

3. Bayesian probabilities

3.1. The hybrid Neyman–Pearson–Bayesian probability

Let δ_{margin} be a parameter for the non-inferiority margin and suppose we have a normal prior $\delta+\delta_{margin}\sim N[\delta^*+\delta^*_{margin},\sigma_0^2]$. Taking the variation of $\delta^*+\delta^*_{margin}$ into account, the probability of obtaining a 'significant' result in the Neyman–Pearson context when testing the null hypothesis $\delta+\delta^*_{margin}\le 0$, $Pr(C_\alpha^{NP})$ is given by

$$Pr\left(C_\alpha^{NP}\right)=\Phi\left[\frac{z_\alpha\tau+\delta^*+\delta^*_{margin}}{\sqrt{\sigma_0^2+\tau^2}}\right]. \tag{5}$$

This probability is often called the 'expected power' [13] or 'strength' [14], but in this paper is called the 'hybrid Neyman–Pearson–Bayesian (hNPB) probability'. The hNPB probability can be obtained by integrating (3) with respect to the prior, or by evaluating the chance of the critical event (3) occurring via the predictive distribution for the statistic T [4]. As $\sigma_0^2\to 0$, the prior tends to a lump on $\delta^*+\delta^*_{margin}$ and $Pr(C_\alpha^{NP})$ tends to the NP power. When σ_0^2 and τ^2 are both non-zero, $Pr(C_\alpha^{NP})$ is always smaller than $Pr(C_\alpha^{NP}|\delta^*+\delta^*_{margin})$. In this way, the hNPB probability formula allows us to take account of an available prior and to critically evaluate the NP power formula.

3.2. The conditionally Bayesian probability

Unlike the hNPB probability, for which the Neyman–Pearson critical region is utilized, we now wish to carry out a fully Bayesian design, analysis, interpretation and report, in which the prior is explicitly incorporated. Given the data, it is possible to calculate the probability of obtaining a 'significant' Bayesian result when testing the null hypothesis $\delta+\delta_{margin}\le 0$. We shall denote such 'Bayesian significance' as $C_\alpha^B\equiv Pr(\delta+\delta_{margin}\le 0|\text{data})<\alpha$.

Suppose we have the same normal prior $\delta+\delta_{margin}\sim N[\delta^*+\delta^*_{margin},\sigma_0^2]$ and likelihood $T\sim N[\delta+\delta_{margin},\tau^2]$ as in Section 2. The posterior distribution is obtained as

$$\delta+\delta_{margin}|T\sim N\left[\frac{\frac{\delta^*+\delta^*_{margin}}{\sigma_0^2}+\frac{T}{\tau^2}}{\frac{1}{\sigma_0^2}+\frac{1}{\tau^2}},\frac{1}{\sigma_0^2}+\frac{1}{\tau^2}\right], \tag{6}$$

and C_α^B will occur when T satisfies

$$T>-z_\alpha\left(\frac{1}{\sigma_0^2}+\frac{1}{\tau^2}\right)^{1/2}\left(\tau^2-\delta^*+\delta^*_{margin}\right)\frac{\tau^2}{\sigma_0^2}. \tag{7}$$

Hence C_α^B will occur with probability

$$Pr\left(C_\alpha^B|\delta^*+\delta^*_{margin}\right)=\Phi\left[z_\alpha\left(\frac{1}{\sigma_0^2}+\frac{1}{\tau^2}\right)^{1/2}\tau+\frac{\delta^*+\delta^*_{margin}}{\tau}\left(\frac{\tau^2}{\sigma_0^2}+1\right)\right]. \tag{8}$$

In this paper this probability is called the 'conditionally Bayesian (cB) probability. With an increasingly vague prior opinion, $\sigma_0^2\to\infty$ and $Pr(C_\alpha^B|\delta^*+\delta^*_{margin})$ approaches the NP power given by (3).

3.3. The unconditionally Bayesian probability

The probability of obtaining a 'significant' Bayesian result, not conditional on the pre-specification of $\delta^*+\delta^*_{margin}$, is obtained by allowing $Pr(C_\alpha^B|\delta^*+\delta^*_{margin})$ to vary and integrating (8) with respect to the prior, yielding

$$Pr\left(C_\alpha^B\right)=\Phi\left[\frac{z_\alpha\left(\frac{1}{\sigma_0^2}+\frac{1}{\tau^2}\right)^{1/2}\tau^2+\left(\delta^*+\delta^*_{margin}\right)\left(\frac{\tau^2}{\sigma_0^2}+1\right)}{\sqrt{\sigma_0^2+\tau^2}}\right]. \tag{9}$$

Here this probability is called the 'unconditionally Bayesian (uB) probability.

