J. Sci. Math. Lett. UPSI Volume 3 (2015) 24 - 30
ISSN 2462-2052

A Narrative Review of Two-Arm and Three-Arm Non-Inferiority Clinical Trials

Tinjauan Naratif Ujian Klinikal Tidak Inferior bagi 2-Kumpulan dan 3-Kumpulan Rawatan

Zurina Mohd Yusof[1] & Nor Afzalina Azmee[2]
Department of Mathematics, Universiti Pendidikan Sultan Idris, 35900 Tanjung Malim, Perak
[1]zurinakhairi75@gmail.com & [2]afzalina@fsmt.upsi.edu.my

Abstract

Clinical trial is a research conducted to test the effectiveness and safety of the experimental drugs or treatments or diagnostic tools before approval using consenting human subjects as sample size. Non-inferiority test in clinical trial is performed to demonstrate that the experimental treatment is not worse than the reference treatment by more than a pre-defined margin. This type of design has accumulated much attention from the researchers as it is viewed as an alternative to the existing superiority trials. However, issues in the design and analysis of non-inferiority trials are still at large and highly debatable. Thus, the object of this paper is to address the gap in the literature, by providing a concise, narrrative review of some selected papers related to non-inferiority trials. The review of 162 published papers indicates potential studies related to two-arm and three-arm non-inferiority trials, focusing on the implementation of Bayesian design and analysis.

Keywords non-inferiority trial, two-arm, three-arm, Bayesian analysis

Abstrak

Ujian klinikal merupakan satu penyelidikan yang dijalankan bagi menguji keberkesanan dan keselamatan ubat atau rawatan atau alat diagnostik, sebelum ia diluluskan dengan menggunakan subjek manusia sebagai saiz sampel. Ujian klinikal tidak inferior bertujuan untuk membuktikan yang rawatan percubaan tidak kurang baik berbanding dengan rawatan piawai. Dengan pengurangan keberkesanan haruslah tidak melebihi had tertentu yang ditetapkan. Reka bentuk ujian ini telah mendapat perhatian ramai penyelidik dan ia dianggap sebagai alternatif kepada ujian klinikal yang bertujuan untuk membuktikan bahawa rawatan percubaan lebih unggul daripada rawatan piawai. Walau bagaimanapun, isu-isu yang berkaitan dengan reka bentuk dan analisis ujian tidak inferior masih banyak diperbahaskan. Justeru, objektif kertas ini adalah untuk menangani jurang literatur dengan menyediakan ulasan naratif yang ringkas dan padat berkenaan ujian tidak inferior. Tinjauan sebanyak 162 kertas penyelidikan ini menunjukkan yang kajian lanjut diperlukan dalam pelaksanaan reka bentuk dan analisis Bayesian bagi ujian tidak inferior untuk 2-kumpulan dan 3-kumpulan rawatan.

Keywords ujian tidak inferior, 2-kumpulan, 3-kumpulan, analisis Bayesian

INTRODUCTION

Clinical trial is a scientific experiment that is bounded by ethical regulations in order to protect the patients' interest. For example, patients can only be recruited in a trial after they have given their informed written consent and these patients must have been clearly informed about the risks or the side effects they might experience when taking the medications. These scientific experiments are governed by well-recognized, regulatory bodies such as the International Conference of Harmonisation, the Commitee for Medical Products for Human Use, the U.S. Food and Drugs Administration and the Agency for Healthcare Research and Quality (see details in ICH, 1998; CHMP, 2005; FDA, 2010; AHRQ, 2012).

Any type of clinical trial can possibly have the objective of showing either superiority, equivalence or non-inferiority of the experimental treatment with respect to the current, reference treatment. In the late 1990s, non-inferiority trials started to garner interest among the researchers in either statistical or medical background. This type of trial is often sought after by the researchers to demonstrate that although the experimental treatment is non-inferior to reference with respect to efficacy, it possesses other advantages such as being cheaper, easier to be administered or has fewer side effects as opposed to reference (see discussion by Pigeot et al., 2003; Koch & Rohmel, 2004; Tang & Tang, 2004). It is important to note that non-inferiority trials can only be conducted if a good reference treatment exists.

