uOttawa

L'Université canadienne
Canada's university

FACULTÉ DES ÉTUDES SUPÉRIEURES ET POSTOCTORALES

uOttawa

L'Université canadienne
Canada's university

FACULTY OF GRADUATE AND POSDOCTORAL STUDIES

Darcie Dow

AUTEUR DE LA THÈSE / AUTHOR OF THESIS

M.Sc. (Epidemiology)

GRADE / DEGREE

Department of Epidemiology and Community Medicine

FACULTÉ, ÉCOLE, DÉPARTEMENT / FACULTY, SCHOOL, DEPARTMENT

Analysis and Review of Analytic Approaches to Dealing with Post-Randomized Ineligible Patients Resulting from a Change in Post-Randomization Disease-State Ascertainment

TITRE DE LA THÈSE / TITLE OF THESIS

Shawn Aaron

DIRECTEUR (DIRECTRICE) DE LA THÈSE / THESIS SUPERVISOR

Dean Fergusson

CO-DIRECTEUR (CO-DIRECTRICE) DE LA THÈSE / THESIS CO-SUPERVISOR

EXAMINATEURS (EXAMINATRICES) DE LA THÈSE / THESIS EXAMINERS

George Wells

Shi Wu Wen

Gary W. Slater

Le Doyen de la Faculté des études supérieures et postdoctorales / Dean of the Faculty of Graduate and Postdoctoral Studies

Analysis and review of analytic approaches to dealing with post-randomized ineligible patients resulting from a change in post-randomization disease-state ascertainment

Darcie Dow

Thesis submitted to the
Faculty of Graduate and Postdoctoral Studies
In partial fulfillment of the requirements
For the MSc degree in Epidemiology and Community Medicine

Epidemiology and Community Medicine
Faculty of Medicine
University of Ottawa

Your file Votre référence
ISBN: 978-0-494-25764-7
Our file Notre référence
ISBN: 978-0-494-25764-7

Abstract:

In a series of three parts, this thesis explores the available approaches to analyze post-randomized ineligible patients. Part one: a comprehensive systematic review yielding 18 unique articles from the Medline and the Cochrane Methodology Register databases. Part two: the methods found from the review were applied to a dataset where 23% of the patients in the experimental treatment group were found to be post-randomized ineligible patients. The results of this exercise revealed that little can be done when these patients exist and the best approach to dealing with them is to carefully choose an approach and accurately report how these patients were analyzed. Part three: Influenza was explored to further illustrate the importance of accurate reporting and the broader impact of this analytic issue. In sum, no firm conclusion could be reached on one "best" approach to take for randomized ineligible patients, however prevention and accurate reporting were surrogate recommendations.

Executive Summary

Rationale: Post-randomized ineligible patients can occur in many acute-care clinical trials and threaten to dilute treatment effects, contribute to misleading results, and introduce bias into trial results if they are not appropriately handled during data analysis.

Objective: First, to explore the analytic approaches available to deal with post-randomized ineligible patients and illustrate their various strengths and weaknesses using an actual dataset containing such patients. Second, to explore the broader clinical and public health impacts of this issue using a second example.

Methods and Results: A comprehensive search of Medline and the Cochrane Methodology Register yielded 18 unique papers which explained the various strengths and weaknesses of eight analytic approaches for dealing with issues such as randomized ineligibles. To further illustrate the merits of the findings, each approach was applied to the original dataset of a recently completed trial on Cystic Fibrosis patients. In this trial 23% of the patients in the treatment arm experienced a change in disease-state that rendered the treatment potentially ineffective. Using an intention-to-treat analysis of the trial data, no significant difference was found between the intervention and the control arms. When each alternative analytic approach that accounted for patients' change in disease-status was applied to the dataset, the trial results remained non-significant in all cases. However, the direction of benefit and the confidence intervals did reflect the theorized strengths

and weaknesses of each analytic approach.

In order to ensure the lack of effect was not due to an unrepresentative sample, major prognostic baseline factors were bootstrapped (3000 iterations) to assess the normality of the data. Results revealed normal baseline data.

As a final step to investigate the implications of the issue of randomized ineligible patients, an additional systematic review of influenza treatment trials was conducted. Results from this review showed that out of 11 influenza treatment trials identified, nearly all investigators excluded randomized but uninfected patients from their primary analysis.

Conclusion: It was concluded that although many approaches to data analysis exist for issues such as non-compliance and misclassification – few of these approaches are appropriate for a post-hoc analysis of randomized ineligible patients. Investigators should be well aware of the potential pitfalls of their clinical trial methodology and make substantive efforts to prevent randomization of potentially ineligible patients. In the event this cannot be prevented, analytic techniques should be chosen a priori to adjust and account for randomized ineligible patients. Investigators should be well schooled in the strengths and limitations of their analytic approach as well as their interpretation.

Acknowledgments

I long to accomplish a great and noble tasks, but it is my chief duty to accomplish humble tasks as though they were great and noble. The world is moved along, not only by the mighty shoves of its heroes, but also by the aggregate of the tiny pushes of each honest worker.
Helen Keller

I would like to extend my sincere gratitude to my supervisors Dr. Shawn Aaron and Dr. Dean Fergusson for their outstanding contribution to this manuscript. These two helped me learn how to write and think and persevere. I would also like to thank Dr. Aaron for the funding he so generously provided for this project. Secondly, I would like to thank the Department of Epidemiology and Community of Medicine and the faculty therein. They provided me with the skill sets to complete such a work at this. To the Women's Health Research Unit, thank you for the opportunity to work not only in the real world of research, but on a project that will impact so many people. Steve Doucette, thank you for teaching me SAS all over again and for making yourself so available for meetings and questions. Doug Coyle, thank you for showing me how to bootstrap in such a user-friendly way. Finally, I would also like to thank Keith and all of my loved ones who have no idea what this thesis is about and probably never will. Your support was unwavering and without it, I would never have finished. Thank you all for your prayers, and thank you God, for hearing them.

Table of Contents

List of Tables

Chapter 2

Chapter 3

Chapter 4

List of Figures

Chapter 2

Chapter 3

Chapter 4

Chapter 1: General Introduction

1.1 Background

Research involving human subjects is, by its nature, more difficult to control than research involving animals or cellular models (1). In animal experiments, the investigators can maintain tight control over study subjects and environments, such as gender, weight, feeding times and room temperature. Thus, they are able to create an environment where only the experimental agents vary; a standard of control that is not possible in human randomized control trials (RCT). Clinical trials involving human subjects can also be imperfect due to patient noncompliance with trial protocol, or due to randomization of ineligible patients, or due to the variable and shifting nature of diseases that affect humans. Human clinical trials can be biased by flaws in methodology or by particular analytic issues that can arise during the course of the study. This thesis is mainly concerned with the issue of dilution of treatment effect as it relates to changes in disease status that are ascertained after randomization is complete and treatment is initiated.

In order to ensure terminology is clear throughout the thesis, a number of definitions are required. To begin, compliance encompasses a large number of assessment procedures and schedules. Throughout this thesis, non-compliance will refer to a patient's failure to adhere to the treatment to which they were randomized. This failure to adhere to treatment may be caused by switching therapies (treatment crossover), supplementing therapies (co-intervention and/or contamination), or not taking the prescribed doses of the therapy. Another

important term to clarify up front is misclassification, which is an error in determining disease status including a "change in disease-status" which means a change in patient diagnosis which renders the treatment partially or wholly ineffective and may result in a failure to meet inclusion criteria post-randomization.

It appears that the methodological issue of post-randomization change in disease status is specific to the acuity of the medical condition under study. For instance, acute medical conditions that require immediate intervention are vulnerable to the problem of misclassification via an inaccurate diagnosis of the disease status of the patients at the time of randomization. This is not a flaw in study design, but a scenario that reflects the real-life clinical decision-making that occurs every day. When patients present symptoms that suggest influenza, the physician does not have time to wait 7 days for confirmation from the lab, but must choose to treat or not to treat the patient immediately. Such is the nature of acute infection; treatment cannot be delayed until diagnostic confirmation. As a result, the investigators are left with enrolled patients who no longer meet study inclusion criteria. These patients present a unique set of problems during trial analysis and interpretation such as dilution of effect, loss of study power, bias, and a misrepresentation of the cost-benefit ratio.

From a clinical perspective, including these patients in the analysis will dilute the effectiveness of the intervention, leaving the efficacy of the intervention unknown. In most of these cases, an intention-to-treat analysis of the results is followed up by a sensitivity or efficacy analysis, removing those patients whose

disease-status had effectively changed in the time between pre-randomization and post-randomization.

Several issues arise when patients are excluded from the final data analyses. The first major issue is loss of study power. Original sample size calculations for clinical trials are based on the suspected difference between the treatment and control groups, as well as the desired level of power. When patients are removed from the analysis the study power is decreased thus reducing the ability to detect a true treatment effect, if one exists.

A second important issue is that excluding patients from the primary intention-to-treat analysis can introduce bias into the study. Differential or even non-differential exclusion of selected study patients may result in trial results that favour one treatment group over the other. Whenever patient exclusions are permitted after randomization, selection bias can occur and the internal validity of the trial may be compromised.

From a public health perspective, the driving force behind clinical trials is the desire for information. It follows then, that when patients are excluded from a trial, not only is potential for bias introduced, but useful (and sometimes expensive) information is lost. Also, research shows that shifting disease status and non-compliance are not considered when financial cost benefit ratios are calculated (7). Similar to this, the personal risk and time involved in agreeing to be a trial participant is agreed upon under the assumption that there is potential to benefit from treatment. Consequently, when patients have little to no likelihood of

benefiting from treatment, the cost benefit ratio is misrepresented. Failing to accurately report and publish results for all patients is similar to failing to publish trials with negative results. Data is misrepresented and leads to reporting bias.

1.2 *Objective and Aims:*

Overall Objective: The overall objective of this thesis was to identify the available analytic approaches to dealing with post-randomized ineligible patients and explore their usefulness for minimizing the issues of dilution, loss of study power, bias, and unrepresentative results.

Aims:

1

1. **To determine what approaches are available to adjust for the deviation between pre-randomization and post-randomization disease-state ascertainment as well as to identify the respective strengths and limitations of each approach.**

Research Questions:

- **What are the approaches available to address these issues?**
- **What are the principles on which these approaches based on?**
- **How does each approach affect the measure and precision of effect?**
- **What are the strengths and weaknesses of each approach?**
- **What are the underlying assumptions of each approach?**
- **For what type of study is each approach appropriate?**

2. **To determine the degree to which a deviation in disease-state ascertainment affects the robustness of estimates in the MCBT trial.**

Research Questions:

- **How do the different approaches affect the study power of the trial?**
- **What is the magnitude of the difference in the point estimates when each approach is applied to the Multiple Combination Bactericidal Test trial?**

- **What are the implications of these results? (Should clinical practice be changed?)**
- **Does bootstrapping the baseline characteristics of the MCBT population offer any explanation for the trial results?**

1.3 Format of the Thesis

Chapter 2: A comprehensive systematic review of the literature concerning the issue of post-randomized ineligible patients and analytic approaches to deal with post-randomized exclusions is conducted. In this section, the approaches are identified and discussed in detail.

Chapter 3: In this section the Multiple Combination Bactericidal Test (MCBT) dataset is introduced and the approaches which emerged from the systematic review are applied to the MCBT dataset. As a final precaution, bootstrapping is performed on the major baseline prognostic factors of the MCBT dataset to ensure the data are normal.

Chapter 4: To further illustrate the issues that arise from post-randomized ineligible patients, as well as highlight the broader implications of these issues, a second systematic review is conducted on influenza trials with an in-depth discussion on safety, economic and practical implications.

Chapter 5: In this section the previous three chapters are considered together to draw final conclusions regarding the issue of patients with post-randomization changes in disease status.

Chapter Two: Comprehensive Systematic Review of Analytic Approaches for Post-randomization ineligible patients

2.1. *Introduction and Objective of Review*

Approaches to deal with post-randomized ineligible patients, in particular, those with time-related post-randomization changes in disease-status in clinical trials, are worthy of further investigation. In order to further understand these issues, as well as determine how frequent randomized ineligibles were in the literature, an exhaustive systematic review of statistical and methodological analytic approaches to adjust for discrepancies was conducted.