4. Sample size calculations

4.1. Example: a non-inferiority trial for granisetron 1 mg to 3 mg

A randomized, double-masked, controlled trial of two intravenous doses of granisetron (1 mg vs. 3 mg) for the prophylaxis of chemotherapy-induced nausea and vomiting in cancer patients is now planned. The main objective is to investigate the non-inferiority of 1 mg of

Table 1
The NP sample size n^{NP} and the hNPB sample size n^{hNPB} for $\delta^* = -0.075, -0.05$, and 0

δ^*	π_1^*	π_0^*	n^{NP}	n^{hNPB}					
				$\gamma=0.05$	$\gamma=0.01$	$\gamma=0.001$	$\gamma=10^{-6}$	$\gamma=10^{-12}$	$\gamma=10^{-20}$
−0.075	0.025	0.1	1923	16,093	4568	3822	3052	2094	2019
	0.075	0.15	3310	27,703	7862	6579	5254	3603	3476
	0.125	0.2	4529	37,903	10,757	9001	7188	4930	4755
	0.175	0.25	5580	46,692	13,252	11,089	8856	6074	5858
	0.225	0.3	6463	54,070	15,348	12,843	10,256	7034	6785
	0.275	0.35	7177	60,065	17,046	14,264	11,391	7812	7535
	0.325	0.4	7723	64,650	18,345	15,350	12,258	8407	8109
	0.375	0.45	8102	67,796	19,242	16,101	12,858	8819	8506
	0.425	0.5	8312	69,560	19,741	16,519	13,192	9047	8727
	0.475	0.55	8354	69,912	19,841	16,603	13,258	9093	8771
	0.525	0.6	8228	68,854	19,542	16,352	13,058	8956	8638
	0.575	0.65	7934	66,386	18,842	15,767	12,591	8636	8329
	0.625	0.7	7471	62,534	17,746	14,849	11,858	8133	7844
	0.675	0.75	6841	57,244	16,247	13,595	10,857	7446	7182
	0.725	0.8	6042	50,571	14,351	12,009	9590	6577	6344
	0.775	0.85	5076	42,460	12,053	10,086	8055	5525	5329
	0.825	0.9	3941	32,966	9358	7831	6254	4289	4137
	0.875	0.95	2638	22,061	6264	5242	4186	2871	2769
−0.05	0.05	0.1	578	4836	1373	1149	918	630	607
	0.1	0.15	915	7630	2170	1816	1451	995	960
	0.15	0.2	1209	10,098	2869	2401	1918	1316	1269
	0.2	0.25	1461	12,214	3469	2903	2318	1590	1534
	0.25	0.3	1671	13,977	3968	3321	2652	1819	1755
	0.3	0.35	1839	15,388	4368	3655	2919	2002	1931
	0.35	0.4	1965	16,446	4668	3906	3119	2139	2063
	0.4	0.45	2049	17,151	4867	4073	3253	2231	2152
	0.45	0.5	2091	17,504	4967	4156	3319	2277	2196
	0.5	0.55	2091	17,504	4967	4156	3319	2277	2196
	0.55	0.6	2049	17,151	4867	4073	3253	2231	2152
	0.6	0.65	1965	16,446	4668	3906	3119	2139	2063
	0.65	0.7	1839	15,388	4368	3655	2919	2002	1931
	0.7	0.75	1671	13,977	3968	3321	2652	1819	1755
	0.75	0.8	1461	12,214	3469	2903	2318	1590	1534
	0.8	0.85	1209	10,098	2869	2401	1918	1316	1269
	0.85	0.9	915	7630	2170	1816	1451	995	960
	0.9	0.95	578	4836	1373	1149	918	630	607
0	0.05	0.05	100	831	237	199	159	109	105
	0.1	0.1	190	1560	448	375	300	206	199
	0.15	0.15	268	2242	637	533	426	292	282
	0.2	0.2	337	2794	797	668	534	366	353
	0.25	0.25	395	3272	934	782	625	429	414
	0.3	0.3	442	3676	1047	877	700	481	464
	0.35	0.35	479	3978	1134	949	759	521	502
	0.4	0.4	505	4205	1197	1002	801	549	530
	0.45	0.45	521	4330	1234	1033	825	567	547
	0.5	0.5	526	4381	1247	1044	834	572	552
	0.55	0.55	521	4330	1234	1033	825	567	547
	0.6	0.6	505	4205	1197	1002	801	549	530
	0.65	0.65	479	3978	1134	949	759	521	502
	0.7	0.7	442	3676	1047	877	700	481	464
	0.75	0.75	395	3272	934	782	625	429	414
	0.8	0.8	337	2794	797	668	534	366	353
	0.85	0.85	268	2242	637	533	426	292	282
	0.9	0.9	190	1560	448	375	300	206	199
	0.95	0.95	100	831	237	199	159	109	105