In general, the non-inferiority trials can be commonly categorized as the two-arm trial and three-arm trial. A simple two-arm non-inferiority trial involves experimental and refererence arms, whereas a three-arm non-inferiority trial includes an extra placebo arm. More than 30 years has past since the idea of non-inferiority trial was first introduced and it is high time to revisit the issues surrounding the conduct of non-inferiority trials. The narrative review of these selected papers is hoped to impart an adequate background study to novice researchers and to identify potential areas for further studies.

METHOD

The aim of of this paper is to provide an overview of two-arm and three-arm non-inferiority trials, hence termed as a narrative review. In this case, explicit or systematic literature search protocol was not implemented in selecting and appraising evidence. Selected articles were extracted from clinical and non-clinical non-inferiority trials published between 1982 and early 2014. Those articles were obtained via a literature search, using the following keywords; non-inferiority, sample size determination, assurance. The following criteria such as general information related to authors, publication year, journal name, main theme discussed, type of non-inferiority clinical trials conducted, the sample size methods and the statistical analyses considered were noted. Some of the articles were not accessible and for such a problem, the review was based on just the abstract.

The Narrative Review of Non-Inferiority Trials

One of the earliest problem in conducting non-inferiority trials is the perplexity of analyzing such trials. The trials are often, wrongly analyzed by using the conventional hypothesis testing for superiority trials, where non-rejection of the null hypothesis is implied as successfully

J. Sci. Math. Lett. UPSI Volume 3 (2015) 24 - 30
ISSN 2462-2052

showing that an experimental treatment is non-inferior to reference. Blackwelder (1982) offered a solution to this problem, proposing that the null hypothesis should be stated in terms of a specified difference, δ. Although the fundamental idea of testing non-inferiority was given much earlier in Blackwelder (1982), it was not immediately picked up until early 2000s (see Figure 1).

During that period of 1980s – 1990s, non-inferiority trials were at the early stage of development and lack of understanding was common among practitioners. The issues commonly raised include how the non-inferiority margin should be chosen or how assay sensitivity can be assessed. Defining a non-inferiority margin is an important part of the methodology in the two-arm non-inferiority trials. This margin represents the reduction in efficacy in the experimental treatment that is thought to be tolerable. Because subjective and divergent opinions may arise with respect to either statistical or medical reasons, determining a margin becomes a debatable and controversial issue (see Hwang & Morikawa, 1999; D'Agostino et al., 2003). On the other hand, assay sensitivity is defined as the ability of a trial to differentiate between effective and ineffective treatments. This problem is inherited in two-arm non-inferiority trials and leads to a challenge of assuring that the experimental treatment is at least as good as the standard treatment (Hwang & Morikawa, 1999; Laster & Johnson, 2003). A solution to this problem is proposed by Blackwelder (2004), stressing the importance of maintaining a high degree of adherence protocol in the equivalence and non-inferiority trials and demonstrating a strong interim evidence which may be an indicator that an experimental treatment is superior to reference.

