Vol.5, No.3, 364-374 (2013)
http://dx.doi.org/10.4236/health.2013.53049

Hypothesis testing by simulation of a medical study model using the expected net benefits criteria

Ismail Abbas[1*], Joan Rovira[2], Josep Casanovas[1]

[1]InLab FIB, Universitat Politècnica de Catalunya BarcelonaTech, Barcelona, Spain;
[*]Corresponding Author: ismail.i.abbas@gmail.com
[2]Universitat de Barcelona, Barcelona, Spain

Received 4 January 2013; revised 5 February 2013; accepted 18 February 2013

ABSTRACT

Introduction: This work investigates whether to conduct a medical study from the point of view of the expected net benefit taking into account statistical power, time and cost. The hypothesis of this paper is that the expected net benefit is equal to zero. Methods: Information were obtained from a pilot medical study that investigates the effects of two diagnostic modalities, magnetic resonance imaging (MRI) and computerized axial tomography scanner (CT), on patients with acute stroke. Statistical procedure was applied for planning and contrasting equivalence, non-inferiority and inequality hypotheses of the study for the effectiveness, health benefits and costs. A statistical simulation model was applied to test the hypothesis that conducting the study would or not result in overall net benefits. If the null hypothesis not rejected, no benefits would occurred and therefore the two arms-patterns of diagnostic and treatment are of equal net benefits. If the null hypothesis is rejected, net benefits would occur if patients are diagnosed with the more favourable diagnostic modality. Results: For any hypothesis design, the expected net benefits are in the range of 366 to 1796 per patient at 80% of statistical power if conducting the study. The power depends on the monetary value available for a unit of health improvement. Conclusion: The statistical simulations suggest that diagnosing patients with CT will provide more favourable health outcomes showing statistically significant expected net benefits in comparison with MRI.

Keywords: Statistics; Simulation; Hypothesis Testing; Expected Net Benefits

1. INTRODUCTION

As a general rule private and public researchers in medicine and health care, such as medical or pharmaceutical companies, centers for research and development, can be assumed to decide whether or not to carry out a study on the basis of the anticipated net benefits of health improvements obtained by a product or new medical indication. Health system organizations, link (or should link) their decisions to the expected future benefits in terms of patient's health improvements and cost savings. All parties, however, are likely to consider the cost of reaching the expected benefits, such as the cost of the study, and they are usually interested in shortening the study duration as much as possible, since this will mean that an effective diagnosis or treatment will be available sooner thus contributing to health improvement and increasing the time of marketing their product exclusively and, hence, the product's lifecycle benefits. Obtaining statistically significant evidence of the superiority, equivalence or non-inferiority of a given modality of diagnosis and treatment in relation to other modality increases their chances of their application in clinical practice.

Optimal clinical studies design will avoid unnecessary use of resources and increase benefits. Some knowledge and information, gathered from the execution of studies, can positively contribute toward improving the results of such trials. A clinical trial can be viewed as an uncertain experiment whose design depends on the problem being addressed. Some studies fail to answer the questions that need to be addressed [1-3] due to design issues, which mean that the resources used are wasted. Conventional statistical methods for designing clinical studies are widely used to make decisions, basically, on the sample size or power requirements used to test the hypothesis that there is a difference between two options. These models do not take into account the future costs and benefits of the decisions that for example might follow

study findings; instead they rely on arbitrary decision criteria. Other methods have been used with similar objectives [4-8], but the authors applied those assuming deterministic relationships and so only point results could be assessed, which do not take into account variability and uncertainty involved in the decisions.

Patel and Ankolekar [9] established a hybrid classical-Bayesian model at the design stage to determine the sample size while assuming that the data resulting from the study were analyzed based on the classical Neyman Pearson formula. A combination of the Bayesian and the classical approach was developed by Wang and Leung [10] and Shao *et al*. [11] to optimize the power of clinical studies. Their model combined the traditional statistical procedures with prior distribution of accepting a new treatment. Kikuchi *et al*. [12] developed a similar model which focused on estimating the subsequent distribution of treatment response variance. Jiang *et al*. [13] applied simulation to show which design was more efficient for the maximum tolerated dose in the treatment of a cancer disease. Huang *et al*. [14] used simulation to assess a better design for a parallel study design. The simulations showed that their suggested design conserved the same sample size, had better power, and assigned doses to patients more efficiently.

A variety of models have been published focusing on the cost and benefit but with different objectives. Baker and Heidenberger [15] applied the Monte Carlo simulation to estimate the sample size that maximizes expected health benefits based on expected discounted life years gained when the decision maker is constrained by the cost of the studies. Spiegelhalter and Best [16] developed Bayesian approaches to cost-effectiveness based on discrete-state Markov models with Monte Carlo simulation. They used data from a single clinical study and performed a probabilistic sensitivity analysis based on first and second order Monte Carlo simulation concepts that were discussed in Brigs and Sculpher [17], because they were uncertain about the cost-benefit estimation of their study design. As Willan and Pinto [18], we consider the monetary value of health benefits multiplying the quality of life obtained by society's willingness to pay for the benefits that would result from implementing the superior technology as the standard therapeutic option in a health system. Using information from a pilot medical study, our work addresses whether from the point of view of the expected net benefit (ENB) it would have made sense to carry out a clinical or medical trial assuming statistical significant result are reached at certain errors. In contrast to the previous methods, our model can be applied to discrete or continuous outcomes. Moreover, as we will show later on, it integrates other aspects of the study such statistics, e.g. testing hypothesis, economic and, e.g. the cost of search, the rate of enrolment, and managements, e.g. number of centers from which patients belong. Taking into account variability and uncertainty, the overall hypothesis of this work is that the expected net benefit that would result from a study is equal to zero or not comparing the effectiveness of two diagnostic images.