granisetron compared to 3 mg in its efficacy for treating chemotherapy-induced nausea and vomiting, and thus the primary endpoint is complete control of a nausea/vomiting-related event. No significant differences in efficacy between the doses have been reported in [15], and none of the investigators taking part in the trial raise

Table 2
The NP sample size n^{NP} and the hNPB sample size n^{hNPB} for $\delta^*=0.05$, 0.1, and 0.15

Δ^*	π_1^*	π_0^*	n^{NP}	n^{hNPB}					
				$\gamma=0.05$	$\gamma=0.01$	$\gamma=0.001$	$\gamma=10^{-6}$	$\gamma=10^{-12}$	$\gamma=10^{-20}$
0.05	0.1	0.05	65	517	151	127	102	70	68
	0.15	0.1	102	839	241	202	161	111	107
	0.2	0.15	135	1104	318	266	213	147	141
	0.25	0.2	163	1339	384	322	258	177	171
	0.3	0.25	186	1544	441	369	295	203	195
	0.35	0.3	205	1692	484	406	324	223	215
	0.4	0.35	219	1810	517	433	346	238	230
	0.45	0.4	228	1897	540	452	362	248	240
	0.5	0.45	233	1927	551	461	369	253	244
	0.55	0.5	233	1927	551	461	369	253	244
	0.6	0.55	228	1897	540	452	362	248	240
	0.65	0.6	219	1810	517	433	346	238	230
	0.7	0.65	205	1692	484	406	324	223	215
	0.75	0.7	186	1544	441	369	295	203	195
	0.8	0.75	163	1339	384	322	258	177	171
	0.85	0.8	135	1104	318	266	213	147	141
	0.9	0.85	102	839	241	202	161	111	107
	0.95	0.9	65	517	151	127	102	70	68
0.1	0.15	0.05	46	384	110	92	73	51	49
	0.2	0.1	66	541	156	131	104	72	69
	0.25	0.15	83	686	196	165	132	91	87
	0.3	0.2	98	792	229	192	154	106	103
	0.35	0.25	110	887	257	216	173	119	115
	0.4	0.3	119	969	279	234	188	129	125
	0.45	0.35	125	1039	296	248	198	136	131
	0.5	0.4	129	1070	306	256	205	141	136
	0.55	0.45	131	1063	307	257	206	142	137
	0.6	0.5	129	1070	306	256	205	141	136
	0.65	0.55	125	1039	296	248	198	136	131
	0.7	0.6	119	969	279	234	188	129	125
	0.75	0.65	110	887	257	216	173	119	115
	0.8	0.7	98	792	229	192	154	106	103
	0.85	0.75	83	686	196	165	132	91	87
	0.9	0.8	66	541	156	131	104	72	69
	0.95	0.85	46	384	110	92	73	51	49
0.15	0.2	0.05	35	289	83	70	56	38	37
	0.25	0.1	47	382	111	93	74	51	49
	0.3	0.15	57	468	135	113	90	62	60
	0.35	0.2	66	523	153	129	103	71	69
	0.4	0.25	72	599	171	143	114	79	76
	0.45	0.3	77	642	183	153	123	84	81
	0.5	0.35	81	653	189	159	127	88	85
	0.55	0.4	82	685	195	163	131	90	87
	0.6	0.45	82	685	195	163	131	90	87
	0.65	0.5	81	653	189	159	127	88	85
	0.7	0.55	77	642	183	153	123	84	81
	0.75	0.6	72	599	171	143	114	79	76
	0.8	0.65	66	523	153	129	103	71	69
	0.85	0.7	57	468	135	113	90	62	60
	0.9	0.75	47	382	111	93	74	51	49
	0.95	0.8	35	289	83	70	56	38	37