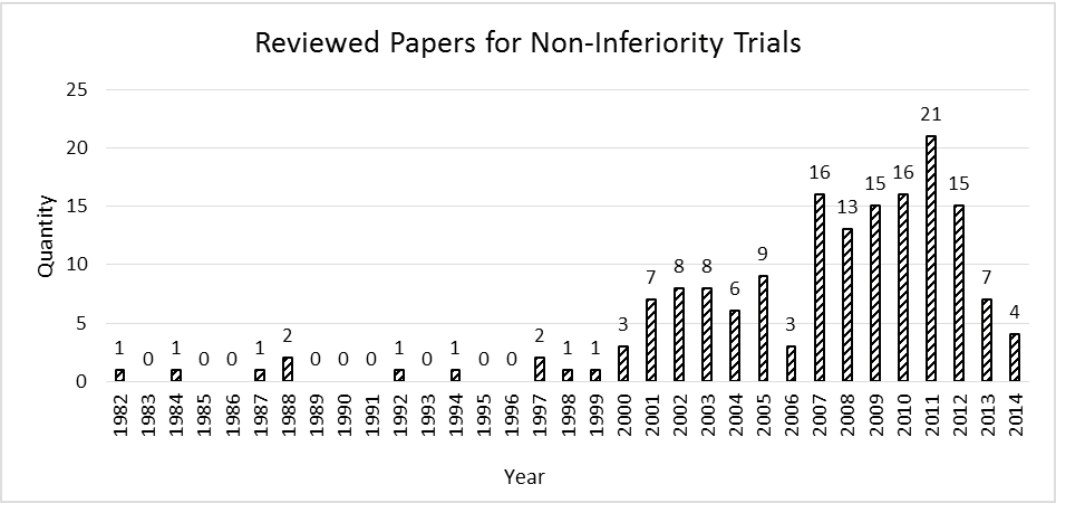

Figure 1 Classification of 162 overall published papers across the years

Across the years, the debates regarding the pros, the cons, the barriers and the challenges of conducting non-inferiority trials continously appeared in academic journals (see Snappin, 2000; Pocock, 2001; Hung et al., 2001; D'Agostino et al., 2003; Wangge et al., 2012). Some other studies were devoted to designing the protocol and reporting the non-inferiority trials (Weins, 2002; Brittain & Lin, 2005; Le Henanff et al., 2006; Piaggio et al., 2006). In particular, Piaggio et al. (2006) recommended some examples

on how to write an effective report on non-inferiority trials to avoid unethical issues and misunderstanding in the trials. Ethical issues have also been raised in Garratini & Bertele' (2007) and Hung et al. (2007). For example, the non-inferiority trials have been viewed by some as a platform to justifiably enter patients into a trial that will not provide them any advantage. The non-inferiority trials are also seen as means to merely cut down the cost and shorten the time. Others argued that non-inferiority is still needed in some clinical trials such as in the treatment of diseases for tuberculosis, leukimia and treatment involving antibiotics (see Temple & Ellenberg, 2000; Nunn et al., 2008; Chuang-Stein, 2008).

As given in Figure 2, the design of three-arm non-inferiority trial started to receive considerable attention from the year 2003 onwards, due to the unresolved problems found in the design of two-arm non-inferiority trials, which includes assuring assay sensitivity and defining a proper margin. The inclusion of a placebo arm allows for direct proof of efficacy of the new experimental treatment, with respect to placebo. In situations where delaying the current reference treatment is not going to cause irreversible morbidity, three-arm non-inferiority trial is seen as an appealing choice (Pigeot et al., 2003; CHMP, 2005; Britton, 2007; Munk et al., 2007; Dette et al., 2008; Ghosh et al., 2011; Hida & Tango, 2011). In particular, Koch & Rohmel (2004) summarized clear-cut situations where the inclusion of a placebo arm may well be supported; such as when the reference treatment is traditional, weak or volatile or that the trial is conducted to cure a disease that is not fully understood yet. The hot topics being discussed in the context of three-arm non-inferiority trials include methods of data analysis and sample size determination.