2.2 *Systematic Review Methods*

2.2.1 Search Strategy

A broad search strategy was developed to best capture methodological and clinical articles addressing the issue of a post-randomization change in disease-status. In order to be included articles had to present an analytic approach which dealt with the issue of dilution of treatment effects. This issue as it pertains to patients with a change in disease status post-randomization has not yet been explicitly discussed in the methodological literature and so the searches were directed towards articles that presented approaches to dealing with patients found to be ineligible, non-compliant or mis-classified, post-randomization. The Cochrane Methodology Register was searched using the following pre-determined

index terms: bias and compliance, bias and intention-to-treat, bias and treatment analysis, bias and accrual and sample size, bias and statistical analysis, bias and study design. The register was also searched for the keywords compliance and analysis. In addition to the Cochrane search, an electronic search strategy was developed in consultation with clinical and methodological experts for the Medline database (Appendix Ai). For the Medline database, the Dickerson Randomized Control Trial filter was used to efficiently identify randomized clinical trials. These results were then searched using truncated versions of the following keywords: adjust, correct, compliance, adherence, analysis, misclassify, eligibility, and exclusion. Both electronic searches were not limited based on publication status however they were limited to English language articles. Moher et al.(9) and Juni et al (10) found that excluding trials in languages other than English did not significantly bias measures of effectiveness in randomized control trials, but both stated this as a cautionary result and recommended inclusive search strategies. It is not known if this finding can be applied to methodological searches and should be investigated further.

The Cochrane Methodology Register was searched on September 16 2005, and the Medline database search was completed from 1966 to September, Week 2, 2005. In addition to the electronic searches, a convenience sample of clinical trial text books and their references available in the library were screened for relevant chapters and additional articles. Again, the references of relevant articles were screened for further articles. Five clinical epidemiology and clinical trials journals

were hand-searched by table of contents: Journal of Chronic Disease (1970-1986); Biometrika (2001-2004); Journal of Clinical Epidemiology (2004-2005); Controlled Clinical Trials (2004-2005); and Statistics in Medicine (2004-2005). The latter 4 journals were hand-searched in order to update the hand-searching already completed in the compilation of the Cochrane Methodology Register. The selection of journals and conference proceedings were not based on hard criteria, but on a desire to update the searches on those journals most likely to carry articles pertaining to treatment effect dilution. The journals and conference proceedings and dates searched for this register are located in Appendix Aii.

Finally, ten content experts were formally written and e-mailed concerning this methodological issue. Relevant articles and information pertaining to how these patients should be dealt with were requested (Appendix Aiii-Av). These experts were selected based on the frequency of their name appearing in the literature, as well as the direction provided by Dean Fergusson, also a content expert in this area.

2.2.2 Screening and Selection

After initial identification of potential articles, the titles and abstracts of the papers were screened for relevancy in an unblinded fashion. According to Berlin (11) there is inconclusive evidence that blinding introduces bias into the screening process of randomized control trials. As with the limitation to English articles, this finding may not be applicable to methodological searches.

The Cochrane abstracts were screened by three independent reviewers (DD,

DF, SA). The Medline titles and abstracts were initially screened by one reviewer (DD), narrowing the search down to 41 potentially relevant abstracts. These were then reviewed by two additional reviewers (DF, SA) for inclusion criteria. Hand-searching and reference checks were done by one reviewer (DD) and identified articles were then screened by two additional reviewers for eligibility (DF, SA). Articles identified by content experts were reviewed by all three reviewers. Disagreements on eligibility were resolved by open consensus. All titles and abstracts meeting eligibility criteria were downloaded for full-text review. Full-text review was completed by three reviewers (DD, DF, SA) based on the same criteria used for the primary screening.

2.2.3 Eligibility Criteria

All titles, abstracts and full-text articles were screened for relevancy based on the following search objective:

The objective of the search is to identify relevant articles pertaining to patients who experience a change in disease-state ascertainment between pre-randomization and post-randomization in an attempt to estimate intervention efficacy within the context of a randomized controlled trial.

Therefore, relevant articles had to present and justify an analytic approach for dealing with these particular patients within randomized controlled trials. Articles were excluded for one or more of the following reasons: general irrelevance to topic of interest, failing to discuss analytic approach, or techniques not appropriate for

randomized controlled trial analysis.

Pilot studies were not completed to determine if appropriate search terms were incorporated into the search criteria. Few articles were expected to make it into the final analysis and therefore it is unlikely that a pilot study would have provided much insight into the appropriateness of the search terms.

2.2.4 Data Abstraction

One reviewer (DD) independently abstracted data from all articles that met the eligibility criteria using a pre-printed standardized data collection form located in Appendix Avi. The data collection form was developed in consultation with SA, and DF. Information pertaining to methodological issue, type of outcome (binary, continuous, time-to-event), underlying assumptions, approach strengths and weakness and details of the analytic approach (formulae), as well as the conditions under which the approach is best applied was collected. In addition, any further discussion or points of import presented by the authors was collected.

2.3 Systematic Review Results

A flow diagram summarizing the article selection process is presented in Figure 2.3.1. A total of 1715 citations were downloaded from electronic databases into RefWorks, reference managing web-based software. Of these, 18 duplicate articles were identified and subsequently removed. Two abstracts could not be retrieved. Hand-searching yielded an additional 10 potentially relevant citations and textbook and article reference checking revealed an additional 21 potentially

relevant articles. Of these, one could not be retrieved. Nine of the ten content experts suggested 16 unique references; and 13 were obtained. The remaining three references could not be located. In all, 1744 abstracts were screened for general relevancy and 1655 were excluded due to general irrelevancy or because the full-text articles were printed in a language other than English. The remaining 89 articles were identified as potentially eligible and underwent full-text review. Sixty-nine articles were excluded for the following reasons: no analytic approach discussed (n=20), inappropriate, non-applicable analytic approach discussed (n=34), or irrelevant subject matter (n= 14). Of the 20 eligible articles, 9 articles were ultimately retrieved from the Cochrane Methodology Register search, 3 from the Medline search, 4 from textbook reference lists, one from hand-searching, and 2 from the content experts. For a flow chart of content expert references see Figure 2.3.2. For a brief summary of content expert responses please refer to Appendix Av. A total of 19 articles met inclusion criteria and were included in the final analysis of this review.

Figure 2.3.1 Flow Diagram of Screened Studies

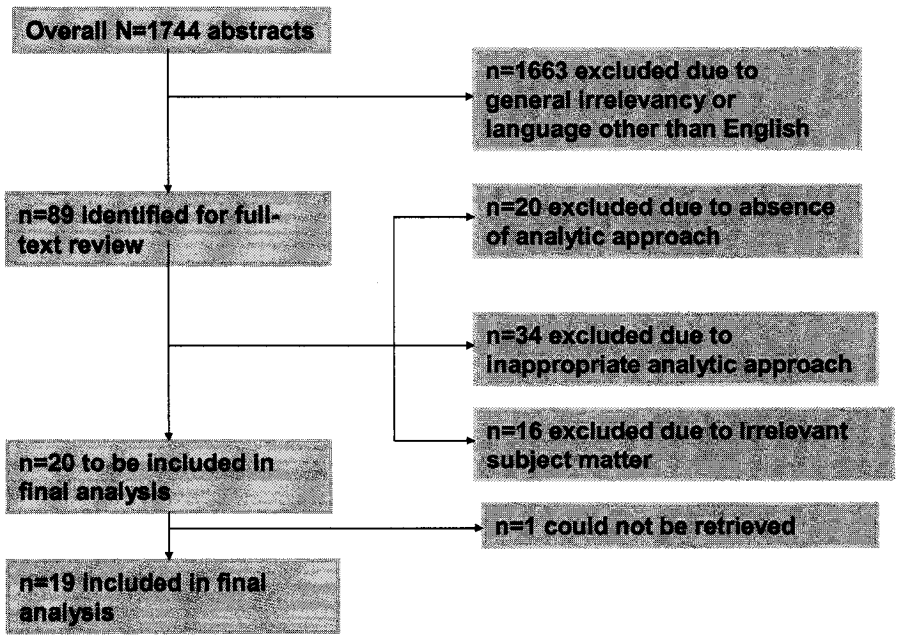
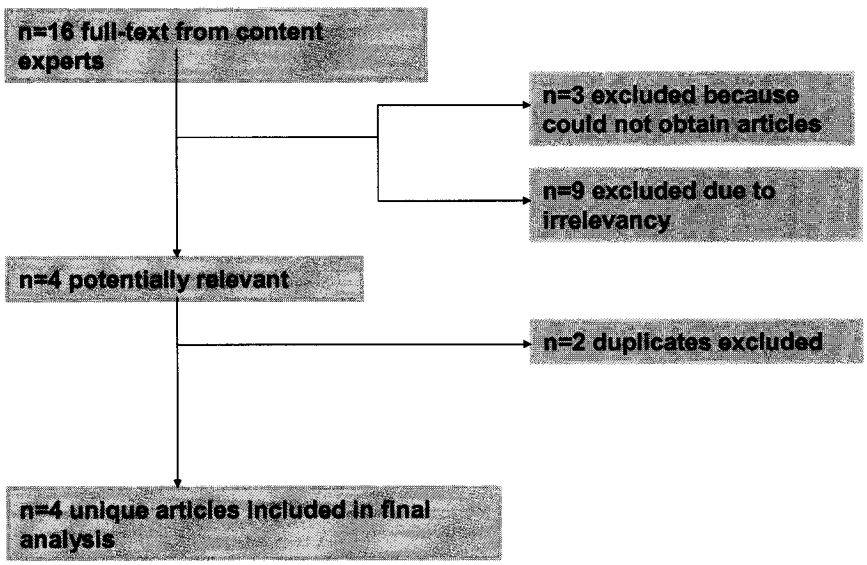


Figure 2.3.2 Flow Diagram of Content Expert Suggested Studies



2.3.1 Qualitative Summary of Included Studies

All 19 of the included papers were published as methodological pieces. Five

papers focused on the issue of non-compliance (12-16), four on misclassification or mis-measurement of disease status (17-20) one on treatment change (21), one on post-randomization exclusions (6), one on bias in analyses (22), and 5 on various approaches to data analysis (e.g., intention-to-treat (ITT), efficacy, per-protocol)(23-27). A summary of the characteristics of the included studies is provided in Table 2.3.1.

As illustrated in Table 2.3.1, the majority of the papers were from investigators based in the United States (n= 13). Two papers were from Canada and the United Kingdom and one each from Germany and the Netherlands. All of the papers were written between the years of 1973 and 2005.

Table 2.3.1 Characteristics of Included Articles

Source	Journal	Pub. Year	Issue	Outcome	General Description of Approach
Cochrane Methodology Register	Statist. Med	1997	Non-compliance and contamination	Binary with and without non-compliance and time-to-event	Instrumental Variable Analysis (IVA)
Cochrane Methodology Register	J Chron Dis	1973	Non-compliance	Time-to-event	Intention-to-treat vs. Secondary Analysis
Cochrane Methodology Register	Biometrics	2000	Non-compliance and dependent censoring	Time-to-event	Inverse Probability Weighting Censored
Cochrane Methodology Register	Statist. Med.	2000	Non-compliance	Not specified	Adjusted Treatment Received
Cochrane Methodology Register	Statist. Med.	1993	Non-compliance	Time-to-event	Intention-to-treat vs. Per-Protocol and Treatment Received
Cochrane Methodology Register	BMJ	2005	Post-randomization exclusions	Not-specified	Exclusion Principles
Cochrane Methodology Register	Controlled Clinical Trials	2000	Statistical considerations in the ITT principle	Not-specified	Intention-to-treat vs. Efficacy
Cochrane Methodology Register	Statist. Med.	2001	Outcome misclassification and measurement error	Binary	Intention-to-treat in Equivalence Trials
Cochrane Methodology Register	Circulation	1981	Bias in analysis	Not-specified	Exclusion Principles
Textbook Reference	J Clinical Pharm & Therap.	1995	ITT and goals of clinical trials	Not-specified	Instrumental Variable Analysis
Textbook Reference	Statist. Med.	1991	Analysis by treatment-received	Not-specified	Treatment Received
Textbook Reference	Statist. Med.	1991	Estimating Efficacy	Binary	Instrumental Variable Analysis
Medline	Controlled Clinical Trials	2002	Disease status and regression effect	Binary	Validation data
Textbook Reference	Thrombosis and Haemostasis	1979	Qualification and Disqualification of Patients	Time-to-event	Exclusion Principles
Medline	Statist. Med.	1996	Independent Multiple diagnostic classifications	Binary	Validation data
Medline	Statist. Med.	1988	Misclassification	Time-to-event	Validation data
Handsearch	Statist. Med.	2005	ITT and PP results in equivalence trials	Not-specified	Intention-to-treat vs. Per Protocol in equivalence trials
Content Expert	Statist. Methods in Medical Research	2005	Uses and Limitations of randomization-based efficacy estimators	Not-specified	Instrumental Variable Analysis
Content Expert	Book: Statistical Issues in Drug Development	1996	False Inclusions and ITT	Not-specified	Intention-to-treat and post-randomization ineligibles

2.3.2 Methodological Issues

Because the issue of altering disease-states and the appropriate method for dealing with randomized ineligible patients is not discussed in current methodological literature, our search objective was broadened to encompass articles dealing with all causes of treatment dilution. As a result our analysis is no longer about approaches designed to deal with patients experiencing a post-randomization change in disease status, but with all phenomena that result in dilution of the treatment effect. It was difficult to determine what aspects of the current medical literature could be applied to this issue of changing disease state. Particularly difficult were papers dealing with treatment non-compliance. According to Senn (28), lack of compliance can take on many forms, from drop out and dishonesty about taking study medication to physicians breaking protocol by admitting patients into a trial before inclusion criteria are satisfied. All of these issues lead to problems in deciding which patients should be included in the analysis. The included papers focus on a number of different analytic issues and various approaches to dealing with these issues that may in some way be applicable to our issue as well. Six papers focus on approaches to dealing with non-compliance in clinical trials. Non-compliance or non-adherence is defined by "a failure to observe the requirements of the protocol" (29).