Table 3
The cB sample size n^{cB} and the uB sample size n^{uB} for $\delta^* = -0.075, -0.05$, and 0

Δ^*	π_1^*	π_0^*	n^{cB}				n^{uB}			
			$\gamma=0.10$	$\gamma=0.05$	$\gamma=0.01$	$\gamma=0.001$	$\gamma=0.10$	$\gamma=0.05$	$\gamma=0.01$	$\gamma=0.001$
−0.075	0.025	0.1	1492	1194	1	1	>100,000	12,086	1	1
	0.075	0.15	2567	2054	1	1	>100,000	20,805	1	1
	0.125	0.2	3513	2810	1	1	>100,000	28,465	1	1
	0.175	0.25	4327	3462	1	1	>100,000	35,064	1	1
	0.225	0.3	5012	4010	1	1	>100,000	40,604	1	1
	0.275	0.35	5566	4453	1	1	>100,000	45,109	1	1
	0.325	0.4	5990	4793	1	1	>100,000	48,554	1	1
	0.375	0.45	6283	5027	1	1	>100,000	50,913	1	1
	0.425	0.5	6446	5157	1	1	>100,000	52,238	1	1
	0.475	0.55	6479	5184	1	1	>100,000	52,503	1	1
	0.525	0.6	6381	5105	1	1	>100,000	51,708	1	1
	0.575	0.65	6153	4923	1	1	>100,000	49,853	1	1
	0.625	0.7	5794	4636	1	1	>100,000	46,964	1	1
	0.675	0.75	5305	4245	1	1	>100,000	42,989	1	1
	0.725	0.8	4686	3749	1	1	>100,000	37,979	1	1
	0.775	0.85	3936	3149	1	1	>100,000	31,884	1	1
	0.825	0.9	3056	2445	1	1	>100,000	24,755	1	1
	0.875	0.95	2046	1637	1	1	>100,000	16,565	1	1
−0.50	0.05	0.1	449	359	1	1	>100,000	3631	1	1
	0.1	0.15	709	567	1	1	>100,000	5726	1	1
	0.15	0.2	937	750	1	1	>100,000	7581	1	1
	0.2	0.25	1133	907	1	1	>100,000	9171	1	1
	0.25	0.3	1296	1037	1	1	>100,000	10,496	1	1
	0.3	0.35	1427	1141	1	1	>100,000	11,556	1	1
	0.35	0.4	1524	1220	1	1	>100,000	12,351	1	1
	0.4	0.45	1590	1272	1	1	>100,000	12,881	1	1
	0.45	0.5	1622	1298	1	1	>100,000	13,146	1	1
	0.5	0.55	1622	1298	1	1	>100,000	13,146	1	1
	0.55	0.6	1590	1272	1	1	>100,000	12,881	1	1
	0.6	0.65	1524	1220	1	1	>100,000	12,351	1	1
	0.65	0.7	1427	1141	1	1	>100,000	11,556	1	1
	0.7	0.75	1296	1037	1	1	>100,000	10,496	1	1
	0.75	0.8	1133	907	1	1	>100,000	9171	1	1
	0.8	0.85	937	750	1	1	>100,000	7581	1	1
	0.85	0.9	709	567	1	1	>100,000	5726	1	1
	0.9	0.95	449	359	1	1	>100,000	3631	1	1
0	0.05	0.05	78	62	1	1	>100,000	624	1	1
	0.1	0.1	147	118	1	1	>100,000	1168	1	1
	0.15	0.15	208	167	1	1	>100,000	1683	1	1
	0.2	0.2	261	209	1	1	>100,000	2095	1	1
	0.25	0.25	306	245	1	1	>100,000	2453	1	1
	0.3	0.3	343	274	1	1	>100,000	2758	1	1
	0.35	0.35	371	297	1	1	>100,000	2983	1	1
	0.4	0.4	391	313	1	1	>100,000	3155	1	1
	0.45	0.45	404	323	1	1	>100,000	3248	1	1
	0.5	0.5	408	326	1	1	>100,000	3287	1	1
	0.55	0.55	404	323	1	1	>100,000	3248	1	1
	0.6	0.6	391	313	1	1	>100,000	3155	1	1
	0.65	0.65	371	297	1	1	>100,000	2983	1	1
	0.7	0.7	343	274	1	1	>100,000	2758	1	1
	0.75	0.75	306	245	1	1	>100,000	2453	1	1
	0.8	0.8	261	209	1	1	>100,000	2095	1	1
	0.85	0.85	208	167	1	1	>100,000	1683	1	1
	0.9	0.9	147	118	1	1	>100,000	1168	1	1
	0.95	0.95	78	62	1	1	>100,000	624	1	1