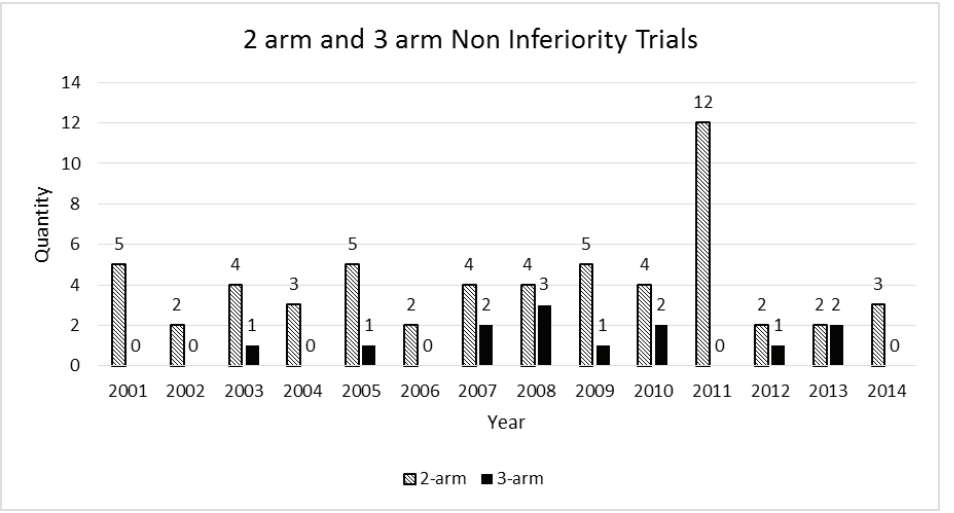

Figure 2 Comparison of published papers specifically related to two-arm and three-arm non-inferiority trials.

In three-arm non-inferiority trials, multiple comparisons arise due to the interest of comparing both the experimental and the reference with respect to placebo and then assessing the relative efficacy between the two active treatments. To have a meaningful interpretation of non-inferiority, it is important to establish superiority of both active treatments with respect to placebo. To conduct several multiple tests simultaneously will

J. Sci. Math. Lett. UPSI Volume 3 (2015) 24 - 30
ISSN 2462-2052

lead to a problem of multiplicity and this is discussed in Pigeot et al. (2003) and Koch & Rohmel (2004). The papers reviewed also indicates the tendency of investigators to adopt the Bayesian approach in data analysis (for example Britton, 2007; Ghosh et al., 2011). A Bayesian approach was implemented in Britton (2007) to examine the deviance information criterion used in camparing various models. The evaluation took into account twelve different prior distributions to produce posterior distribution, in order to assess the best fit model. On the other hand, Ghosh et al. (2011), considered a Bayesian analysis, which incorporated parametric as well as semi-parametric models. The benefit of the proposed Bayesian method was assessed via simulation, allowing for the conditions presumed in the study protocol.

Sample size determination is another important element in the design stage of clinical trial. Sample size that is either too small or too large may be judged unethical and may lead to serious consequences. As an example, a trial that has a small sample size may have little chance of showing the difference between the mean populations of two treatment groups, should the difference truly exists. On the other hand, a study that has a too large sample size could have met the objectives of the trial before the end of the study. In a latter case, some patients may have unnecessarily entered the trial and results in waste of resources. Papers advocating the frequentist method for sample size calculation for two-arm and three-arm include Pigeot et al. (2003), Julious (2004) and Friede & Kieser (2008). In particular, Pigeot et al. (2003) proposed an optimal sample size allocation across the three distinct groups, assuming the case of normally distributed variables with homogeneity of variances. Other studies who proposed a Bayesian approach had one unified theme, that is Bayesian approach allows the uncertainty of estimation be represented by using proper prior distributions. The constructions of these priors have to be based on evidence and sound judgement and subjectively vary from trial to trial. Among the earliest paper found is in Joseph et al. (1997), who gave a note on Bayesian methods in sample size calculation based on length and coverage criteria. An application of variation of that methods, such as the average length criteria (ALC) or the average coverage criteria (ACC) in non-inferiority trials is given in Wang & Stamey (2010). Assurance is another method that falls under Bayesian category. It is an interesting concept because it is close to the idea of power, that is by setting assurance at a certain level, one should find the sample size that fulfills the desired assurance. The application of assurance method was first introduced by O'Hagan & Steven (2001) in the study of cost-effective comparison between competing interventions, but the application in non-inferiority trials is seen later in O'Hagan et al. (2005). The study related to application in one sided superiority, two sided superiority, equivalence and non-inferiority trials with examples of simple problems of normal, binary and gamma distributed. Recently, a study in Azmee et al. (2013) demonstrated the application of Bayesian Clinical Trial Simulation in finding the required sample size via assurance for the case of three-arm non-inferiority trials, with normally distributed variables and homogeneity of variances.