In Section 2, we first present a brief description of a conducted pilot clinical study results, and then we present a brief background of existing statistical procedures for testing hypotheses and at the same time to re-estimate the resulting power for a given design and sample size. Then, giving the information obtained from the pilot trial we construct the expected net benefits model that allow helping making decisions of the clinical study at design stage. The variability and uncertainty of the expected net benefits was quantified estimating the probability distributions. The model is executed many times to simulate simultaneously all hypotheses testing of the trial and of the expected net benefits. In Section 3, we present the results of the simulation model. In Section 4 we discuss the methods and the results of this application.

2. METHODS

2.1. The Clinical Study

Information were provided from a pilot randomized medical trial that compares the overall consequences in term of effectiveness and health benefit of two diagnostic options, CT and MRI option as the initial diagnostic test for patients with suspected acute stroke [19]. The alternative hypothesis was that patients initially diagnosed with MRI would show a 15% more favourable health outcome than those diagnosed with CT. The patients were recruited from those admitted to a hospital with suspected acute stroke. All patients who met the inclusion criteria were selected for the study. Statistical analysis of the severity of the patient's condition and other characteristics such as age and gender showed that there were no statistical significant differences between the two groups, therefore both groups can be considered similar based on a clinical judgment. They were diagnosed with either MRI or CT and followed-up for three months. After sixteen months of recruitment, 160 patients had been recruited of whom 130 were included in the study and 30 (0.19) were excluded. The Rankin scale assessment with categorical values (0, 1, 2, 3, 4, 5, and 6) was used to evaluate the health state of each patient at the admission to hospital. Each patient were subsequently followed-up for a period of three months, at the end her/his health state was evaluated again with the Rankin scale. Two outcomes were used to value the health states of patients: effectiveness and utility.

2.1.1. Effectiveness

In order to quantify the effectiveness at the end of the

study (three months after they have been diagnosed and treated, the categorical variable resulting from using the Rankin scales were converted into 3 health states for each patient as follows: 1) levels 0 - 2 were considered as an independent health state in which patients are assumed to have normal life, 2) scale levels 3 - 5 were considered as a dependent health state in which patients are assumed to need health care, and 3) scale level 6 corresponds to the death health state. The results of the 130 patients showed that 87 of them had been diagnosed with CT, from which a proportion of 0.506 (p_2) were in an independent health state; the remaining 43 had been diagnosed with MRI, from which a proportion of 0.429 (p_1) achieved an independent state.

The results of the effectiveness of the two diagnostic options are shown in **Table 1**.

The difference in effectiveness, 0.077, illustrates that using CT provided a more favourable outcome. The hypothesis is tested applying (1), the estimated value (0.82) is lower than the upper percentile of α, z_α (for $\alpha = 0.05$, $z_\alpha = 1.96$), and therefore there was no evidence to reject the null hypothesis that there is no difference in effectiveness between the two groups.

2.1.2. Health Benefits

The health benefits were quantified by converting the Rankin scale into an indicator of quality of life that ranges from −0.02 (patient is worse than dead state) to 1 (patient is alive and in very good state) of patients in two moments: at the time of admission to hospital (baseline) and after three months (end of study) follow-up. The conversion of the Ranking scale levels into a quality of life index was obtained according to the results of previous studies done by Fagan *et al*. [20] and Pinto *et al*. [21]. Scale level 0 was converted to 0.9, scale level 1 was converted to 0.68, scale level 2 was converted to 0.47, scale level 3 was converted to 0.2, scale level 4 was converted to 0.07, scale level 6 was converted to 0 and scale level 5 was converted to −0.02. Then the within group health benefits in utility was calculated as the difference between the two measures of utilities (value of quality of life at the end minus the value at the diagnostic state) within each group of diagnostic, MRI and CT (see **Table 2**).

2.1.3. Costs

The costs of the trial were: 1) the treatments costs that could be avoided if all patients were moved to the less costly diagnostic option, that is, the difference between the cost of treatment under CT and MRI (C_{TC}-C_{MRI}); these costs were estimated from the data collected in the original study (see **Table 2**), and 2) the cost of completing the study until significant results were found. This is estimated as the additional cost of the research, considered to be 200 monetary units per patient.

Table 1. Observed results on effectiveness after diagnosing patients with MRI or CT.

	MRI ($n_1 = 43$)	CT ($n_2 = 87$)
Independence (Rankin < = 2)	0.429	0.506
Dependence (Rankin 3 - 5)	0.429	0.412
Death	0.142	0.082

Table 2. Observed results on cost and benefits of treatment after diagnosing patients with MRI or CT.

	MRI ($n_1 = 43$)	CT ($n_2 = 87$)
	Mean (SD)	Mean (SD)
Costs of treatment	6184 (2413)	6129 (2546)
Health benefits	0.1123(0.2815)	0.1638 (0.2286)

2.1.4. Results

The pilot trial's results were analyzed before the planned sample was recruited and the outcomes were evaluated. It happened because the initial results indicate that effectiveness associated with CT is higher than MRI, the time results obtained from the recruited patients so far suggested that further investigation's expected duration to reach statistical significance in effectiveness exceeds the time originally planned for the whole research. Based on this initial results, there was no change in the actual practice (patients continued be diagnosed with MRI or CT), no additional research expenditures were needed but there will be no benefits from adopting the best option taking into account their effectiveness, health benefits and cost saving. However, the money invested in the study was somehow wasted, as no useful information was used to provide evidence on the best option. The decision made may not be totally accurate since not rejecting the null difference is not evidence of null difference. Moreover, the non-significant result might indicate that the inequality design of the trial was not appropriate, and might have been designed as equivalence or non-inferiority designs in effectiveness as primary outcome and gathering information on health benefits.