an objection against it, i.e., the investigators pre-specify $\pi_1^*=\pi_0^*=0.75$ and agree on $\delta^*=\pi_1^*-\pi_0^*=0$. However, their opinions do vary concerning the non-inferiority margin. In fact, the sample mean and variance of the margins specified by 12 investigators are 0.125 and 0.004773, respectively. We can set these values for δ^*, δ^*_{margin} and σ_0^2 in the prior: $\delta^*=0$, $\delta^*_{\text{margin}}=0.125$, and $\sigma_0^2=0.004773$. Let the sample sizes such that $Pr(C_\alpha^{\text{NP}}|\delta^*+\delta^*_{\text{margin}})$, $Pr(C_\alpha^{\text{NP}})$, $Pr(C_\alpha^{B}|\delta^*+\delta^*_{\text{margin}})$ and $Pr(C_\alpha^{B})$ exceed 0.9, be denoted by n^{NP}, n^{hNPB}, n^{cB} and n^{uB}, respectively. Assuming equal sample sizes per group, we can calculate n^{NP}, n^{hNPB}, n^{cB} and n^{uB}, and these turn out to be 253, 1256, 134 and 791, respectively. As shown in the calculated sample sizes, $n^{\text{hNPB}}=1256$, i.e., about five times the size of n^{NP}, is required to convince all 12 investigators of the significant result in the NP hypothesis testing. To convince them of the Bayesian significance, $n^{\text{uB}}=791$, about six times the size of $n^{\text{cB}}=134$, is required. Furthermore, it is noted that n^{cB} is about half as large as n^{hNPB}.

4.2. A numerical example: from a futility viewpoint

As mentioned in the discussion of Bayesian probabilities in Section 3, suppose that investigators arrive at a consensus on δ^* and δ^*_{margin}, and that we have the normal prior $\delta+\delta_{\text{margin}}\sim N[\delta^*+\delta^*_{\text{margin}},\sigma_0^2]$. Then we can take another view of the variation of δ^* and δ^*_{margin}, namely, we can consider a small value γ such that

$$\Pr\left(\delta+\delta_{\text{margin}}<0\right)=\gamma.$$

This implies that $\gamma=1-\Phi\ [(\delta^*+\delta^*_{\text{margin}})/\sigma_0]$ and means that the prior is centered on a pre-specified alternative hypothesis in (1), that is, H_A^* and with a low chance (say, 5%) that $\delta+\delta_{\text{margin}}$ is negative. Hence we can interpret γ as quantifying the 'futility' of a trial.

We consider the following values for the type I error, the difference, and the margin:, $\alpha=0.025$ $\delta^*=-0.075$, -0.05, 0, 0.05, 0.10, 0.15, and $\delta^*_{\text{margin}}=0.10$, and calculate the NP sample sizes (n^{NP}), the hNPB sample sizes (n^{hNPB}), the cB sample sizes (n^{cB}), and the uB sample sizes (n^{uB}). Again, n^{NP}, n^{hNPB}, n^{cB} and n^{uB}, respectively are such that $Pr(C_\alpha^{\text{NP}}|\delta^*+\delta^*_{\text{margin}})$, $Pr(C_\alpha^{\text{NP}})$, $Pr(C_\alpha^{B}|\delta^*+\delta^*_{\text{margin}})$ and $Pr(C_\alpha^{B})$ exceed 0.9. For calculation of sample sizes based on the hNPB probability $Pr(C_\alpha^{\text{hNPB}})$, γ is set at the following values: 0.05, 0.01, 0.001, 10^{-6}, 10^{-12}, and 10^{-20}. For calculation of sample sizes based on both of the cB probability $Pr(C_\alpha^{B}|\delta^*+\delta^*_{\text{margin}})$ and the uB probability $Pr(C_\alpha^{B})$, $\gamma=0.10$, 0.05, 0.01, 0.001.