CONCLUSION

Non-inferiority trials are considered to be a new area in medical statistics and some of the issues in the design and analysis are still debatable and being researched. These certainly illustrate the needs and motivations to continue exploring those areas in non-inferiority

trials. The review of literature indicates that the frequentist approach has been the dominant approach in both design and analysis of data. However, there is a considerable trend of switching from the conventional approach to Bayesian approach, a substantial gap that can be addressed in the near future. Example of potential studies include the implementation of Bayesian analysis, allowing for covariates or the implementation of Bayesian approach in sample size determination in three-arm non-inferiority trials, assuming normally distributed variables with heterogeneity of variances.

REFERENCES

AHRQ - Agency for Healthcare Research and Quality (2012), *2012 National Health Care Quality Report.* [Online]. Available from: http://www.ahrq.gov/ research/findings/nhqrdr /nhqr12/ index.html [Accessed 25th September 2014].

Azmee, N.A., Zulkifley, M. & Ahmad, A. (2013) *Determination of the required sample size with assurance for three-arm non-inferiority trials*. Jurnal Teknologi. 63: 89-93.

Blackwelder, W.C. (1982). *Proving The Null Hypothesis in clinical trials.* Controlled Clinical Trials. 3:345-353.

Blackwelder, W.C. (2004). *Current Issues in Clinical Equivalence Trial*. Journal of Dental Research. 83:113-115.

Brittain, E & Lin, D. (2005) *A comparison of intent-to-treat and per-protocol results in antibiotic non-inferiority trials*. Stat Medicine. 24:1-10.

Britton, M.C. (2007). *Bayesian Approach to Three-arm Non-Inferiority Trials*. Unpublished Mathematics Thesis, Georgia State University. [Online]. Available from: http://scholarworks.gsu.edu/cgi/viewcontent.cgi?article=1026&context=math_theses [Accessed 25th September 2014].

Chuang-Stein, C. (2008). *Assay Sensitivity*. Wiley Online Library.

CHMP – Committee for Medicinal Products for Human Use (2005). Guideline on the choice of non-inferiority margin. [Online]. Available from: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003636.pdf [Accessed 25th September 2014].

D'Agostino, R.B Sr., Massaro, J.M. & Sullivan L.M. (2003). *Non-inferiority trials: design concepts and issues - the encounters of academic consultants in statistics.* Stat Med. 22:169-186.

Dette, H., Trampisch, M. & Hothorn, L.A. (2008*). Robust Design in Non-Inferiority Three-arm Clinical Trials with Presence of Heteroscedasticity.* [Online]. Available from: http://www.ruhrunibochum.de/imperia/md/content/mathematik3/publications/revision_unblinded_final__3_.pdf [Accessed 25th September 2014].

FDA (2010). *Guidance for Industry: Non-Inferiority Clinical Trials* [Online]. Availabe from: http://www.fda.gov/downloads/Drugs/Guidances/UCM202140.pdf [Accessed 25th September 2014].

Friede, T. & Kieser, M. (2008) *Sample Size Reestimation in Non-Inferiority Trials.* [Online]. Available from: http://www.efspi.org/PDF/activities/international/non_inf_2008/7_re-estimation_TF.pdf [Accessed 25th September 2014].

Garattini, S. & Bertele', V. (2007). *Non-inferiority trials are unethical because they disregard patients' interests.* Lancet. 270:1875-1877.

Ghosh P.. Nathoo, F. Gonen, M. & Tiwari, R.C. (2011). *Assessing non-inferiority in a three-arm trials using the Bayesian Approach.* Stat Med. 30: 1795-1808.

Hida, E. & Tango, T. (2011). *On the three-arm non-inferiority trial including a placebo with a prespecified margin.* Stat Med. 30:224-231.

Hung, H.M.J., Wang S.J., Tsong Y, Lawrence J & O'Neil RT. (2001). *Some Fundamental Issues with Non-Inferiority Testing In Active Controlled Trial.* Stat Med. 22:213-225.

Hung, H.M.J., Wang, S.J. & O'Neill, R. (2007*). A regulatory perspective on choice of margin and statistical inference issue in non-inferiority trial*s. Biometrical Journal, 47 :28-36, 2007.

Hung, H.M.J.,. Wang, S.J. & O'Neill, R (2007*). Issues with statistical risks for testing methods in noninferiority trial without a placebo arm*. Journal of Biopharmaceutical Statistics, 1:213.