2.2. The Model

2.2.1. The Statistical Process for Inequality Design

Suppose that a protocol for a double blind randomized clinical trial is designed in order to compare the effects of two diagnostic products (two diagnostic images), MRI and CT on patients with a given disease. In order to test on whether or not the effectiveness of the two products is or not equal, in the case of comparing two proportions, two-sided hypothesis $H_0 : \pi_2 - \pi_1 = 0$ versus

$H_1 : \pi_2 - \pi_1 \neq 0$ are planned assuming α and β. A sample of patients, N, recruited from the population were assigned randomly either to MRI or to CT, in which the probabilities that an individual has a successful outcome is designated as p_1 and p_2, respectively. As soon as the data are available from the trial for evaluation, a test statistic can be applied to compare two proportions of success. As in inequality design, using the test statistic with the information of the pilot trial we would obtain the observed value of the statistic as follow:

$$\hat{z} = \frac{p_2 - p_1}{\mathrm{SE}(p_2 - p_1)} \to N(0,1) \qquad (1)$$

where,

$$\mathrm{SE}(p_2 - p_1) = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

If the resulting value of $\hat{z}$ is higher than $z_{1-\alpha/2}$ (=1.96, $\alpha = 0.05$) or p-value is smaller than α, the null hypothesis is rejected with the conclusion that CT provides more favourable outcome. If the resulting value of $\hat{z}$ is lower than $-z_{1-\alpha/2}$ (= −1.96, $\alpha = 0.05$) or p-value is smaller than α, the null hypothesis is rejected with the conclusion that MRI provides more favourable outcome. If the resulting value of $\hat{z}$ is lower than $-z_{1-\alpha/2}$ and higher than $-z_{1-\alpha/2}$ (1.96, $\alpha = 0.05$) or p-value is higher than α, the null hypothesis is not rejected with the conclusion that there is not enough evidence that one of the associated diagnostic provides more favourable outcome.

After providing the success probabilities obtained from the pilot trial, the sample size of n_1 and n_2 that are necessary for the new trial can be calculated using formula 2 taking into account α, β or power.

$$\begin{aligned}\text{Power} &= 1 - \text{prob}\left[Z \le \frac{z_{1-\alpha/2}\mathrm{SE}(p_2 - p_1) - (p_2 - p_1)}{\mathrm{SE}(p_2 - p_1)} \right] \\ &= 1 - \text{prob}\left[Z \le z_{1-\alpha/2} - |\hat{z}| \right]\end{aligned} \qquad (2)$$

Suppose that $n_1 = 650$, $n_2 = 650$ (assuming $n_1 = n_2$), $p_1 = 0.429$, $p_2 = 0.506$ (obtained from the Trial). Applying (1) $\hat{z} = 2.8$ (p-value = prob ($Z > \hat{z}$) = 0.0026) indicates that there is enough evidence to reject the null hypothesis at significant level of 0.0026. The power of the trial, $1 - \beta$, is expected to be as initially planned (e.g. 80%):

$$\text{Power} = 1 - \text{prob}\left[Z \le 1.96 - \frac{(p_2 - p_1)}{\mathrm{SE}(p_2 - p_1)} \right] = 1 - \text{prob}\left[Z \le 1.96 - \frac{(0.506 - 0.429)}{\sqrt{\frac{0.506(1-0.506)}{650} + \frac{0.429(1-0.429)}{650}}} \right]$$

$$= 1 - \text{prob}\left[Z \le 1.96 - |2.8| \right] = 1 - \text{prob}\left[z \le -0.84 \right] = 0.8.$$

Now, suppose that ($n_1 = 490$) <> ($n_2 = 980$) and applying (2), the expected power of 0.8 would be reached assuming unequal allocation as follow:

$$\text{Power} = 1 - \text{prob}\left[Z \le 1.96 - \frac{(p_2 - p_1)}{\mathrm{SE}(p_2 - p_1)} \right] = 1 - \text{prob}\left[Z \le 1.96 - \frac{(0.506 - 0.429)}{\sqrt{\frac{0.506(1-0.506)}{980} + \frac{0429(1-0.429)}{490}}} \right]$$

$$= 1 - \text{prob}\left[Z \le 1.96 - |2.8| \right] = 1 - \text{prob}\left[z \le -0.84 \right] = 0.8.$$

2.2.2. The Statistical Process for Equivalence Design

Suppose that a protocol for a double blind randomized clinical trial is designed in order to compare the effects of two medical products (two diagnostic images), MRI and CT on patients with a given disease. In order to test on whether or not the effectiveness of the two products is equivalent, in the case of comparing two proportions, the equivalence hypotheses $H_0 : \pi_2 - \pi_1 > |\Delta|$ and $H_1 : \pi_2 - \pi_1 \le |\Delta|$ are planned. A sample of patients, N, recruited from the population were assigned randomly either to MRI or to CT, in which the probabilities that an individual has a successful outcome is designated as p_1 and p_2, respectively. As soon as the data are available from the trial for evaluation, a statistical test can be applied to compare two proportions of success. Thus, using the test statistic with the information of the pilot trial we would obtain the observed value of the statistic as follow:

$$\hat{z} = \frac{\Delta - (p_2 - p_1)}{\mathrm{SE}(p_2 - p_1)} = \frac{0.15 - 0.077}{0.026} = 2.8$$

Since $\hat{z} > z_{1-\alpha/2}$, or p-value is smaller than α the hypothesis of inequivalence is rejected. Applying the two

sided confidence interval for the observed difference assuming acceptable margin of difference Δ, if the re- sulting two limits of the confidence interval lie within the range [−Δ, +Δ], the hypothesis of inequivalence is rejected, and otherwise, is accepted.