In Tables 1 and 2, we present the NP sample sizes (n^{NP}) and the hNPB sample sizes (n^{hNPB}). In Tables 3 and 4, we present the cB sample sizes (n^{cB}) and the uB sample sizes (n^{uB}).

As shown in Tables 1, 2, 3 and 4, n^{NP}, n^{hNPB}, n^{cB} and n^{uB} decrease as δ^* becomes larger. For fixed δ^*, as π_0^* becomes larger, n^{NP}, n^{hNPB}, n^{cB} and n^{uB} increase, reaching their maximum around, $\pi_0^*=0.5$ and decrease symmetrically. As γ becomes smaller, n^{hNPB}, n^{cB}, and n^{uB} decreases remarkably. Tables 1 and 2 show that the hNPB sample size n^{hNBP} tends to the NP sample size n^{NP} as γ becomes quite small, but n^{hNPB} is always larger than n^{NP}. That is, the hNPB sample size allows us to take a conservative view of the NP sample size.

As shown in Tables 3 and 4, the cB sample size n^{cB} is small as compared with the corresponding NP sample size n^{NP} (shown in Tables 1 and 2). This result implies that, if we commit ourselves to a Bayesian design, analysis, interpretation and report, and explicitly incorporate the investigators' opinions or evidence from previous trials as a prior, we could considerably save on the sample size required to obtain Bayesian significance, as compared with the NP sample size. However, for quite small γ ($\gamma=0.01$, 0.001), any trial will necessarily lead to a significant Bayesian result, that is, $n^{\text{cB}}=1$. In this case, it is not realistic to apply the cB probability to the sample size calculation. Like the cB sample size, for quite small ($\gamma=0.01$, 0.001), the uB sample size turns out to be $n^{\text{uB}}=1$, whereas for large γ ($\gamma=0.10$), it is so large as to be impractical.

5. Discussion

In a non-inferiority trial, the margin, as well as the true difference to be detected between the two groups that has clinical meaning, or the alternative hypothesis in the context of Neyman–Pearson (NP) hypothesis testing, needs to be discussed and pre-specified, but it often happens that investigators taking part in the trial have different opinions as to what these pre-specified values should be. Therefore, for the purpose of taking this variation of the margin and/or the difference into consideration, in this article, we have presented the hybrid Neyman–Pearson–Bayesian (hNPB) probability, the conditionally Bayesian (cB) probability and the unconditionally Bayesian (uB) probability, and have considered a sample size calculation for a non-inferiority test of two independent binomial proportions in a clinical trial, based on these probabilities. The proposed approach here can be applied to other types of endpoint such as the hazard ratio and the odds ratio as well as values of a continuous variable since it is based on large-sample approximation and the conjugate normal analysis.

Table 4
The cB sample size n^{cB} and the uB sample size n^{uB} for $\delta^*=0.05$, 0.1, and 0.15