Four papers dealt with outcome or exposure misclassification or mis-measurement, six other papers explored the various consequences of conducting

intention-to-treat and efficacy analyses, and one paper discussed independent diagnostic classifications. The final 3 papers focused directly on the problem of post-randomization exclusions and warn of the various potential biases associated with these exclusions.

2.3.3 Primary clinical outcomes addressed

The clinical outcome measure discussed in each paper depended on the specificity of the approach and conditions under which the approach was best applied. The statistical methods were designed using formulae for specific primary outcomes (e.g., binary, time-to-event). The design allocation approaches did not apply formulae but grouped patients. These approaches were not developed for a particular primary outcome but could be applied to any outcome measure.

2.3.4 Design Allocation Approaches

a) Intention-to-treat

The intention-to-treat approach to data analysis is the most widely accepted approach in randomized controlled trials. A conservative approach in superiority trials, this approach serves to analyze all randomized patients according to their original treatment assignment, regardless of compliance or change in disease status. In the event of substantial non-compliance or changes in original diagnosis, the treatment effect will be considerably diluted. Despite this characteristic, ITT remains the "gold standard" for randomized controlled trials.

Freedom from Bias

Perhaps the strongest argument in favour of this approach is that it is "virtually free of the influence of any of the many possible biases due to patient selection that may operate in [an] efficacy subset analysis" (24). ITT provides an unbiased estimate of treatment effectiveness, naturally incorporating compliance, concomitant drug-use, withdrawal due to adverse events, etc. (23). In short, it maintains the randomization of all unknown confounding variables between the treatment groups. In addition, allowing exclusions introduces an "arbitrary element" into the trial. This may cause problems since everyone may not agree on who should be excluded (28). Furthermore, more information is better than less. Knowing that the eligibility criteria are not perfect and if in the physicians' objective opinion, the patients are suitable for trial inclusion, there is little reason to discard patient information (28). In all, ITT is said to provide "the most realistic and unbiased answer to the more relevant question of clinical effectiveness" (24) and provide a "better estimate of the practical consequences of allocating a treatment" (28).

Biologic Efficacy vs. Effectiveness

With respect to our specific issue of ineligibility, all randomized patients who have a change in disease-status post-randomization would be analyzed with the treatment group to which they were allocated, regardless of whether this change rendered treatment ineffective. As a result, the treatment effect would be diluted by those patients randomized to a treatment group who do not have the disease of interest or by those whose disease status has changed post-randomization. This will

have considerable impact on study power. This is arguably the greatest weakness of the intention-to-treat approach (24); it does not measure biologic efficacy of the treatment. Because compliance or misclassification or change in disease-status, are all ignored in this approach, a superior treatment could be deemed ineffective simply because the true treatment effect was washed out or diluted (13). This may lead to confusion for the reader when attempting to interpret the results of a study and may lead to the dismissal of potentially efficacious interventions (26).

Quality of Trials

Another common argument in favour of exclusions is that inclusion criteria are not affected by treatment, and so excluding these patients will not bias the results. In other words, as long as eligibility is determined by a blinded reviewer, it is independent of the randomization to treatment process and therefore, exclusions based on eligibility criteria will not be biased in this way. However, Senn writes that "rectification schemes are known to be a poor approach to quality control. In the long-run, allowing the statistician to exclude patients in the analysis will encourage the physician to be less than vigilant in deciding who is included" (28). On the other hand, some investigators argue that ITT permits trialists to become much less stringent with monitoring patient compliance since this data is essentially ignored with this analytic approach (26).

b) Per-Protocol

This approach is also known as adherers only (AO). Among other reasons, this approach excludes patients who did not take a given amount of the treatment drug they were assigned and would also exclude patients who were considered to be ineligible if they no longer met protocol inclusion criteria after they were randomized . Many trials require the patient to receive at least one dose of the treatment drug in order to be included in the analysis. Some trials require a certain percentage of the prescribed dosage to be taken (e.g., 80%). According to Sheiner and Rubin, this approach is concerned with the question "what are the differences between average [experimental] outcomes for patients who choose to adhere to recommended [experimental] treatment and average [control] outcomes for patients who choose to adhere to [control]"(26). In other words, the per-protocol approach explores the treatment effects between a sub-population of trial participants who chose to adhere to their assigned therapy.

Study power

This approach will generally have better power for detecting treatment effects than an ITT approach and serves to detect the "pure" treatment effect in the set of patients who complied with therapy (23), provided the decrease in sample size is not enough to counter-influence the power. Due to these exclusions, a per-protocol (PP) approach also assumes that any information provided by patients later found to not meet inclusion criteria, is irrelevant (28).

For patients with a change in disease state post-randomization, this approach would exclude them based on the reasoning that they are not eligible for the study provided this exclusion is specifically stated in trial protocol eligibility criteria.

Loss of Information

Unfortunately, patients with "potentially informative exposures" may be excluded from the per-protocol analysis because they are censored before the primary endpoint is reached (23). Whenever exclusions are permitted the potential for introducing bias into the study is increased. Just as compliance and misclassification may or may not be random, a change in disease-status may not be random as well, but a result from systematic phenomena caused by unmeasured confounding variables. For these reasons, a per-protocol approach may be excluding patients differentially from the groups. If there are a substantial number of patients who deviate from protocol, a great deal of study power is lost along with the information on the exclusions. For instance, if a patient is assigned a study drug that causes intolerable side effects and chooses to skip dosages or violate the protocol in some other way, the per protocol analysis that excludes this patient makes the drug effectiveness appear better than it really is because the patient who suffered adverse effects without apparent therapeutic benefits is not represented in the analysis. Similarly, by excluding this patient, the per-protocol analysis can also provide a misrepresentation of the safety of the study. In this case, the intolerable side effects experienced by this patient will not be represented in the safety analysis

and the results will only indicate the mild side effects felt by other patients who did not discontinue treatment.

c) Treatment Received

Treatment received (TR) basically analyzes each patient according to which treatment they took over the course of the study.

Study Power

This is not an often-used approach because of its high potential for bias. However, since all randomized patients are included in the analysis, it does not suffer from the problem of maintaining power that plagues the ITT and PP approaches. This approach analyzes patients according to the treatment they received, which may or may not be the treatment to which they were randomized (26). In this way, all threat of dilution of effect due to treatment crossover is removed. Although this approach may not be appropriate for a primary analysis, it has been said that "a carefully defined, appropriate and cautiously interpreted secondary analysis which examines subgroups of patients defined by factors not influenced by treatment can be useful and may offer hypotheses for subsequent evaluation" (25).

In the case of post-randomized ineligible patients, this approach places those patients with a change in disease status from the experimental group into the control group since this change effectively renders the experimental treatment

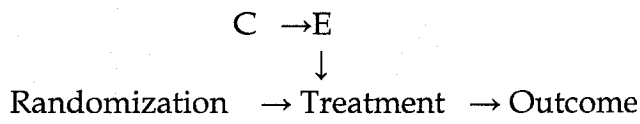
ineffective. It is assumed that these patients received no benefit from the treatment because they no longer had the condition of interest and therefore did not receive the experimental treatment as intended. In the event that the trial is not placebo-controlled, this approach could serve to minimize the true treatment effect because some benefit would be received from the control/conventional therapy. Furthermore, most statistical procedures are based on the assumption of randomization and therefore this approach also undermines the validity of the statistical tests (25).

2.3.5 Statistical Approaches Dealing with Treatment Change

No statistical approach was found that directly concerned the issue of disease status alteration. Approaches were evaluated closely to determine if there were techniques that could be applied to the issue of disease-status alteration .

a) Adjusted Treatment Received

Nagelkerke et al. proposed an adjusted treatment received method. Here, the investigator uses a third variable labeled "E" which represents the effect of the confounding factors on the treatment outcome. "E" is therefore treated as "the" confounder and as indicated by Nagelkerke, "essentially absorbs all of the non-treatment effects which may act as a confounder for the relationship between treatment and outcome" (13). An important property of "E" is that it is orthogonal or uncorrelated to randomization (R).



As indicated above, E represents all causes of T (Treatment) except R. That is, R and E together determine all causes of T (13). Treatment effects are calculated by including two covariates; a dummy variable for treatment and the other being the residual E. From this point it makes no difference if one uses linear, logistic regression, Poisson regression or the Cox proportional hazards model. For our purposes, the "E" would be the change in disease-status caused by unknown confounder C.

b) Instrumental Variable Analysis

The instrumental variable analysis was created as a model-based analysis that incorporates information on both the intended and received treatment (26). According to Sheiner and Rubin "to make intelligent individual therapeutic decisions, it is at least as essential to know potential benefit given full compliance as it is to know expected benefit averaging over rates of compliance in a particular trial" (26). For our purposes, we will use these methods to deal with patients who undergo a change in disease-status which, like compliance, may render the treatment potentially ineffective, regardless of true efficacy. This approach, with some modification, divides the randomized population into those patients who will have the expected disease-status for the experimental treatment (E) when assigned to it and those who will have an altered disease-status and will thus revert to the

control treatment (C) after a short time. Therefore, the changing patients are essentially "assigned" to the control group. In most trials, although the compliance for those assigned to E can be known, it cannot be known whether the control patients would have complied with E treatment had they been assigned it. As a result, the compliance rate of the E-compliers is applied to both the treatment and control group and the two groups are compared to one another, providing an unbiased estimate of method-effectiveness in the sub-population of compliers (26). For our purposes, the disease-status of all patients is available and it will not be necessary to infer the compliance of the control group had they been assigned to the experimental treatment. Next, the approach analyzes those in the treatment group with no bacterial change against with those in the control group with no bacterial change since they would have received bactericidal treatment had they been assigned to it. Thus, our analysis focuses on the efficacy of the intervention in the subpopulation of patients with no change in disease-status. The assumption that the exposure to treatment was short enough to justify the claims that the outcomes of treatment group non-compliers are the same as those in the control group is not enough to justify an as-treated analysis (26). This is mostly due to the likelihood that the treatment non-compliers would experience the control intervention very differently than the treatment compliers would had they been assigned to it. In contrast to the intention-to-treat design, the instrumental variable analysis "encourages designs that carefully monitor patients' experience thereby enhancing the total knowledge available from clinical trials."(26). The papers presenting this

approach assert that this method has the same power as the intention-to-treat method, but provides estimates of treatment effect size which are not diluted by lack of compliance or a change in disease-status, or contamination, as they are in a standard intent-to-treat analysis"(12). However, as stated above, this approach requires additional assumptions compared to standard analytic approaches. The first of these is that the outcomes are known for all patients. Secondly, observed outcomes for E non-compliers are the same as if they were exposed to C rather than E (12). In other words, the observed outcomes for patients with an altered disease-status are the same as those who received the control intervention. This is a very large assumption to satisfy and is unlikely work in practice. In most cases, a patient will receive at least partial benefit from a treatment therapy even if they are not fully compliant. However, in the case of randomized ineligible patients, they may receive no therapeutic benefit at all and subsequently have poorer outcomes than either treatment or control groups. A third assumption is that the rate of compliance in the two groups is assumed to be the same (16). Similarly, it is assumed that the rate of disease-status alteration will be non-differential, if randomization was successful. Therefore, both treatment and control arms would have these subgroups of status-altered patients who would have similar measures of the outcome since in both treatment arms these patients have no expected potential to benefit from the treatment (16).