Suppose that $n_1 = 730$, $n_2 = 730$ ($n_1 = n_2$), $p_1 = 0.429$, $p_2 = 0.506$, $\Delta = 0.15$, applying (2) the power is expected to be:

$$\text{Power} = 1 - \text{prob}\left[Z \le 1.96 - \frac{\Delta - (p_2 - p_1)}{\text{SE}(p_2 - p_1)}\right] = 1 - \text{prob}\left[Z \le 1.96 - \frac{0.15 - 0.077}{\sqrt{\frac{0.506(1-0.506)}{730} + \frac{0.429(1-0.429)}{730}}}\right]$$

$$= 1 - \text{prob}\left[Z \le 1.96 - |2.8|\right] = 1 - \text{prob}\left[z \le -0.84\right] = 0.8.$$

Now, suppose ($n_1 = 550$) <> ($n_2 = 1100$) and applying (2), the expected power of 0.8 would be reached assuming unequal allocation.

$$\text{Power} = 1 - \text{prob}\left[Z \le 1.96 - \frac{\Delta - (p_2 - p_1)}{\text{SE}(p_2 - p_1)}\right] = 1 - \text{prob}\left[Z \le 1.96 - \frac{0.15 - (0.506 - 0.429)}{\sqrt{\frac{0.506(1-0.506)}{1100} + \frac{0.429(1-0.429)}{550}}}\right]$$

$$= 1 - \text{prob}\left[Z \le 1.96 - |2.8|\right] = 1 - \text{prob}\left[z \le -0.84\right] = 0.8.$$

2.2.3. The Statistical Process for Non-Inferiority Design

Suppose that a protocol for a double blind randomized clinical trial is designed in order to compare the effects of two medical products (two diagnostic images), MRI and CT on patients with a given disease. In order to test on whether or not CT would provide worse outcome than MRI, in the case of comparing two proportions, the one sided non-inferiority hypotheses $H_0 : \pi_2 - \pi_1 < -\Delta$ and $H_1 : \pi_2 - \pi_1 \ge -\Delta$ are planned. A sample of patients, N, recruited from the population was assigned randomly either to MRI or to CT, in which the probabilities that an individual has a successful outcome is designated as p_1 and p_2, respectively. As soon as the data are available from the trial for evaluation, a test statistic can be applied to compare two proportions of success. Using formula (2), we will be able to calculate the sample size or power for equal and unequal allocation.

Suppose that $n_1 = 578$, $n_2 = 578$ ($n_1 = n_2$), $p_1 = 0.429$, $p_2 = 0.506$, $\Delta = 0.15$, Applying (1) assuming non-inferiority hypothesis.

$$\hat{z} = \frac{\Delta - (p_2 - p_1)}{\text{SE}(p_2 - p_1)} = \frac{0.15 - 0.077}{0.026} = 2.8. \text{ Since } \hat{z} > z_{1-\alpha},$$

the hypothesis of inferiority is rejected. Using the two sided confidence interval for the observed difference assuming acceptable margin of difference Δ. If the resulting low limit of the confidence interval is higher than the negative margin of [−Δ, +Δ], the hypothesis of the inferiority is rejected, and otherwise, is accepted.

Applying (2) the power is expected to be

$$\text{Power} = 1 - \text{prob}\left[Z \le 1.645 - \frac{\Delta - (p_2 - p_1)}{SE(p_2 - p_1)}\right] = 1 - \text{prob}\left[Z \le 1.645 - \frac{0.15 - 0.077}{\sqrt{\frac{0.506(1-0.506)}{578} + \frac{0.429(1-0.429)}{578}}}\right]$$

$$= 1 - \text{prob}\left[Z \le 1.645 - |2.5|\right] = 1 - \text{prob}\left[z \le -0.84\right] = 0.8.$$

Now, suppose ($n_1 = 420$) <> ($n_2 = 860$) and applying (2), the expected power of 0.8 would be reached assuming unequal allocation.

$$\text{Power} = 1 - \text{prob}\left[Z \le 1.645 - \frac{\Delta - (p_2 - p_1)}{\text{SE}(p_2 - p_1)}\right] = 1 - \text{prob}\left[Z \le 1.645 - \frac{0.15 - (0.506 - 0.429)}{\sqrt{\frac{0.506(1-0.506)}{860} + \frac{0.429(1-0.429)}{430}}}\right]$$

$$= 1 - \text{prob}\left[Z \le 1.645 - |2.5|\right] = 1 - \text{prob}\left[z \le -0.84\right] = 0.8.$$

2.2.4. The Expected Net Benefit

The analysis addresses whether from a health system's perspective it would make sense to conduct a clinical study, considering the expected net benefit (ENB) for the health system. We assume that the cost for the health system would consist of the cost of completing the study, while the benefits are defined as the health benefits and cost savings that would be accrued, assuming a statistically significant result was reached. The analysis, however, does not consider the potential administrative costs of research or any cost related to implementing the decision of changing the diagnostic patterns in line with the study recommendations. Moreover, the CT is a dominant option from a cost-effectiveness perspective, it is less costly (see **Table 2**) and has a better health outcome than MRI (see **Tables 1** and **2**). However, this is not confirmative results because the pilot study did not provide evidence statistically significant of superiority of any of the two diagnostic images.

The ENB hypothesis also requires an assessment in monetary terms of the future health benefits likely to be obtained by using the superior diagnostic modality on all patients, which in the context of this work the health benefits were quantified by estimating the gain in utility as the difference between the two measures of utilities corresponding to CT and MRI. The resulting difference was multiplied by a cost, which is the amount of willing to pay per unit of quality of life. The amount is approximately 30,000 monetary units that have often been assumed as a reasonable cost-effectiveness threshold for accepting a new medical device in the health system [21]. Taking into account these considerations, two-sided hypothesis $H_0 : ENB = 0$ and $H_1 : ENB \neq 0$ are planned to test whether the ENB is or not zero assuming two types of errors α and β. A sample of patients, N ($n_1 + n_2$), recruited from the population were assigned randomly either to MRI or to CT, in which the probabilities that an individual has a successful outcome is designated as p_1 and p_2, respectively. As soon as the data are available from the trial for evaluation, testing this hypothesis, we estimate the ENB that would result from integrating the power estimated with time, costs and health benefits. If ENB is lower than $z_{1-\alpha/2}$ and higher than $-z_{1-\alpha/2}$ ($z_{1-\alpha/2}$ = 1.96, α = 0.05) or p-value is higher than α, the null hypothesis is not rejected with the conclusion that there is not enough evidence of net benefits, otherwise, the hypothesis is rejected and we conclude that it there is statistical evidence of net benefits using MRI or CT diagnostic modalities. If the ENB is positive then the CT provides net benefits, otherwise, MRI. The ENB test is based on the following model:

$$\begin{aligned}\text{ENB} &= \text{power}\left(w\left(U_{\text{CT}} - U_{\text{MRI}}\right) - \left(C_{\text{CT}} - C_{\text{MRI}}\right)\right) \\ &- \left[1/\lambda \times \left(n_1 + n_2\right) / \left(1 - e^{-\lambda S}\right) S + t\right] \text{Cost} / \left(n_1 + n_2\right)\end{aligned} \quad (3)$$

where, w is the monetary value of a unit improvement of utility, U_{CT} and U_{MRI} are the utilities values of CT and MRI option respectively, C_{CT} and C_{MRI} are the cost of diagnosing a patient with CT or MRI option respectively, lambda is the rate of enrolled patients. The total time of the trial as a function of the number of patients, the number of centers and the rate of enrolment, that is $\left[\frac{1}{\lambda} \times \frac{n_1 + n_2}{\left(1 - e^{-\lambda S}\right) S} + t\right]$. Thus, the total cost of research is calculated as the multiplication of the total time by research cost per patient.

To put a simple example, suppose that estimated power = 0.12, $U_{\text{CT}} - U_{\text{MRI}} = 0.05$, $C_{\text{CT}} - C_{\text{MRI}} = 100$, for simplicity let

$$\frac{\left[\frac{1}{\lambda} \times \frac{n_1 + n_2}{\left(1 - e^{-\lambda S}\right) S} + t\right] \text{Cost}}{n_1 + n_2} = 200 \quad \text{(monetary currency)}$$

Per patient, w = 30,000, then ENB = −32. The ENB is negative which means that there will be a loss if the CT is applied. However, if 0.8 of power is reached, the ENB is positive (926 monetary unit) which means that a net benefits are expected if patients are diagnosed with CT for any hypotheses design.

In order to test the hypothesis planned that are or not a net benefit would derived from the study, the probability distribution of ENB is needed. Simulation will allow carrying out the trial as if it comes from real world experiments, given the statistical design and extracting the costs and health benefits from their respective distributions, and then to estimate probability curve of positive expected net benefits.

2.3. Simulations

Following the construction of the model as shown in **Figure 1**, we executed it to simulate results of the trial. The inputs of the model are effectiveness, the number of enrolled per month, time of follow-up, cost of research per month, health benefits, cost of each arm of the study including diagnosis and treatment. These inputs were used through the simulation according to the patients flow shown in **Figure 1** in order to simulate the trial and estimate the ENB, and hence testing the hypotheses of our work.

The states of the simulation model are:

1) Population: the population from which a sample of patients will be selected.

2) Arrivals: patients will be admitted to the study site such as labs, hospital or other research units.

3) Inclusion: patients will be included at random according to criteria of inclusion that is modeled in probabilistic pattern.

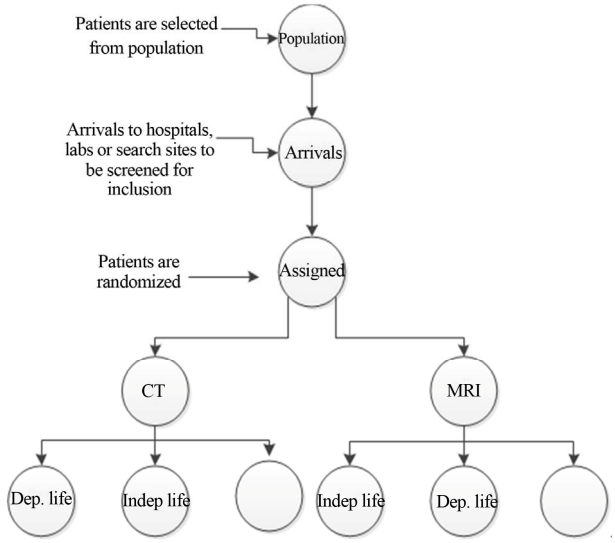

Figure 1. The simulation of the patients flow, **Dep. life**: dependent health state where patients assumed to need health care, **Indep. life**: independent health state where patients assumed to have normal life without need health care.

4) Diagnosis test: the included patients in the trial are assigned randomly to one of the two diagnoses, CT or MRI.

5) Follow-up states: following the diagnostic, the health state of each patient will be valued and classified into one of three health states three months later: independent health state (indep. life), dependent health state (dep. life) and death state (empty). Those patients who died before the end of follow-up are excluded from the analysis since there were no data on the cause of death.

The simulation is run in three steps:

1) Individual simulation (one patient)

The model starts when a patient from the population arrives to the study centre according to an exponential arrivals time. It is screened for subsequent inclusion or exclusion. If the patient is included then s/he is randomly assigned to either MRI or to CT. After that, the patient is randomly moved to one of the three health states (indep. life, dep. life, empty). All patient movements are controlled by their respective probabilities and the random probabilities generated from a uniform distribution between 0 and 1.

2) Sample of patients simulation

In order to estimate variability between patients, the previous individual simulation was repeated for a given number of patients that are selected from the population to participate in the trial. When a number of patients are included in MRI and in CT (n_1 *and* n_2), the simulation is stopped and the number of excluded patients, the number of patients assigned to MRI and CT, proportion of patients in each model state depending on MRI and for CT, are obtained. The cost and benefit for this patient were randomly chosen assuming normal distribution of cost and health benefits. So that the resulting difference in effectiveness, utilities and cost of treatments were calculated.