Δ^*	π_1^*	π_0^*	n^{cB}				n^{uB}			
			γ=0.10	γ=0.05	γ=0.01	γ=0.001	γ=0.10	γ=0.05	γ=0.01	γ=0.001
0.05	0.1	0.05	50	40	1	1	>100,000	385	1	1
	0.15	0.1	79	63	1	1	>100,000	628	1	1
	0.2	0.15	104	84	1	1	>100,000	826	1	1
	0.25	0.2	126	101	1	1	>100,000	1003	1	1
	0.3	0.25	144	116	1	1	>100,000	1158	1	1
	0.35	0.3	159	127	1	1	>100,000	1268	1	1
	0.4	0.35	170	136	1	1	>100,000	1356	1	1
	0.45	0.4	177	142	1	1	>100,000	1423	1	1
	0.5	0.45	181	144	1	1	>100,000	1444	1	1
	0.55	0.5	181	144	1	1	>100,000	1444	1	1
	0.6	0.55	177	142	1	1	>100,000	1423	1	1
	0.65	0.6	170	136	1	1	>100,000	1356	1	1
	0.7	0.65	159	127	1	1	>100,000	1268	1	1
	0.75	0.7	144	116	1	1	>100,000	1158	1	1
	0.8	0.75	126	101	1	1	>100,000	1003	1	1
	0.85	0.8	104	84	1	1	>100,000	826	1	1
	0.9	0.85	79	63	1	1	>100,000	628	1	1
	0.95	0.9	50	40	1	1	>100,000	385	1	1
0.1	0.15	0.05	36	29	1	1	>100,000	289	1	1
	0.2	0.1	51	41	1	1	>100,000	405	1	1
	0.25	0.15	65	52	1	1	>100,000	514	1	1
	0.3	0.2	76	60	1	1	>100,000	591	1	1
	0.35	0.25	85	68	1	1	>100,000	661	1	1
	0.4	0.3	92	74	1	1	>100,000	724	1	1
	0.45	0.35	97	78	1	1	>100,000	779	1	1
	0.5	0.4	100	80	1	1	>100,000	803	1	1
	0.55	0.45	101	81	1	1	>100,000	794	1	1
	0.6	0.5	100	80	1	1	>100,000	803	1	1
	0.65	0.55	97	78	1	1	>100,000	779	1	1
	0.7	0.6	92	74	1	1	>100,000	724	1	1
	0.75	0.65	85	68	1	1	>100,000	661	1	1
	0.8	0.7	76	60	1	1	>100,000	591	1	1
	0.85	0.75	65	52	1	1	>100,000	514	1	1
	0.9	0.8	51	41	1	1	>100,000	405	1	1
	0.95	0.85	36	29	1	1	>100,000	289	1	1
0.15	0.2	0.05	28	22	1	1	>100,000	245	1	1
	0.25	0.1	37	29	1	1	>100,000	322	1	1
	0.3	0.15	44	36	1	1	>100,000	396	1	1
	0.35	0.2	51	41	1	1	>100,000	439	1	1
	0.4	0.25	56	45	1	1	>100,000	508	1	1
	0.45	0.3	60	48	1	1	>100,000	545	1	1
	0.5	0.35	63	50	1	1	>100,000	550	1	1
	0.55	0.4	64	51	1	1	>100,000	582	1	1
	0.6	0.45	64	51	1	1	>100,000	582	1	1
	0.65	0.5	63	50	1	1	>100,000	550	1	1
	0.7	0.55	60	48	1	1	>100,000	545	1	1
	0.75	0.6	56	45	1	1	>100,000	508	1	1
	0.8	0.65	51	41	1	1	>100,000	439	1	1
	0.85	0.7	44	36	1	1	>100,000	396	1	1
	0.9	0.75	37	29	1	1	>100,000	322	1	1
	0.95	0.8	28	22	1	1	>100,000	245	1	1

We showed that the calculated sample sizes changed dramatically depending on the three probabilities. In other words, this result implies that the sample size required to answer a research question could be changed, depending on whether investigators' differing opinions on appropriate values for the margin and the difference are explicitly considered or not in the design of the non-inferiority clinical trial. Furthermore, the

selection of three probabilities has impacts upon the analysis, interpretation and report as well as the sample size calculations. So it would be worthwhile to try the sample size calculations based on all three probabilities. In particular, we might expect the hNPB and cB probabilities to be applied in some situations; for example, the former can be used to critically evaluate the NP power, and the latter enables us to reduce the NP sample size required for a trial if we intend to do a fully Bayesian analysis, interpretation and report. However, it is noted that both the cB and uB probabilities vary remarkably according to the prior variance or the futility. Such a tendency for the 'strength' of a prior is a well-known weak point in applying a Bayesian approach to the design. In other situations, for example, if a chief investigator won over all other investigators, we might advise him to use the largest of the sample sizes calculated based on the three probabilities.

However, it is obvious that the determination regarding which probability (or sample size) is best differs per situation or trial's goal. In principle we recommend using the three probabilities for the sample size calculation and making compensation for the NP power, rather than ranking these probabilities and the NP power. It will be helpful to perform some more simulations if the investigators select the best sample size.

In the proposed approach, we adopt a normal prior centered on the alternative hypothesis that is used in Neyman–Pearson hypothesis testing, and change the prior variance or futility. In particular, the elicitation of the prior in an example shown in Section 4.1 follows the methodology shown in [16]. As a matter of course, the appropriateness for the prior needs to be checked, and if it is not reliable, appropriate priors, e.g., a Beta prior and a log-normal prior, will need to be assumed for the margin and/or difference. This issue is beyond the scope of this article.