Aside from these additional assumptions, there is a weakness when this approach is employed in a placebo-controlled trial. As suggested by Sommer,

estimates of biologic efficacy will be biased if there are substantial non-biologic impacts of delivering the treatment regimens such as the "placebo effect" (16). It is not possible to control for differential placebo effects, particularly in those patients with a change in disease-status that renders any treatment ineffective. However, a placebo effect will impact the results of an intention-to-treat analysis as well.

c) Inverse Probability Censoring Weighted

This method was developed as a tool to reanalyze the data from an Acquired Immunodeficiency Syndrome (AIDS) clinical trial. Results for a particular intervention were negative, leading the authors to two possible conclusions. The first was that the intervention truly had no biological effect on survival and the second being that there actually was a beneficial effect but that the trial had insufficient power to detect it using the standard intention-to-treat analysis (15). Consequently, these investigators created a new method for analyzing the data. This approach artificially regards patients as dependently censored the first time a patient voluntarily stops therapy, switches therapy or is lost to follow-up. After developing a model and adjusting for major time-dependent prognostic factors for censoring, each censored patient is assigned an estimate of the conditional probability of remaining uncensored at time "t" that is based on the model. They then apply the inverse probability of censoring weighted (IPCW) Kaplan-Meier estimator and log-rank test of Robins (1993) that adjusts for dependent censoring by utilizing data collected on time-dependent risk factors for failure and censoring (15).

This series of analyses attempts to estimate what the survival curves would have been: if the possibility of crossover to the other treatment arm after the development of the outcome has been eliminated from the treatment protocol; if no subject had voluntarily stopped therapy without medical indication; and if all subjects were forced to stay on their assigned therapy (15). The fundamental assumption of this approach is that there are no unmeasured confounding variables for censoring.

2.3.6 Other Approaches

a) Validation data

In most acute care trials that assess the efficacy of a treatment, a diagnosis prior to randomization and treatment is required (20). As such, disease screening or clinical suspicion is used to determine eligibility for the study. Because diagnosis may be subject to error, some researchers have proposed the use of validation data to estimate the misclassification rate. Investigators are able to incorporate an expected misclassification rate into the sample size calculations to ensure that accurate estimates of treatment effectiveness can be calculated.

This approach is dependent on the availability of appropriate and unbiased classification estimates. The same concept of validation studies could also be applied to patients with changes in disease status that render them ineligible for the trial since they may experience no benefit from the intervention.

According to Lin, Lyles, and Williamson (20), if the screening test for disease status was perfect and not subject to misclassification and the disease could not

resolve on its own, the probability of having a second positive test for the same untreated disease would be equal to one. However, most, if not all tests are not perfect and thus the probability is somewhat less than one. In the same vein, one would think the probability of a placebo-treated patient having a positive test before randomization and after randomization would be equal since each test should be independent. However this is not necessarily the case for two reasons: first, extreme clinical and test result values are subject to regression to the mean (20) and second, in almost all instances, clinical diagnoses of diseases are not black and white. Most often, a continuum of measurable and unmeasurable traits on which a diagnosis is based will range from "clear absence of the disease to clear presence of the disease with an intermediate zone of uncertainty" (17). In diseases where a dichotomous diagnosis must be made based on continuous data, the more extreme symptoms or traits will regress towards the mean in subsequent measurements. Consequently, the patient may be misclassified as disease-free, thereby attenuating the treatment difference (20). Brenner adds that this is especially the case for those patients in the intermediate zone of uncertainty; close to the cut-points for diagnostic classification (17). Consequently, this leads to a positive correlation of diagnostic error if multiple diagnostic measurement strategies are used (17). A common assumption, convenient for statistical purposes but unrealistic for clinical diagnoses is the independence of "multiple test results conditional on the individuals' true status" (17). Unfortunately, little is known about the impact of correlation of errors has on the error rate of diagnosis when based on multiple

tests(17).

In order to apply validation studies, a preliminary scan of the literature is necessary to obtain estimates of rates of compliance, misclassification, change in disease-statues, and other issues related to compliance or misclassification that may influence the final analysis. Following this, explicit provision should be made in the trial protocol for what steps will be taken to deal with these patients. If they are to be excluded, the estimates should be built into the sample size calculations to ensure the desired level of study power is maintained. If these patients are to be retained in the analysis, the minimum important difference or alpha level may have to be adjusted to maintain power.

b) Exclusion Approach

A final approach to patients with a change in disease-status after randomization is simply to exclude them. Some of the papers included in the analysis focused on the issue of post-randomization exclusion of patients with an unconfirmed clinical diagnosis, and the implications of excluding such patients. Fortunately, no bias is incurred so long as the same criteria are used to exclude patients in all treatment arms (6). The criteria should be evaluated by a central adjudicating committee that is blinded to treatment assignment to avoid bias in the process of exclusion (27). If precise criteria are not strictly adhered to, even equal numbers of excluded patients from each treatment arm cannot guarantee valid group comparisons because the reasons for exclusion may differ between groups

(22). Excluding patients is never without some consequence. In the clinical setting, unless definitive eligibility requirements can be ascertained before the first treatment dose, it will be necessary for patients without the condition of interest to take the study drugs. Consequently, by excluding these patients the analysis no longer addresses the effectiveness of the drug in everyone who receives it (6). In addition, "retrospective exclusion of a large number of patients who would not be expected to benefit from the treatment creates a potentially misleading impression of the overall effect" (6).

A summary of the strengths and weaknesses of each approach is provided in Table 2.3.2.

Table 2.3.2 Summary of Approach Strengths and Limitations

Approach	Strengths	Limitations
Intention-to-Treat	• conservative for superiority trials • unknown confounders remain randomly distributed between groups • most realistic and unbiased estimate • naturally incorporates non-compliance, adverse events and other factors • addresses "treatment policy" question	• treatment effect diluted by protocol violations and randomized ineligibles • question of biologic efficacy not addressed • assumes compliance patterns and/or changes in disease-status will be the same in the clinical setting • sensitive to placebo effects if outcomes are soft • anti-conservative for equivalence trials
Per-Protocol	• measures the "pure" treatment effect in those who comply	• very limited validity • can suffer from length-sampling bias • ignores patient information • may have decreased power due to large numbers of excluded patients • comparison groups may not be comparable • unknown confounders not longer randomized between groups
Treatment Received	• addresses question of biologic efficacy • uses all patient information	• comparison groups may no longer be comparable • unknown confounders no longer randomized between groups • limited confidence can be placed in results • power may be lost because of smaller sample sizes • blurs definition of compliance
Adjusted Treatment Received	• can be used with any outcome measure • allows for some measure of biologic efficacy	• requires more thought than ITT
Instrumental Variable Analysis	• same power as ITT without dilution of treatment effects • allows for more realistic confidence intervals to be determined for treatment effects • respects randomization • encourages designs that carefully monitor disease-status, compliance, etc. • uses all available information	• complex formulas • require additional assumptions that may be difficult to determine • difficult for reader to interpret results • not typically used, thus limiting comparability with other similar trial results
Inverse Probability Weighted Censored	• estimates what survival curves would have looked like if not patient crossover • takes compliance/change in disease-status into account • maintains randomization • maintains power	• assumes no unknown confounding variables • highly complex statistical calculations • difficulty in result interpretation • reader difficulty in result interpretation • no other trials for comparison
Validation Studies	• allows one to determine probability of still having same disease-status post-randomization • effective for when binary diagnosis is determined by continuous data • effective when extreme disease-status observations are subject to regression to the mean	• done post hoc which limits options once information is known • require more calculations

2.4 *Discussion*

Utilizing any of these approaches will have implications on the interpretation of a study's results. There are four central implications; power, explanation of conflicting results, comparability and generalizability. Although discussed separately, these issues are intertwined on multiple levels. The discussion will finish with an overview of some recommendations for future studies.

2.4.1 Power

Two of the primary ways to lose study power post hoc are through events that affect sample size and factors that dilute treatment effects. The first of these is usually caused by drop-outs and withdrawals, which are inevitable in almost every study and most investigators enroll more patients than necessary in order to maintain their desired study power. Typically, post-randomization exclusions occur because some unforeseen event has taken place and a number of patients no longer fit the required study inclusion and exclusion criteria, and this can also have a detrimental effect on study results and study power.

As stated, a second way to lose study power is to dilute the treatment effect. Dilution of treatment effect occurs mainly through the inclusion of patients who crossover to the alternative treatment, fail to comply with their assigned treatment, or undergo a change in trial eligibility. Essentially, the issue is the same in all three cases; the interventions are not being received as intended and therefore the ability to detect a true difference between the interventions is compromised.

To further illustrate this issue of power loss, some new and potentially superior diagnostic screening tools and treatment interventions may be abandoned because of an inadequately powered trial. In the MCBT cystic fibrosis trial example, 23% of the patients in the treatment group underwent a change in disease status that rendered the intervention potentially ineffective. Those patients were included in the ITT analysis and no treatment difference was detected between groups. However, due to the amount of dilution, the trial was not adequately powered to detect a difference, even if a true difference did exist. This example serves to show that if no further investigation into trial results is considered, an intervention may be dismissed as ineffective when there was no chance of it being deemed effective from the start.

2.4.2 Explanation of conflicting results

Although conducting an intention-to-treat analysis as well as a sensitivity analysis to tease out the treatment efficacy is perfectly acceptable, inadequately reporting these results raises some important issues; particularly in the event that the two analyses have conflicting conclusions. Confusion can arise if investigators do not provide enough information for the reader to fully understand and interpret the trial results. Readers need to be informed as to the rationale for and limitations of the analytic approaches employed in the study as well as the degree to which the results should impact clinical practice. If the intention-to-treat results conflict with the secondary efficacy analysis, the trial discussion should be focused on explaining

these results and why they are seemingly in opposition. Sheiner and Rubin (26) make the point that confusion is possible if the null hypothesis is not rejected in an ITT analysis. They feel that unless otherwise informed, one does not know what step to take next. If the intervention failed because too many subjects had a change in disease-status then further investigation into the reasons for this should be conducted, but if the intervention failed because it was, in fact, inferior to the other treatment, it should be abandoned. In order to remedy this, investigators should fully report and interpret all presented analyses.

2.4.3 Comparability

Analytic approaches which exclude patients post-randomization potentially limit their ability to compare treatment groups. Even if the same exclusion criteria are used on both treatment groups, it is possible for systematic error to occur. For example, a trial dictates that 80% compliance with study medication must be maintained in order for a patient to be included in the analysis. Within this trial the patients in the treatment group experience severe adverse events and the sickest patients only comply with 45% of the treatment dosage and are consequently excluded from the analysis. Moreover, the healthiest patients in the control group consider themselves well and decide to discontinue study medication after they have only taken 60% of study drugs. Although the same exclusion criteria were applied to both groups, the patients left in the analysis are no longer comparable. The analysis would compare the healthiest patients in the treatment group to the

sickest patients in the control group. Therefore, comparability is vitally important for the validity of test results as well as the generalizability of these results. If the comparability of groups is not maintained, the results are not valid and therefore cannot be generalized to the underlying population.

2.4.4 Generalizability

Excluding patients may also limit the external validity of the trial, or the degree to which the results can be generalized to patients outside of the trial. If a true difference is detected but is only applicable to a small subset of a clinical population, these results cannot be applied to the greater population from which this subset was drawn. Consequently, allowing exclusions may not be the most effective or economic option for therapy.

In order to maximize the amount of information provided by trial results, many investigators perform an ITT analysis, maintaining randomization as thus internal and external validity and subsequently perform a sensitivity analysis on the various subsets of the trial population. The information from a sensitivity analysis, although not generalizable to the entire trial population, can be informative for generating hypotheses for future research.

2.4.5 Economic Implications

Trials are very expensive to conduct and require countless hours of investment by investigators and patients. If follows then, that if a trial is underpowered, these dedicated hours and financial resources are somewhat wasted

when a trial is destined to fall short of its objective from the start. Therefore, determining a trial sample size *a priori* which takes into account the possible changing nature of the disease status and its effects on the results of the study is very important. A proper sample size calculation that accounts for possible post-randomization exclusions, and dilution of treatment effects, will prevent the study from enrolling too few patients. This will also help to prevent repetition of underpowered studies. Re-conducting a trial because of power issues is frustrating and arduous and certainly not cost effective.