3) Replications

In order to estimate variability between samples, the second step was repeated 10,000 times, applying 10,000 different sequences of random numbers. Consequently, 10,000 different results were generated. So that ten thousands differences in effectiveness were compared according to the corresponding hypothesis testing. Differences in mean cost of treatment and in mean health benefits between MRI and CT are also obtained for each execution.

The hypothesis contrast, power and expected net benefits are processed as follows:

a) For each replication, the statistical value for the inequality hypotheses design was estimated dividing the observed mean differences in effectiveness by the expected standard error. If the calculated statistic is higher than the absolute value of $z_{1-\alpha/2}$, or the p-value is smaller than α, we rejected the null hypothesis, otherwise, we accepted it.

b) For each replication, the statistical value for the equivalence hypotheses design was estimated assuming $\Delta = 0.15$ is an acceptable margin of difference for equivalence design. If the calculated statistic lies outside the area of $[-z_{1-\alpha/2}, +z_{1-\alpha/2}]$, we reject the null hypothesis of non-equivalence, otherwise we accept it.

c) For each replication, the statistical value for non-inferiority design was assuming also $\Delta = 0.15$ is an acceptable margin of difference for equivalence hypotheses. If the calculated statistic lies outside the area of $[-z_{1-\alpha/2}, +z_{1-\alpha/2}]$, we rejected the null hypothesis of inferiority, otherwise we accepted it.

d) In all these hypothesis design, the power is calculated as the number of rejecting the null hypothesis divided by the total number of replications (in this work they are 10,000 ENBs). Once the expected power of the trial is obtained, the expected net benefits are calculated for each replication. The 10,000 ENBs obtained allow constructing the empirical distribution of the mean, and testing the hypothesis of this work, subsequently, the power of ENB is calculated as the number of rejecting the that the ENB = 0 divided by 10,000.

3. RESULTS

Table 3 shows the results of simulations for the three studies hypothesis under the same conditions. It is clear that variability (SE) of effectiveness, benefits and costs; and their 95% confidence interval ranges decreases as the sample size increases. There are statistical significant differences comparing the effectiveness and benefits of MRI and CT ($p < 0.05$ at 0.8 of power). There are no differences in treatment cost following diagnosing pa-

Table 3. Simulation results for equivalence, inequality and non-inferiority hypotheses design.

		CT				MRI				Differences: CT-MRI					Expected net benefits				
Hypothesis Design		**Mean**	**SE**	**IC95L**	**IC95U**	**Mean**	**SE**	**IC95L**	**IC95U**	**Mean**	**SE**	**Test**	**p-value**	**power**	**Mean**	**SE**	**Test-enb**	**p-value**	**power**
Equivalence	**Effectiveness**	0.5059	0.0183	0.4701	0.5417	0.4289	0.0185	0.3927	0.4652	0.0769	0.0258	2,8377	0,0023	0,8096	1098.29	342.62	3.2055	0.0007	0.8945
1460 (730-730)	**Benefits**	0.1638	0.0085	0.1471	0.1805	0.1122	0.0103	0.0919	0.1324	0.0516	0.0134	3,8458	0,0001	0,9711					
	Costs	6129.31	94.20	5944.67	6313.95	6183.93	89.26	6008.98	6358.87	-54.6188	129.6885	-0.4212	0.6632	0.0654					
Inequality	**Effectiveness**	0.5065	0.0196	0.4681	0.5448	0.4291	0.0193	0.3912	0.4670	0.0773	0.0277	2.7923	0.0026	0.7941	1074.98	356.08	3.0189	0.0013	0.8549
1300 (650-650)	**Benefits**	0.1638	0.0089	0.1463	0.1813	0.1121	0.0110	0.0906	0.1337	0.0517	0.0142	3.6347	0.0001	0.9516					
	Costs	6129.94	99.74	5934.45	6325.44	6184.34	94.56	5998.99	6369.68	-54.3910	138.0996	-0.3939	0.6532	0.0630					
Non-inferiority	**Effectiveness**	0.5062	0.0209	0.4651	0.5472	0.4288	0.0208	0.3880	0.4696	0.0774	0.0295	2.4611	0.0069	0.7938	1068.56	373.80	2.8586	0.0021	0.8165
1156 (578-578)	**Benefits**	0.1639	0.0149	0.1346	0.1931	0.1123	0.0095	0.0937	0.1310	0.0515	0.0149	3.4521	0.0003	0.9354					
	Costs	6130.88	107.35	5920.46	6341.29	6183.22	100.91	5985.43	6381.01	-52.3427	147.5638	-0.3547	0.6386	0.0543					

tients with MRI or CT ($p > 0.05$). The null hypothesis of the expected net benefits is rejected since for any design the statistical significance level lies below 5% at a power higher than 80%. Moreover, the probability of rejecting the null hypothesis increases as the amount of monetary amount of utilities increases, in which case diagnosing patients with CT would be beneficial for the patient and the hospital that would apply it (**Figure 2**). Giving the results of the simulated hypotheses, the equivalence hypothesis is more efficient because the expected power of the expected net benefits is the highest.

Simulations results are similar to the analytical ones, confirming the accuracy of power and the expected net benefits estimated by the statistical procedure. Moreover the parameters estimated with simulation that we know they can be also available with conventional statistical methods, however, the standard error of ENB and standard deviation (within and between groups variability) are estimated, and thus the accumulated probability distribution of ENB could be obtained. In the case of hypothesis testing, there is a significant evidence that the expected net benefits are higher than zero since that the resulting value of observed statistical test is higher than the reference one $t_{1-\alpha/2}$ (for $\alpha = 0.05$, $t_{0.975}$) at an estimated power of more than 80%. Hence, the probability of benefits moving patients to the CT or MRI diagnostic modality could be taken into account for different trial designs.