Furthermore, we have used an identical α for both of the Neyman–Pearson and Bayesian critical regions for the null hypothesis, but the choice of α needs to be discussed, based on the philosophy concerning statistical inference in each of the Neyman–Pearson and Bayesian schools.

Acknowledgements

The author would like to thank Byron Brown Jr., PhD– one of the foremost biostatisticians in the U.S.A. and former chair of the Department of Health Policy and Research, Stanford University– who unfortunately has passed away, for supervision and many valuable suggestions that led to the refinement of the original manuscript.

Appendix A. List of abbreviations and symbols

cB	Conditionally Bayesian.
hNPB	Hybrid Neyman–Pearson–Bayesian.
NP	Neyman–Pearson.
uB	Unconditionally Bayesian.
$Pr(\bullet)$	Probability.
X_{ij}	Binary response from the jth subject in the ith treatment group.
π_i	True proportion in the ith treatment group.
n_i	Sample size in the ith treatment group.
δ	Difference between the two true proportions.
δ^*_{margin}	Non-inferiority margin.
τ^2	Sample variance.
σ_0^2	Prior variance.
H_0	Null hypothesis.
H_A	Alternative hypothesis.
α	Type I error.
β	Type II error.
γ	Futility index.
C_α	Critical region of the null hypothesis at significance level α.
n^{NP}	Neyman–Pearson sample size.
n^{hNPB}	Hybrid Neyman–Pearson–Bayesian sample size.
n^{cB}	Conditionally Bayesian sample size.
n^{uB}	Unconditionally Bayesian sample size.

References

[1] Dudewicz EJ. Confidence intervals for power with special reference to medical trials. Aust J Stat 1972;14:211–6.
[2] Goldstein M. A Bayesian criterion for sample size. Ann Stat 1981;9:670–2.
[3] Gould AL. Sample sizes required for binominal trials when the true response rates are estimated. J Stat Plan Inference 1983;8: 51–8.
[4] Spiegelhalter DJ, Freedman LS. A predictive approach to selecting the size of a clinical trial, based on subjective clinical opinion. Stat Med 1986;5:1–13.
[5] Joseph L, Du Berger R, Bélisle P. Bayesian and mixed Bayesian/ likelihood criteria for sample size determination. Stat Med 1997;16:769–81.
[6] Katsis A, Toman B. Bayesian sample size calculations for binominal experiments. J Stat Plan Inference 1999;81:349–62.
[7] Stamey JD, Seaman JW, Young DM. Bayesian sample-size determination for inference on two binominal populations with no gold standard classifier. Stat Med 2005;24:2963–76.
[8] European Medicines Agency (EMEA). Guideline on the choice of the non-inferiority margin; 2005.
[9] Blackwelder WC. Proving the null hypothesis in clinical trials. Control Clin Trials 1982;3:345–53.

[10] Laster LL, Johnson MF. Non-inferiority trials: the 'at least as good as' criterion. Stat Med 2003;22:187–200.
[11] Hung HMJ, Wang S-J, Tsong Y, Lawrence J, O'Neil RT. Some fundamental issues with non-inferiority testing in active controlled trials. Stat Med 2003;22:213–25.
[12] Chow S-C, Shao J, Wang H. Sample size calculations in clinical research. New York: Marcel Dekker; 2003.
[13] Brown BW, Herson J, Atkinson N, Rozell ME. Projection from previous studies: a Bayesian and frequentist compromise. Control Clin Trials 1987;8:29–44.
[14] Crook JF, Good IJ. The powers and strengths of tests for multinominals and contingency tables. J Am Stat Assoc 1982;77: 793–802.
[15] Jordan K, Hinke A, Grothey A, Voigt W, Arnold D, Wolf HH, Schmoll H-J. A meta-analysis comparing the efficacy of four 5-HT3-receptor antagonists for acute chemotherapy-induced emesis. Support Care Cancer; 2006.
[16] Chaloner K. Elicitation of prior distributions. In: Berry DA, Stangl DK, editors. Bayesian Biostatistics. New York: Marcel Dekker; 1996. p. 141–56.