2.4.6 Limitations

Although in this review every effort was made to identify all relevant articles, there are some limitations to this search. First, some of the standards for RCT systematic reviews may not apply to methodological searches. For example, only articles written in English were considered for full-text review and those written in other languages (n = 22) were not eligible, however the justification provided by Moher and Juni may not hold true. Second, data abstraction was performed by only one reviewer. Consequently, extractor bias, where information is copied down with a systematic error or inaccuracy, may have been introduced into the review. Finally, there was also relatively low agreement between reviewers in deciding which articles to include in the systematic review due to the somewhat ambiguous inclusion criteria. These loosely defined criteria may have resulted in selector bias, since the criteria alone did not determine which articles would be included and

some aspect of personal choice was present. However, the necessity for the loose eligibility criteria can partly be explained by the paucity of medical literature devoted to this issue. A second reason for the low level of initial agreement was the varying expertise of the reviewers. Some articles were very technical in nature, containing many complex formulae and notation. Consequently, the ability to determine eligibility was varied amongst the reviewers.

2.4.7 Recommendations

In order to choose the approach that best fits the data and produces the most powerful, unbiased estimate of treatment effect, the investigator must possess a certain degree of knowledge regarding potential issues in analysis before the study even begins. Perhaps the best way to obtain this prior knowledge is to conduct a pilot study with a small sample of patients. In doing so, investigators can observe phenomena such as changes in disease-status and estimate the degree to which this may occur in their full-scale study. This will facilitate an *a priori* decision on how these patients will be dealt with, the terms of which should be specified in the study protocol. In doing so, the likelihood of biasing study results by exclusion of these patients is somewhat diminished and study power need not be compromised. Furthermore, investigators should be familiar with the different analytic approaches and their respective strengths and weaknesses.

A more accurate interpretation of study results will be facilitated by a deeper understanding of limitations of various statistical approaches, including the effect

that post-randomization exclusions will have on result comparability. The focus of this thesis is not to advocate for any one approach, but to summarize the various strengths and limitations of each and stress the importance of reporting these alongside ITT results in order to equip the reader to accurately interpret them, particularly if these results are conflicting.

Chapter 3: MCBT Data

3.1. Background

The Multiple Combination Bactericidal Antibiotic Testing (MCBT) trial was designed to test the effectiveness of combination antibiotic susceptibility compared to the current standard of care. February 2005 marked the completion of this highly promising intervention trial for the treatment of Cystic Fibrosis (CF).

Cystic Fibrosis patients suffer from chronic bacterial lung infection. This smoldering chronic infection is punctuated by acute pulmonary deterioration, termed acute CF exacerbation. These acute episodes require immediate antibiotic treatment. Many of these bacterial isolates are resistant to single antibiotics so combinations of two or more are typically prescribed. Problems arise when an additional antibiotic is added to a previously bactericidal combination which can occasionally result in antagonism or growth of the organism as opposed to its destruction. Typically, antibiotics are individually tested on the cultured organisms but this does not provide the necessary information. MCBT is the first trial to test multiple antibiotic combinations on cultured organisms in an attempt to determine which are bactericidal and whether therapy with bactericidal therapy would affect clinical outcomes.

The trial enrolled 251 clinically stable patients from CF centres over an 18 month period. These patients had sputum cultures performed every three months until they experienced an acute exacerbation. The sputum samples from all stable patients underwent conventional culture and sensitivity testing as well as MCBT

testing. Based on these test results, prescriptions were written in the event the patient experienced an acute exacerbation. The MCBT physician wrote an order based on the MCBT results and the CF physician wrote an order based on the conventional test results. In all, investigators collected data on 414 patient variables. These variables included personal characteristics (e.g., age, weight, date of birth), current and past medical history (e.g., previous medication use, medical allergies, and blood cell counts), dates of exacerbation, hospitalization, adverse events as well as details regarding their randomized therapy.

The first 132 patients to experience an acute exacerbation were randomized to either treatment group (MCBT-directed therapy) or to the control group (standard culture and sensitivity-directed therapy) based on the results of their most recent sputum culture. Sputum samples were also taken at the time of the initial acute exacerbation and put through MCBT testing, although the results were not available for ten days. After the patients completed the 14-day course of IV antibiotics they were monitored by monthly phone calls and medical records until a subsequent exacerbation. The primary outcome measure of the trial was time to next exacerbation and was analyzed as a Kaplan-Meier survival analysis using the log-rank test to compare the curves and Cox Proportional Hazards modeling to control for baseline characteristics (5).

The analysis revealed no difference between treatment and control groups; a disappointing result given the rationalized hypothesis. However, there is one issue that may have affected the overall results; a difference in disease-state ascertainment

between pre-randomization results and post-randomization results. MCBT testing is close to 100% effective at determining bactericidal antibiotic combinations. The intriguing issue is that bacteria can mutate and develop altering resistance to antibiotics. As a result, only 77% of those in the treatment group received bactericidal antibiotic combinations because of mutation and alteration of resistance profiles in the patients' bacteria between the time when patients provided sputum and the time point at which they were randomized. As of yet, this is an unavoidable aspect of CF exacerbation treatment. Sputum samples are only collected every three months and require 7 days for conventional lab results and 10 days for MCBT results to become available. Due to the nature of the disease, treatment of an acute pulmonary exacerbation cannot be delayed and drugs are prescribed based on old laboratory results.

When patients in the treatment group presented different bacterial profiles post-randomization, their disease status had effectively changed, and this rendered the intervention potentially ineffective. This deviation phenomenon is seen in trials that require urgent treatment that cannot be delayed In the MCBT trial this phenomenon presumably occurred due to a time-lag between the initial pre-randomization sputum culture and subsequent sputum cultures done on the day of randomization (or acute exacerbation). Issues involving time-lag were also seen in influenza drug trials where influenza diagnosis could not be confirmed until after randomization and treatment had already occurred (3, 8). As a result, patients could be excluded from the analysis as legitimate post-randomization exclusions (6).

Although the actual patient diagnosis did not change in the MCBT trial, 23% of the treatment group patients no longer had colonization of the specific multi-resistant bacteria.

The MCBT trial serves as a clear example of the problem of post-randomization change in disease status. Therefore, this trial data is useful to illustrate the analytic issues and methods associated with post-randomization changes in disease state as well as issues related to study power, bias, ethics and economics.

3.2 Methods

3.2.1 Methods Overview

The methods for this section are twofold. First, the application of the different methods found in the systematic review to the MCBT dataset will be applied, followed by bootstrapping of the baseline variables of the dataset.

Because of the relatively small sample size of the MCBT trial, the major prognostic factors for the trial were bootstrapped in order to evaluate the normality of the data and the precision of the confidence intervals around the point estimates. It also provided insight into the altering bacterial profiles of the enrolled patients. These patients diluted the effectiveness of the MCBT intervention and so, determining if the MCBT population was characteristic of the underlying population of patients with cystic fibrosis and multi-resistant bacteria, could have significant impact.

The theory behind bootstrap statistics is relatively simple. Bootstrapping is a

technique for finding the sampling distribution from one random sample (31). Essentially, a new sample is created from the original random sample by sampling with replacement. The new sample is the same size as the original but will not contain all of the exact values as the original because some values will appear more than once (because of replacement sampling). This process is usually done 1000-5000 times, generating new point estimates for the variable of interest. Because the original random sample represents the population from which it was drawn, if we re-sample from the original random sample we can obtain a representation of what we would have observed if we had taken each of the 1000 new samples from the population. By taking the average of these new point estimates we can have more confidence that our sample mean is an accurate estimate of the actual value in the true population. This is referred to as robustness of the point estimate.

The bootstrap technique is especially valuable for studies with smaller sample sizes. By performing bootstrapping, it is possible to obtain a more robust point estimate and more accurate confidence limits without having to recruit more patients. Of course, if the original sample is flawed or not representative of the population from which it came, then bootstrapping cannot "fix" anything. The entire theory hinges on the quality of the original random sample.

3.2.2 MCBT Exploration Methods

After the systematic review was complete and the identified approaches were evaluated (see Chapter 2), applied to the MCBT data. This process served to

highlight the comparative effect on the survival point estimates, number of excluded patients, measurement of bias, and limitations to internal and external validity. The analyses were conducted with the original MCBT closed data set (5) using SAS version 9.1 English.

To bootstrap the data, the original MCBT database was reduced from 414 variables to those pertinent to this issue and was transferred to an Excel file. In order to obtain bootstrap estimates of the variables, a macro was created in Excel Version 10. Binary variables originally coded as 0 or 1 were recoded to read 1 or 2, with the additional 1 subtracted from final mean estimates. The macro was set to perform the re-sampling function 3000 times on 22 variables. This macro can be viewed in Appendix Bi. The rationale behind the number of replications is founded on the fact that the standard error of the bootstrap estimate approaches the standard error of the original sample as the number of replications approaches infinity (32).

3.3 MCBT Results

3.3.1 Design Allocation Approaches

Table 3.3.1, as well as Figures 3.3.1 through Figure 3.3.3, illustrates the differences in the hazard ratio, p-value, confidence interval, and patient exclusions across the design allocation approaches.

Although all results are non-significant, the direction of effect and confidence intervals follow the expected results for each approach. The ITT hazard ratio is in favour of conventional therapy and is clearly illustrated in Figure 3.3.1 with broad confidence intervals revealed in Table 3.3.1. Likewise Figure 3.3.2 suggests that the

PP hazard ratio also favours the conventional therapy over the MCBT directed therapy, however the confidence intervals also are wide and cross zero. In Figure 3.3.3, the Treatment Received approach reveals a hazard ratio in favour of the MCBT directed therapy. This effect is also not significant, with confidence intervals crossing zero. The code for these data analyses is available in Appendix Bii.

Table 3.3.1 Practical Application of Design Allocation Approaches

Approach	Sample Size	Hazard Ratio	p-value	95% Confidence Limits	# patients excluded from analysis
ITT	132	0.858	0.4031	0.599, 1.229	0
Per Protocol	75	0.959	0.8646	0.589, 1.560	57
Treatment Received	121	1.151	0.4680	0.787, 1.682	11