4. DISCUSSION AND CONCLUSION

Medicine and health care are of our primary concerns. The overall objective of our group is to improve the health of patients. More specifically, with the help of scientists, technologists, developers and industry, our aim is to help private and public institution in improving disease diagnosis and to predict the most optimal treatment pattern for patients. Conducting a research study, in first instance, required collecting a sample of patients that are the main objects of research. Adequate design and sample size of clinical or medical studies would yield to obtain sufficient evidence from the expected results.

The overall expected net benefits of a study is an issue, in this work, the statistical and the simulation processes were applied to plan to a new trial taking into account data from conducted pilot trial. The simulation model addresses the issue of whether any benefit would be derived from the medical trial carrying out a larger one based on the pilot data in order to test the hypothesis. We tested some possible hypotheses designs such as inequality, equivalence, non-inferiority. In addition, simulations provided us with information on variability of complex expected net benefits model, allowing for constructing its probability distributions. These results were obtained in two steps. First, we modeled the results of the pilot study; we extend the study to larger sample sizes. Second, we repeatedly tested the hypothesis of the simulated study for a given sample size, to assess whether or not to reject the null hypothesis of any design, inequality, equivalence and non-inferiority estimating the empirical probability distributions of the expected net benefits.

From statistical perspective, with the simulation we were able to provide information on the expected net's benefits variability. Concretely, testing the hypothesis of

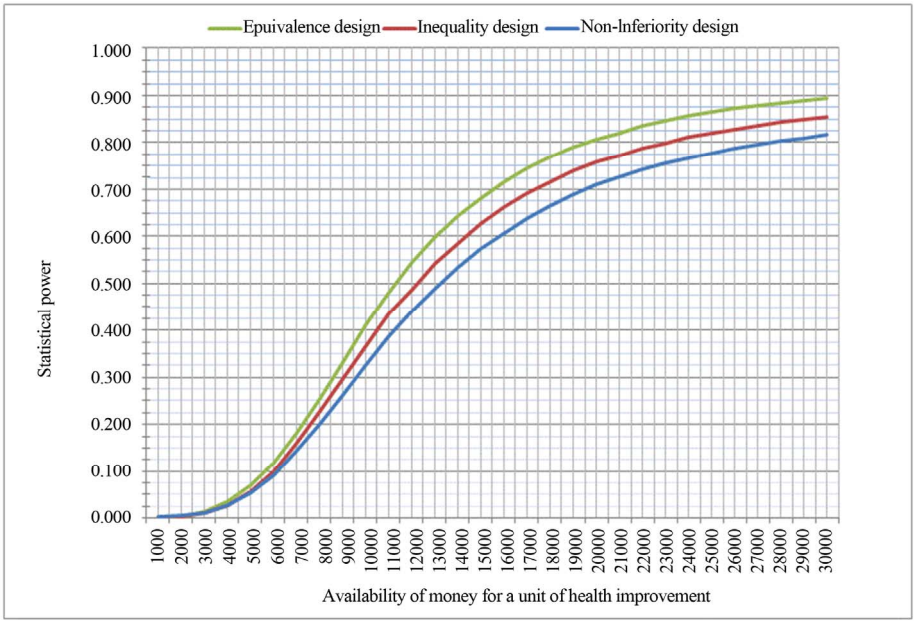

Figure 2. Statistical power of expected net benefits depending on availability of money for a unit of health improvement.

the expected net benefits needed that variability to be estimated, marriaging variability from a range of different probability distributions, assuming independency. These variables are, the enrollment time per patient that follows the inverse of a probability exponential distribution and the mean time distribution of inclusion that follows a normal distribution, the number of patients that reach the primary outcome of the trial that follows a binomial distribution and the mean that follows a normal distribution, the differences in utility within and between samples that were extracted from two normal distributions, the overall cost per patient of the treatment that assumed to follow a normal distribution, the overall differences in cost of treatment between the two studied arms of the study that were extracted from a normal distribution. Marriaging all possible values of each combination; and uploading them to the expected net benefits function generate the empirical distribution of ENB. Having estimated its mean and standard error ended in the construction of the statistic that used to test our hypothesis.

From probabilistic perspective, with simulation we are able to apply this procedure to a range of monetary amounts that a hospital would receive from the health system for a unit of a patient health improvement, or that the health system would pay for a unit of health benefit. Assuming that the trial has shown statistical significant assuming α and β, a probability curve can be generated such as the one produced with this study. Although the approach in this paper applied to one site trial; however, it can be applied to a multisite trial assuming robustness or sensitivity on one or more variables. In the robustness case, the underlying assumption is that the number of sites does not change the effect on health such as the effectiveness or utilities, but might increase or decrease the cost per patient. Under the sensitivity assumption, including more sites will produce different health effects and might change other variables, therefore, the procedure will marriage them within the overall health effects.

From generality perspective, the simulation model is also can be applied taking into account any hypothesis testing. For example, trials that study a continuous variables and employ t-test for two means: trials that study dependency between several variables applying chi-square, studies that compare two variances to confirm homogeneity or heterogeneity, analysis that study correlation or dependency between two variables, or trials that test statistical association between a given continuous response and a given factor.

In conclusion, the clinical trial was designed with the hypothesis that MRI is more costly than CT but would show better health outcomes. According to the results of the pilot trial, the CT is a dominant option from a cost-effectiveness perspective; it is less costly and has a better health outcome than MRI. However, the pilot study did not provide evidence statistically significant showing the favorability of any of the two diagnostic images. The simulated trial for larger sample size shows that for any hypothesis design the MRI and CT will generate the same cost ($p > 0.05$) but the CT provides better health benefits ($p < 0.05$). Furthermore, we have shown that the expected net benefits per patient will be higher than zero statistically significant if CT is used for diagnosing patients with suspected acute stroke. Therefore, we can conclude that CT works better than MRI.