Hwang, I.K. & Morikawa, T. (1999). *Design Issues in Non-Inferiority/Equivalence Trials.* Drug Information Journal. 33:1205-1218.

ICH – International Conference on Harmonization (1998) Note for guidance on statistical principles for clinical trials. [Online]. Available from: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500002928.pdf [Accessed 25th September 2014].

Joseph, L., du Berger, R. & Belisle, P. (1997). *Bayesian and Mixed Bayesian/Likelihood criteria for Sample Size determination.* Statistics In Medicine. 16:769-781.

Julious, S.A. (2004). *Tutorial in Biostastics:Sample sizes for clinical trials with Normal data.* Statistics in Medicine. 23:1921-1986.

Koch, A. & Rohmel, J. (2004) *Hypothesis testing in the "gold standard" design for proving the efficacy of an experimental treatment relative to placebo and a reference*. Journal of Biopharmaceutical Statistics. 14: 315-325.

Laster, L.L. & Johnson, M.F. (2003). *Non-inferiority trials: the 'at least as good as' criterion.* Statistics In Medicine. 22: 187-200.

Le Henanff, A., Giraudeau, B., Baron, G. & Ravaud, P. (2006) *Quality of reporting of noninferiority and equivalence randomized trials*. Journal of American Medical Association. 295:1147-1151.

Munk, A., Meilke, M., Skipka, G. & Freitag, G. (2007). *Testing Non-Inferiority in Three-Armed Clinical Trials Based on The likelihood Ratio Statistics.* The Canadian Journal of Statistics. 35:413-431.

Nunn, A.J., Meredith, S.K., Spigelman, M.K., Ginsberg, A.M. & Gillespie, S.H. (2008). *The Ethics of Non-Inferiority Trials.*Lancet. 371: 895.

O'Hagan, A. & Stevens J.W. (2001), *Bayesian Assessment of Sample Size for Clinical Trials of Cost-Effective*. Medical Decision Making. 21:219-230.

O' Hagan, A., Stevens, J.W. & Campbell, M.J. (2005). *Assurance in Clinical Trial Design.* Pharmaceutical Statistics. 4:187-201

Piaggio, G., Elbourne D.R., Altman, D.G., Pocock, S.J., Evans, S.J. & Group, C. (2006*). Reporting of Noninferiority and Equivalence Randomized Trials An Extension of the CONSORT Statement.* Journal of American Medical Association. 295:1152-1160.

Pigeot, I., Schafer, J., Rohmel, J. & Hauschke, D. (2003). *Assessing Non-Inferiority of a New Treatment in a Three-arm Trial Including Placebo*. Statistics in Medicine. 22:833-889.

Pocock, S.J. (2001). *The pros and cons of noninferiority trials.* Blackwell Publishing. Fundamental & Clinical Pharmacology. 17:483-490.

Snappin, S.M. (2000). *Non-Inferiority trials. Current Controlled Trials.* Cardiovasc. Med. 1:19-21.

Tang, M.L. & Tang, N.S. (2004) *Test of noninferiority via rate difference for three-arm clinical trials with placebo*. Journal of Biopharmaceutical Statistics. 14: 337-347.

Temple, R. & Ellenberg, S.S. (2000). *Placebo-Controlled Trials and Active-Control Trials in the Evaluation of New Treatment*. Part 1: Ethical and Scientific Issues. Annals of Internal Medicine.133:455-463.

Wang, J. & Stamey, J.D. (2010). *A Bayesian algorithm for sample size determination for equivalence and non-inferiority test*. Journal of Applied Statistics. 37: 1749-1759.

Wangge, G., Roes, K.C.B., Boer, A., Hoes, A.W. & Knol, M.J. (2012). *Non-Inferiority trials: methodological and regulatory challenges.* Unpublished thesis, Utretcht University. [Online]. Available from: http://dspace.library.uu.nl/handle/1874/256026. [Accessed 25th September 2014].

Weins, B.L. (2002). *Choosing an equivalence limit for noninferiority or equivalence studies.* Controlled Clinical Trials. 23:774.