Figure 3.3.1 Survival Curves for Intention-to-Treat Analysis

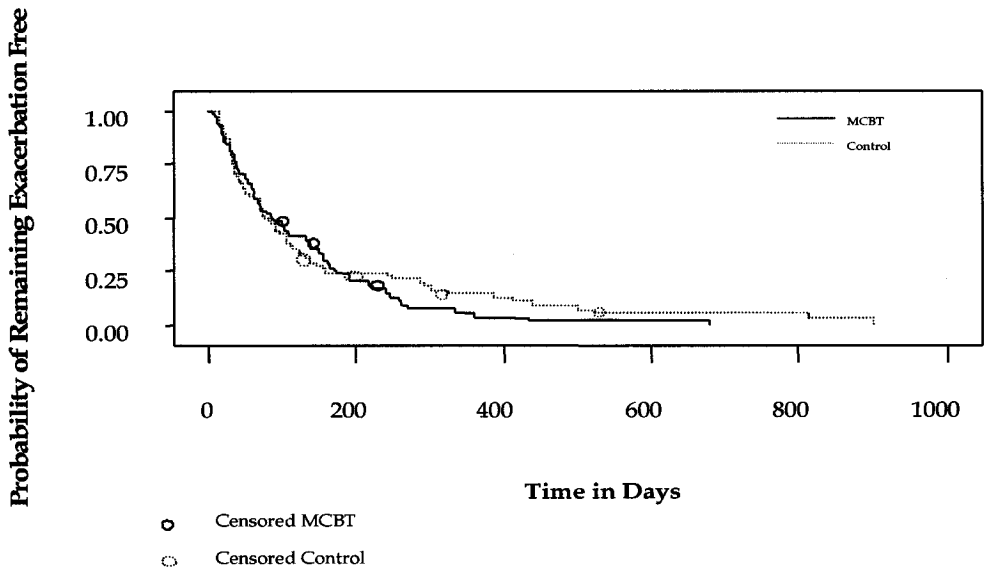


Figure 3.3.2 Survival Curves for Per-Protocol

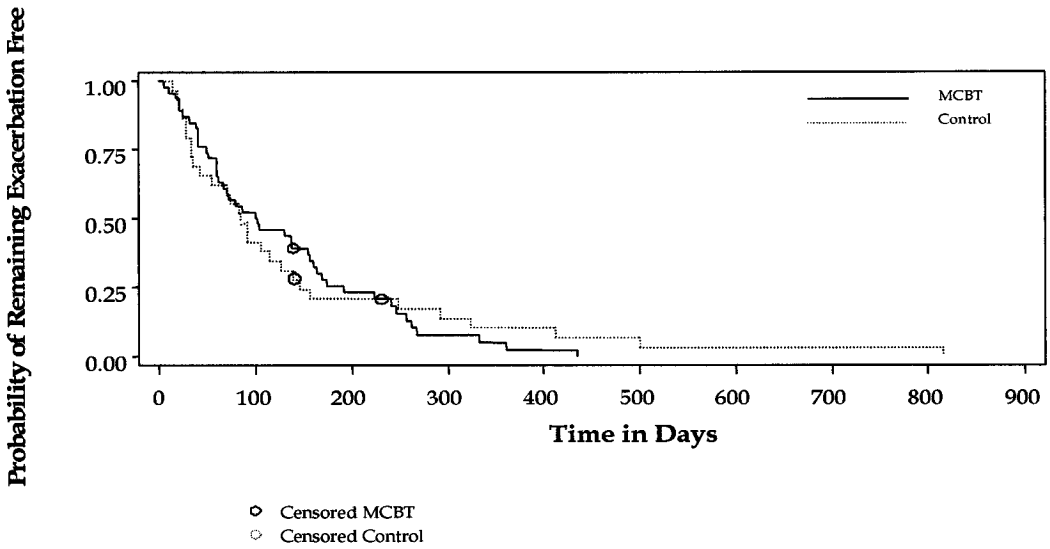
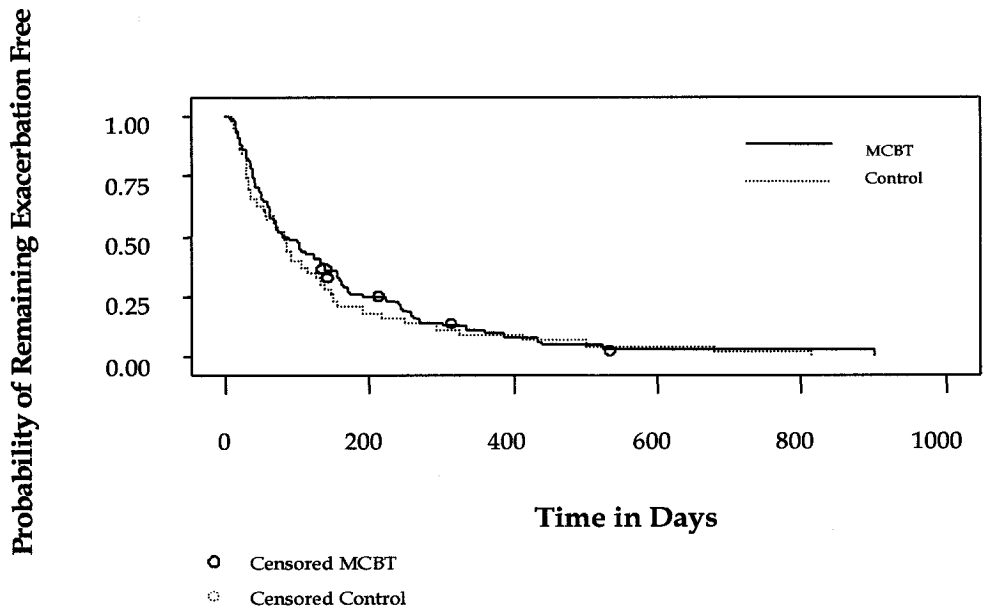


Figure 3.3.3 Survival Curves for Treatment Received



3.3.2 Statistical Techniques for Treatment Change

Instrumental Variable Analysis (IVA) is not appropriate for a survival analysis however it can be used if the outcome is binary. In order to create a binary outcome, two cut points were chosen and the results were generated for patients who exacerbated six months after randomization and twelve months after randomization. A unique feature about the IVA is that not only does it classify individuals by the treatment they actually received but also by the treatment they would have received had they been allocated differently (30). In other words, the approach uses the compliance rate of the experimental group to estimate the compliance rate of the control group had they been assigned to the experimental group. Therefore, the comparison in outcome is between those in the experimental group who complied with the experimental therapy and those in the control group who would have complied with the experimental therapy had they been randomly assigned to it.

Because in most trials, compliance with control treatment cannot be appropriately measured, the IVA approach applies the known compliance rate of the experimental arm to the control arm. In so doing, bias associated with one treatment being more tolerable than the other is avoided. However, the MCBT trial patients suffered a change in bacterial profile not related to treatment, which was measured in both treatment arms, therefore it was not necessary to infer the rate of bacterial change. The MCBT trial data was still protected from bias because the origin of the change was not related to the treatment and so randomization should

have ensured this variable to be equally distributed among the treatment groups.

The first assumption of the IVA is that "compliance" is an all-or-none phenomenon (26). This assumption is satisfied as the MCBT patients experienced bacterial profile change in this fashion. Second, it is assumed that the observed outcome for non-compliers in the treatment arm is the same as those in the control arm (26). This assumption is not so easily satisfied for the MCBT data. This trial was designed in a way that treatment arm patients received bactericidal treatment and those in the control were assumed to have received empirical treatment which may have been bactericidal or non-bactericidal by chance. Therefore, when those in the treatment arm encountered a change in bacterial profile, they received the conventional treatment. In contrast, when those in the control arm encountered a change in bacterial profile, this did not affect their probability of receiving bactericidal or non-bactericidal treatment. Satisfaction of this second assumption is essential if the rate of bacterial profile change is to be inferred. However, as mentioned previously, this data was collected during the trial. Table 3.3.2 provides the notation for the following calculations.

Table 3.3.2 Notation for 2x2x2 table of MCBT results

	Control Bacterial Profile Constant			Treatment Bacterial Profile Constant		
Exacerbation	No	Yes		No	Yes	
No	m_{00}	m_{01}	m_0	n_{00}	n_{01}	n_0
Yes	m_{10}	m_{11}	m_1	n_{10}	n_{11}	n_1
Total			M			N

Table 3.3.3 Results of the MCBT Trial at 6 Months Post-Randomization

	Control Bacterial Profile Constant			Treatment Bacterial Profile Constant		
Exacerbation	No	Yes		No	Yes	
No	8	6	14	3	11	14
Yes	25	22	47	11	35	46
Total	33	28	61	14	46	60

*11 observations were missing data on bacterial profile post-randomization and were therefore excluded.

Table 3.3.4 Results of the MCBT Trial at 12 Months Post-Randomization

	Control Bacterial Profile Constant			Treatment Bacterial Profile Constant		
Exacerbation	No	Yes		No	Yes	
No	5	3	8	1	1	2
Yes	27	26	53	13	45	58
Total	32	29	61	14	46	60

*11 observations were missing data on bacterial profile post-randomization and were therefore excluded.

The formula is expressed by equation 3.1 (16). Because we do not need to infer control cells m_{11} and m_{01}, the relative risk can be used to estimate efficacy.

$$(3.1)\ RR = [n_{11}/(n_{01} + n_{11})]/[m_{11}/(m_{01} + m_{11})]$$

From this equation we see that R is simply the ratio of the observed to expected number of profile changes in the compliant subgroup. The formula for the 95% confidence interval for the relative risk is provided in equation 3.2 (16). The results of the instrumental variable analysis are presented in Table 3.3.5.

(3.2) Standard Error of lnRR = $\sqrt{\frac{(n_{01}/n_{11})}{(n_{11}+n_{01})} + \frac{(m_{01}/m_{11}}{(m_{11}+m_{01})}}$

Confidence Interval

$\ln(RR) \pm 1.96 \times SE_{\ln(RR)}$

Table 3.3.5 Practical Application of Instrumental Variable Analysis

	Risk Ratio	95% Confidence Interval	# of Patients Excluded
6 Months	0.9685	0.74, 1.27	57
12 Months	1.0911	0.96, 1.24	57

Although the risk ratios are not significant, the confidence intervals are much narrower than the per-protocol approach.

The inverse probability of censoring weighted log-rank tests was a tool developed to accurately attribute the trial deaths from AIDS to the appropriate treatment group after the development of PCP - the primary outcome (15). Four different approaches were examined, each using different criteria for dependent censoring of patients. The first was similar to an ITT analysis, attributing all deaths as failures in the assigned groups. The second analysis censored patients by the minimum of time to loss of follow-up and the third by the minimum of time to loss of follow-up as well as time to treatment-crossover. Finally, the fourth analysis censored patients by the above two criteria as well as the time to voluntarily stopping therapy. This step-wise approach was applied to the MCBT data and presented in Table 3.3.6. Normally after this step, a set of time-dependent prognostic factors for dependent censoring would be entered into the survival

analysis model as covariates. Each patient who is censored is assigned a weight that is inversely proportional to an estimate of the conditional probability of remaining uncensored at time "t". This estimate is based on the survival model. This approach is applied under the assumption of no unmeasured confounders for censoring. Normally, this approach would yield a hazard ratio somewhere in between the PP approach and the ITT approach; including the known information of the censored patients and excluding them only at the time of censoring.

Table 3.3.6 Inverse Probability Censoring Weighted Log-Rank Tests

Step	Sample Size	Hazard Ratio	p-value	95% Confidence Interval	# of Patients Excluded
1: No patients excluded	132	0.858	0.4031	0.599, 1.229	0
2: Censored by minimum time to loss of follow-up	132	0.858	0.4031	0.599, 1.229	0
3: Censored by 2 and treatment crossover	75	0.959	0.8646	0.589, 1.560	57
4. Censored by 2, 3 and voluntary stop	75	0.959	0.8646	0.589, 1.560	57

3.3.3 Unapplied Approaches

The Adjusted Treatment-Received approach could not be applied to the MCBT data. Although the approach appeared applicable, there was no instruction as to how the residual "E" was measured nor how to separate those effects on treatment caused by "E" and those caused by "R". The information for the practical application of this approach was sought through a Medline search for other articles by the author and also for articles referring to the Adjusted Treatment Received

method. Three articles were found yet none of these provided any insight or details into the statistical procedure.

Validation data, although an excellent suggestion, is not feasible for post-hoc analyses. As the literature surrounding cystic fibrosis and its variable nature grows, predicting the amount of bacterial change in each group and accounting for this in sample size calculations will become a standard procedure, in all likelihood.

A summary of the approach point estimates and confidence intervals is provided in Table 3.3.7.

Table 3.3.7 Summary Table of Applied Approaches

Approach	Sample Size	Hazard Ratio	P-value	95% Confidence Interval	Direction of Benefit
ITT	132	0.858	0.4031	0.599, 1.229	Conventional
Per Protocol	75	0.959	0.8646	0.589, 1.560	Conventional
Treatment Received	121	1.151	0.4680	0.787, 1.682	MCBT
IPCW					
Step 1	132	0.858	0.4031	0.599, 1.229	Conventional
Step 2	132	0.858	0.4031	0.599, 1.229	Conventional
Step 3	75	0.959	0.8646	0.589, 1.560	Conventional
Step 4	75	0.959	0.8646	0.589, 1.560	Conventional
IVA		**Risk Ratio**			
6 months		0.9685		0.74, 1.27	Conventional
12 months		1.0911		0.96, 1.24	MCBT

3.3.4 Bootstrapping Results

The bootstrapping results from the MCBT data are shown in Figures 3.3.8. These results provide information on the shape, bias and spread of the sampling distribution (31). According to the central limit theorem, if n is large, the sampling distribution of means will be approximately normal. Because re-sampling was done 3000 times, the bootstrap distribution was nearly normal and close to the expected

shape of the sampling distribution (31). The bootstrap distribution means were close to those of the original sample, indicating that the bootstrap distribution did not have much bias as an estimate of the original sample. Finally, the bootstrap distribution provides information on the spread of the data. The bootstrap standard error used to calculate the 95% confidence intervals are made up of the standard deviations of all 3000 re-samples (31). Bootstrapping does not appeal to the central limit theorem to determine normality; this is done visually. In most cases the distribution will be roughly normal in shape and approximate the sampling distribution however the distribution will be centered on the mean of the original sample as opposed to the parameter value (31).