Finally, we would like to communicate that our research is being applied to research studies in order to search for optimal trial's parameters design that will maximize the expected net benefits resulting from testing the hypothesis assuming two types errors of a (the probability of rejecting a hypothesis when it is true), and β (the probability of rejecting alternative hypothesis when it is true) or the power (the probability of accepting a hypothesis when it is true). In our on-going research project, it is trying to apply the optimization approach to aneurysms data.

5. ACKNOWLEDGEMENTS

Lots of thanks to the reviewer for his grateful comments.

REFERENCES

[1] Peck, C.C. (1997) Drug development: Improving the process. *Food and Drug Law Journal*, **52**, 163-167.

[2] Waller, D., Peake, M.D. and Stephens, R.J. (2004) Chemotherapy for patients with non-small cell lung cancer: The surgical setting of the Big Lung Trial. *European Journal of Cardio-Thoracic Surgery*, **26**, 173-182. doi:10.1016/j.ejcts.2004.03.041

[3] Scagliotti, G.V., Fossati, R., Torri, V., Crinò, L., Giaccone, G., Silvano, G., Martelli, M., Clerici, M., Cognetti, F. and Tonato, M. (2003) Randomized study of adjuvant chemotherapy for completely resected stage I, II, or III: A non-small-cell lung cancer. *JNCI Journal of the National Cancer Institute*, **95**, 1453-1461. doi:10.1093/jnci/djg059

[4] Girling, A.J., Lilford, R.J., Braunholtz, D.A. and Gillett, W.R. (2007) Sample-size calculations for studies that inform individual treatment decisions: A "true-choice" approach. *Clinical Trials*, **4**, 15-24. doi:10.1177/1740774506075872

[5] Yin, K., Choudhary, P.K., Varghese, D. and Goodman, S.R. (2007) A Bayesian approach for sample size determination in method comparison studies. *Statistics in Medicine*, **27**, 2273-2289. doi:10.1002/sim.3124

[6] Howard, G. (2007) Nonconventional clinical studies designs: Approaches to provide more precise estimates of treatment effects with a smaller sample size, but at a cost. *Stroke*, **38**, 804-808. doi:10.1161/01.STR.0000252679.07927.e5

[7] Berry, D.A. (2005) Introduction to Bayesian methods III: Use and interpretation of Bayesian tools in design and analysis. *Clinical Trials*, **2**, 295-300. doi:10.1191/1740774505cn100oa

[8] Tan, S.B. and Machin, D. (2002) Bayesian two-stage designs for phase II clinical studies. *Statistics in Medicine*, **21**, 1991-2012. doi:10.1002/sim.1176

[9] Patel, N.R. and Ankolekar, S. (2007) A Bayesian approach for incorporating economic factors in sample size design for clinical studies of individual drugs and portfolios of drugs. *Statistics in Medicine*, **26**, 4976-4988. doi:10.1002/sim.2955

[10] Leung, D.H. and Wang, Y.G. (2001) A Bayesian decision approach for sample size determination in phase II studies. *Biometrics*, **57**, 309-312. doi:10.1111/j.0006-341X.2001.00309.x

[11] Shao, Y., Mukhi, V. and Goldberg, J.D. (2008) A hybrid Bayesian-frequentist approach to evaluate clinical studies designs for tests of superiority and non-inferiority. *Statistics in Medicine*, **27**, 504-519. doi:10.1002/sim.3028

[12] Kikuchi, T., Pezeshk, H. and Gittins, J. (2008) A Bayesian cost-benefit approach to the determination of sample size in clinical studies. *Statistics in Medicine*, **27**, 68-82. doi:10.1002/sim.2965

[13] Jiang, H., Liu, Y. and Su, Z. (2009) An optimization algorithm for designing phase I cancer clinical studies. *Contemporary Clinical Trials*, **29**, 102-108. doi:10.1016/j.cct.2007.06.003

[14] Huang, X., Biswas, S., Oki, Y., Issa, J.P. and Berry, D.A. (2007) A parallel phase I/II clinical studies design for combination therapies. *Biometrics*, **63**, 429-436. doi:10.1111/j.1541-0420.2006.00685.x

[15] Baker, S.G. and Heidenberger, K. (1989) Choosing sample sizes to maximize expected health benefits subject to a constraint on total studies costs. *Medical Decision Making*, **9**, 14-25. doi:10.1177/0272989X8900900104

[16] Spiegelhalter, D.J. and Best, N.G. (2003) Bayesian approaches to multiple sources of evidence and uncertainty in complex cost-effectiveness modelling. *Statistics in Medicine*, **22**, 3687-3709. doi:10.1002/sim.1586

[17] Briggs, A. and Sculpher, M. (1997) Markov models of medical prognosis—Commentary. *British Medical Journal*, **314**, 354-355. doi:10.1136/bmj.314.7077.354a

[18] Willan, A.R. and Pinto, E.M. (2005) The value of information and optimal clinical trial design. *Statistics in Medicine*, **24**, 1791-1806. doi:10.1002/sim.2069

[19] Parody, E.R. (2007) Análisis del coste-utilidad de la resonancia magnética en el manejo del paciente con isquemia cerebral aguda. Universitat Autónoma de Barcelona.

[20] Fagan, S.C., Morgenstern, L.B., Petita, A., Ward, R.E., Tilley, B.C., Marler, J.R., *et al.* (1998) Cost-effectiveness of tissue plasminogen activator for acute ischemic stroke. *Neurology*, **50**, 883-890. doi:10.1212/WNL.50.4.883

[21] Pinto-Prades, J. and Abellán-Perpiñán, J. (2005) Measuring the health of populations: The veil of ignorance approach. *Health Economics*, **14**, 69-82. doi:10.1002/hec.887