Most of the pertinent information from this table is in the confidence intervals. The standard error was used to calculate the confidence intervals and distributions of each bootstrapped variable so that the precision of the point estimate was measured, and not the distribution of values. In addition, non-symmetrical confidence intervals were calculated for the bootstrap distribution to adjust for bias and skewness (31). Calculating the confidence intervals in this way also allowed us to evaluate the variation in the mean across numerous baseline variables, as well as estimate the normality of the data, based on the original sample. In turn, this assisted us in determining whether the original sample was normal enough to allow for approximation to the normal distribution. This allows us to determine if statistical tests, which assume a normally distributed sample, can be applied to our original sample.

From Table 3.3.8 the proportion of patients with an altered bacterial profile in the original MCBT trial (35.6%) was within 1.5% of the bootstrap estimate (34.2%). To illustrate the differences in the shape of the distribution for continuous variables, the original MCBT distribution was plotted against a normal distribution generated by the mean and standard error in Figures 3.3.4 through 3.3.7.

Table 3.3.8 MCBT Bootstrap Results of Baseline Characteristics

	MCBT Trial Population	Bootstrapped Results	Difference
Age **	27 (25.73, 28.37)	27.04 (25.05, 27.34)	0.4
Sex (male/female)	55.0% male	53.67% male (51.52, 66.86)	1.33%
Infecting organism*			
Stenotrophomonous maltophilia	8.3%	8.26% (3.79, 12.88)	0.4%
Achromobacter xylosoxidans	6.8%	9.43% (3.03, 11.36)	2.63%
Burkholderia cepacia complex	40.9%	49.23% (34.85, 51.53)	8.33%
Pseudomonas aeruginosa	62.1%	62.13% (51.52, 68.18)	0.03%
Change in infecting organism	35.56%	34.15% (24.38, 40.87)	1.41%
Baseline Lung Function			
FEV1 in L **	1.63 (1.51, 1.75)	1.63 (1.46, 1.70)	0
FEV1% predicted **	47.69 (44.33, 51.04)	47.42 (44.37, 51.06)	0.27
FVC in L **	2.74 (2.57, 2.90)	2.74 (2.57, 2.92)	0
FEV1/FVC% **	59.21 (57.14, 61.28)	59.10 (56.40, 60.01)	0.11
Lung Function on day of exacerbation			
FEV1 in L **	1.44 (1.33, 1.56)	1.44 (1.34, 1.56)	0
FEV1% predicted **	41.85 (38.72, 44.98)	41.68 (41.61, 48.27)	0.17
FVC in L **	2.44 (2.27, 2.61)	2.44 (2.22, 2.56)	0
FEV1/FVC% **	59.02 (56.93, 61.11)	59.03 (56.65, 60.80)	0.1
Oxygen saturation **	94.08 (93.44, 94.71)	94.01 (92.66, 94.19)	0.07
Medications used chronically			
Inhaled tobramycin	63.64%	64.09% (47.81, 62.12)	0.45%
Inhaled colistin	18.18%	18.29% (16.67, 29.55)	0.11%
Oral antibiotics	40.91%	40.91% (34.85, 50.76)	0%
Co-morbid illnesses			
Diabetes	21.21%	21.03% (18.18, 31.82)	0.18%
Pancreatic insufficiency	96.97%	96.90% (94.70, 100)	0.07%
Liver disease	10.61%	10.73% (12.12, 20.45)	0.12%

* Some patients were infected with more than one species of bacteria
** Brackets indicate 95% confidence intervals

Figure 3.3.4 Age

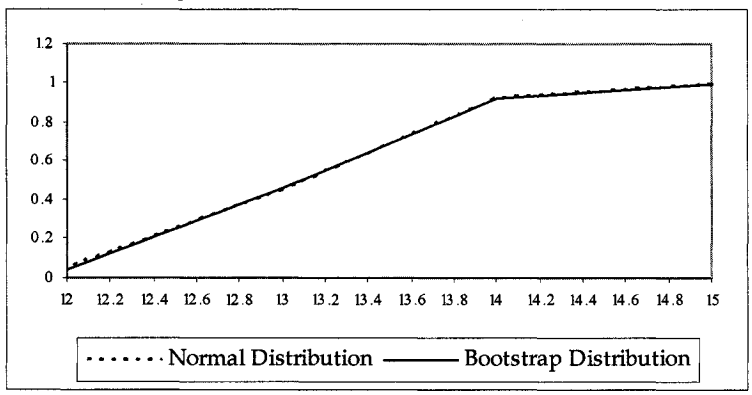


Figure 3.3.5 Baseline FEV1

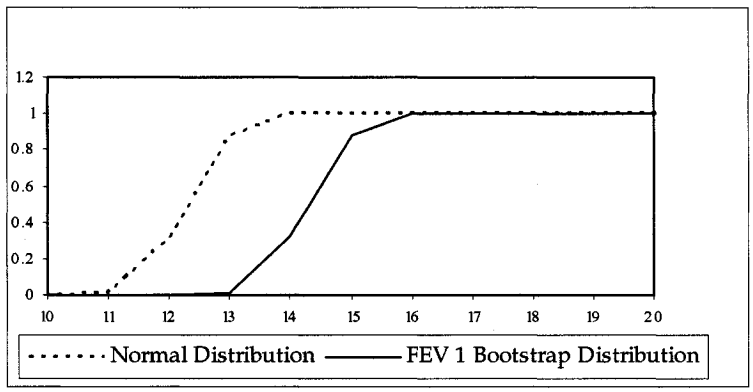


Figure 3.3.6 Baseline FVC

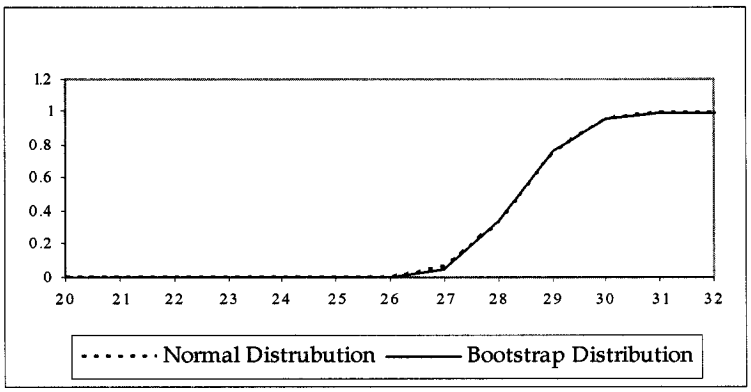
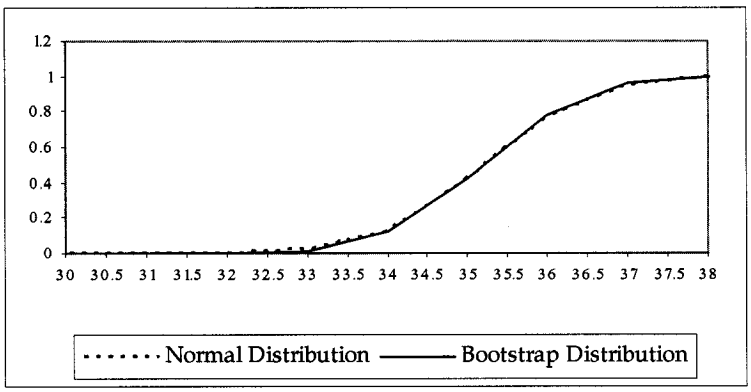


Figure 3.3.7 Baseline FEV1/FVC%



3.4 Discussion

In general, the design allocation approaches provided the most information about the MCBT patients, despite the lack of significance of any treatment effect. The direction of benefit in favour of conventional testing and the broad confidence interval for the ITT approach were likely due to the high amount of dilution in the data and substantial within-group variability. In the PP approach, the direction also favoured conventional testing yet this was contrary to what we would expect given that the dilution from the ineligible patients had been removed. Rather, the true efficacy of therapy had likely been masked by the substantial decrease in sample size. This decrease in effective sample size reduced the power, and therefore despite the reduction in within-group variability, the confidence intervals remained broad. The TR treatment effect was also not significant, however it did suggest that the MCBT-directed therapy produced longer time to event intervals. In so doing, this approach illustrated the merits of reducing within group variability while retaining sample size.

Despite the 57 exclusions using the IVA approach, the power was retained more effectively in this approach because of the reduction in within group variability. The per-protocol approach provided us with an average rate of exacerbation over the entire given trial period, whereas the IVA provided us with risk estimates for exacerbation at two points in time.

The IVA also provided confirmation of the per-protocol results. Although the IVA provided two snapshots of exacerbation rates, the per-protocol survival

analysis provides substantially more information regarding the exacerbation rates in each group. As a result, a considerable amount of valuable information was lost in converting continuous outcome survival data to a binary outcome at two points in time. Also, this approach does not retain any more patients in the analysis than the per-protocol approach. However, the full usefulness of this approach was not explored in our application since we did not need to employ the IVA formulae for inferring compliance rates or for calculating a relative risk with the inferred rates.

The Inverse Probability Censoring Weighted log-rank approach did not work well with the MCBT trial data because there were no patients lost to follow-up, nor were patients on a therapy which could be voluntarily stopped when a change in disease-status occurred as is reflected in the results from Table 3.3.6. In fact, because patients' disease-status changed before treatment was initiated those who did undergo a change in bacterial profile were dependently censored before any time was accrued in their assigned treatment group. Therefore, no further information was gleaned by attempting to use this approach that was not already available from the per-protocol approach to the MCBT trial.

None of the approaches revealed a significant difference between treatment arms for the MCBT trial. The design allocation approaches provided the most information on the time-to-event outcome although the per-protocol approach and treatment received approach are not appropriate for a primary analysis due to the pitfalls of breaking the randomization of unknown confounders. As mentioned, they both provide useful information if they are used as sensitivity analyses

alongside the ITT approach. The instrumental variable analysis did not provide any information other than what was presented in the per-protocol analysis. However, for investigating the effects of compliance on treatment effects for trials measuring binary outcomes, the IVA is an approach appropriate for a sensitivity analysis, provided the additional assumptions can be met.

The IPCW approach, although not appropriate for the MCBT dataset, holds promise for trials dealing the compliance or other issues that cause dilution of treatment effects post-randomization and have time-dependent prognostic factors for censoring. Unfortunately, for the issue of a post-randomization change in disease status, this approach will not provide any information other than what is provided by a per-protocol approach.

There are three potential explanations for why the MCBT clinical trial yielded null results. 1) In all likelihood, as previously discussed, the study was significantly under-powered; 2) The intervention may be truly ineffective. There is some evidence that bactericidal therapy is not a strong predictor of time to next exacerbation (5), and therefore the MCBT test may not be useful to improve clinical outcomes; and 3) The MCBT trial population was not normally distributed and so statistical analyses based on the assumption of normally distributed data would not be valid or applicable. The bootstrapping results indicated there was no pivotal issue.

As indicated by Figures 3.3.4 through 3.3.7, as well as Table 3.3.8, most of the baseline variables of the MCBT data closely follow a normal distribution. In this

way, it is appropriate to take on the assumptions necessary to use statistical procedures based on an approximation to the normal distribution. Moreover, the proportion of change in bacterial profiles was so similar between the original data and the bootstrap estimates, that one can be confident in stating that despite the small MCBT trial sample size, the diagnostic variables of the patients were normally distributed throughout the sample. Therefore, the trial's inability to detect a difference between MCBT treatment and conventional treatment was not due to a misrepresentative sample of patients.

The systematic review and application of the findings to the MCBT trial emphasizes the relative scarcity of options available to a researcher if they anticipate or encounter a sizeable number of ineligible patients post-randomization. In the context of the MCBT trial, the three design allocation approaches to data analysis yielded almost identical point estimates and confidence intervals.

In conclusion, although there appear to be several useful techniques for salvaging data when compliance or misclassification is an issue, these techniques do not easily transfer to the problem of post-randomization ineligible patients. However, the techniques did influence the direction of benefit and the associated 95% confidence intervals.

Chapter 4: Influenza

4.1 Background

Trials assessing the therapeutic effectiveness of influenza antiviral drugs provide an excellent example of the dilemma of post-randomization ineligible patients. Patients enrolled presented influenza-like symptoms yet a significant proportion were found later to be suffering from some other ailment. Unfortunately, confirmation of influenza virus by conventional means can take a week or more; far too long to wait for treatment. Consequently, all patients with influenza-like symptoms were enrolled and took study medication (or placebo); however those without a true influenza virus were excluded from the analysis once laboratory results were available. Patients confirmed not to have flu had a change in disease status - or rather suspected disease status and became ineligible for the trial. Although this change is different from that in the MCBT trial where patients still suffered from Cystic Fibrosis but simply had a change in bacterial profile, this type of change produces the same analytic issue. During a clinical trial, any time treatment is initiated based on an expected disease and the diagnosis changes thus rendering the treatment ineffective, randomized ineligible patients result. The exclusion of these patients in the analysis is referred to as an intention-to-treat-infected (ITTI) analysis (4), sometimes called modified ITT. It utilizes many of the same principles as the intention-to-treat analysis, making no exclusions based on compliance or protocol adherence however it applies these principles only on those patients infected, excluding all others. Influenza serves as a particularly relevant

example of the underlying issues.

4.1.1 Introduction

There is a growing consensus that the world may be overdue for a flu pandemic, possibly due to avian influenza strain H5N1. The World Health Organization reports 186 confirmed cases of H5N1 avian influenza and 105 resultant deaths from 2003 to March 24, 2006 (33). However, even without an outbreak of pandemic influenza this season, epidemics of seasonal influenza occur every winter in North America and are responsible for considerable morbidity and mortality of the population (34) as well as substantial healthcare cost (35). Those particularly at risk for influenza are the elderly as well as anyone who is living with chronic medical conditions or in residential institutions (34). The median duration of illness for influenza is 3 days, however some symptoms like cough and malaise can persist for more than two weeks and significantly affect direct health care costs, losses in worker productivity, and quality of life (36). Other more serious complications associated with the flu include pneumonia, and chronic respiratory disease (36). There are two main lines of defence against the flu; prevention via vaccination and treatment via anti-viral therapy (36). The two most recent anti-viral drugs to appear on the market are nebulised zanamivir, an inhaled neuraminidase inhibitor, and oral oseltamivir, also a neuraminidase inhibitor (36).

The classical influenza symptoms (fever, cough, myalgias, and headache) are common to a whole host of other seasonal viral infections (37). Although a rapid

diagnostic test for influenza exists, its sensitivity and specificity remain unknown and have yet to be evaluated against traditional laboratory methods and therefore have not yet been extensively used. These tests can be performed in the physician's office and yield results in 10 to 30 minutes (38). Conventional laboratory diagnosis is seldom performed on patients and rarely done to confirm influenza infection in the community because of the time required to culture and identify organisms (38). Therefore, patients are still enrolled and randomized to experimental treatment or control in a clinical trial based on a somewhat nebulous clinical or empiric diagnosis of the illness. As a result, a sizable proportion of the patients enrolled in clinical trials with influenza-like-illness will not have the flu and therefore not benefit from therapy (39). This raises a few notable concerns such as the ethics of reporting only the drug effectiveness of the infected (ITTI) population, the potential risk of consuming drugs that are known to have no beneficial effect on an uninfected population, and the potential for adverse medication-related effects in the uninfected population who hold no hope of benefit. Finally other important ethical considerations would include the economic impact of the misuse of anti-viral drugs and the shortage of supply in the event of a pandemic due to the consumption of drug supply by uninfected patients.

The objective of this chapter was to identify the randomized controlled trials conducted on two influenza antiviral therapies, zanamivir and oseltamivir, and determine how the randomized but uninfected patients were handled in the primary analysis. Another objective was to assess the difference in the primary

outcomes between the ITT and ITTI populations and the implications of this difference as a practical illustration of the problem posed by post-randomization ineligible patients.

4.2 *Methods*

4.2.1 Search Strategy

We conducted a systematic search of Medline and Embase databases (Medline, October 28, 2005) (Embase, December 12, 2005) implementing the Dickerson Randomized Controlled Trial filter and the Embase RCT filter, respectively. The search term influenza was then used and followed by either oseltamivir or zanamivir. The initial search was not limited by publication status or language. The results were however, limited to English at full-text review. The complete Medline search strategy is available in Appendix Ci and Embase search strategy in Appendix Cii. In addition, review papers were downloaded to allow for reference checking and secondary source information. Content experts were not contacted however the FDA website as well as the GlaxoSmithKline database were searched for additional trials.

4.2.2 Study Selection

Study selection was based on study design and study objective, namely, randomized controlled trials that investigated the effectiveness of oseltamivir and zanamivir for treatment of naturally acquired influenza. The primary outcome of interest was time-to-alleviation of symptoms. Influenza trials that were not

randomized controlled trials were excluded, as were those using influenza drugs other than oral oseltamivir or zanamivir. Finally, trials involving influenza prophylaxis or treatment of experimental influenza (i.e. treatment of patients deliberately infected with influenza in a laboratory setting) were also excluded since both of these trial types were not affected by post-randomization exclusions of uninfected patients. Abstracts were reviewed by three independent reviewers (DD, DF, SA). Full-text review was completed by two independent reviewers (DD, DF).

4.2.3 Data Abstraction

Data abstraction was done by one independent reviewer (DD) using a pre-formatted data collection form available in Appendix Ciii. This information was put into an Excel spreadsheet. For studies with missing information, attempts were made to contact the authors or the funding organization for the required data. Information on adverse events was not initially collected and so does not appear on the data abstraction forms. The decision to abstract this information was made prior to any analysis.

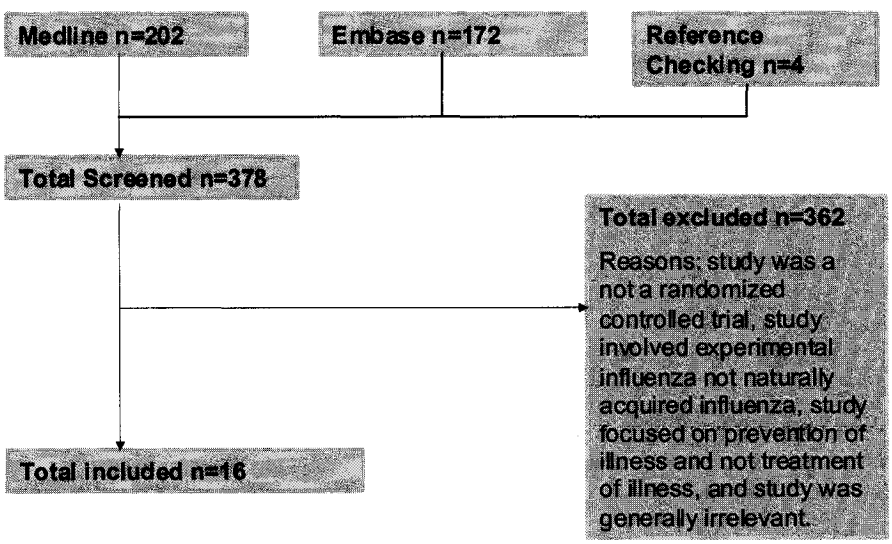
4.2.4 Statistical Analysis Plan

Following data abstraction, heterogeneity will be assessed visually and with the I^2 statistic. If homogeneity could be determined then a fixed effects weighted mean difference model would be used; otherwise a random effects model would be used.

4.3 Results

The Medline search yielded 202 potentially relevant studies. Of these, 8 zanamivir trials (2, 8, 39-44) and 5 oseltamivir trials were identified as eligible (3, 4, 45-47). The Embase search did not yield any additional unique trials. A flowchart illustrating trial inclusion is in Figure 4.3.1. Searching reference lists and databases yielded 1 published (48) and two unpublished zanamivir trials (NAIB2007, NAIA3002) and one unpublished oseltamivir trial (WV15730). Trials were excluded for the following reasons: study was generally irrelevant, study was not a randomized controlled trial, study did not have primary outcome of interest, study was not on naturally acquired influenza, or the study was on prevention of influenza and not treatment. In all, 16 trials were included in the review; 8 zanamivir trials and 5 oseltamivir trials.

Figure 4.3.1 Flow Diagram of Included Influenza Trials



A summary of the characteristics of each individual trial is provided in Table 4.3.1. The primary outcome of each trial was median time to alleviation of symptoms. It is important to note that some trials excluded patients if they had received an influenza vaccine in the past 12 months, and others if vaccinated for the current season. Still, others did not exclude for influenza vaccination if the patient had a positive test for influenza virus.

For each of the zanamivir trials, patients received 10mg of medication or lactose powder placebo via Diskaler twice daily for five days with the exception of Monto et al. where patients also received 6.4 mg intranasal zanamivir or placebo twice daily. Oseltamivir trial patients who were 12 years and older received 75mg of medication or identical placebo twice daily for five days. In the two trials involving children, oseltamivir or placebo was given in 2mg/kg dosages of syrup twice daily for five days.

Table 4.3.1 Characteristics of Relevant Studies

Author, Year, and Country	Source of Funding	Primary Outcome	Definition of Influenza-Like-Illness	Study Drug	Age	Exclusion Criteria	Chronic Conditions	Study Duration
Johnston, 2005, United Kingdom	F. Hoffman-LaRoche	Time to freedom of illness	Temperature >38.7 plus 1 respiratory symptom	Oseltamivir	6-12	Not declared	Asthma	4 weeks
LI, 2004, China	None Declared	Time from symptom onset to time of symptom resolution	Fever >37.8 and two additional symptoms from list	Oseltamivir	18-65	Excluded if vaccinated less than 12 months before study	None	20 days
Nicholson, 2000, Canada	F. Hoffman-LaRoche	Time from start of study drug to time of relief of symptoms	Fever >38 plus 1 respiratory symptom and 1 constitutional symptom	Oseltamivir	18-65	Excluded if vaccinated in past 12 months	None	20 days
Treanor, 2000, United States	F. Hoffman-LaRoche	Time from study drug initiation to time of alleviation of symptoms	Temperature >38 and 1 respiratory symptom and 1 constitutional symptom	Oseltamivir	18-65	Excluded if vaccinated within 12 months prior to study	None	15 days
Whitley, 2001, United States	F. Hoffman-LaRoche	Time from start of treatment to resolution of temperature, cough/congestion and resumption of pre-illness activities	Fever >37.8 and at least 1 respiratory symptom	Oseltamivir	1-12	Vaccination not used as exclusion criterion	None	28 days
Boivin, 2000, Canada	Glaxo-Wellcome	Time to alleviation of symptoms	Fever >37.8 (37.2 for over 65 yoa) and at least 2 more symptoms from list	Zanamivir	12 and older	Not stated	Elderly, respiratory, cardiovascular and renal diseased patients included	Not declared
Hayden, 1997 United States	None declared	Time to alleviation of symptoms	Presence of fever and at least two more symptoms from list	Zanamivir	18 and older	Patients vaccinated for current season were excluded	None	10 days
Hayden, 2000, United States	Glaxo-Wellcome	Proportion of families with at least one initially healthy member in whom symptomatic influenza developed	Fever >37.8, feverishness, cough, headache, sore throat, myalgia	Zanamivir	5 and older		None	Not declared
Hedrick, 2000, United States	Glaxo-Wellcome	Time to alleviation of symptoms	Fever > 37.8 and no clinical evidence of bacterial infection	Zanamivir	5-12	If vaccinated for current season must have positive rapid test	None	14-28 days
Makela, 2000, Europe	Glaxo-Wellcome	Time to alleviation of symptoms	Fever >37.8 (37.3 for over 65 yoa) and at least to more symptoms from list in last 24 hours	Zanamivir	12 and older	No exclusions based on vaccination	None	12 days

Author, Year, and Country	Source of Funding	Primary Outcome	Definition of Influenza-Like-Illness	Study Drug	Age	Exclusion Criteria	Chronic Conditions	Study Duration
MIST group, 1998, Australia	Glaxo-Wellcome	Time to alleviation of symptoms	Fever 37.8, feverishness or both and at least two more symptom from list	Zanamivir	12 and older	Not excluded provided they met entry criteria	None	28 days
Monto, 1999, United States	Glaxo-Wellcome	Time to alleviation of symptoms	Feverishness and two more symptoms from list	Zanamivir	13 and older	Vaccinated patients had to have positive confirmation of infection prior to dose 1	>65, chronic illness including CV, respiratory, endocrine, and metabolic diseases included	21 days
Murphy, 2000, United States	Glaxo-Wellcome	Time to alleviation of clinically significant symptoms	Fever >37.8 and two more symptoms from list	Zanamivir	13 and older	No exclusions based on vaccination status	Asthma and Chronic Obstructed Pulmonary Disease	24 days
NAIA 3002, United States	Not reported	Time to alleviation of symptoms		Zanamivir		Not declared	None	
NAIB 2007, Australia	GlaxoSmithKline	Percent of patients with symptom alleviation by day 4		Zanamivir	13 and older	Not declared	None	10 days
Puhakka, 2003, Finland	GlaxoSmithKline	Time to alleviation of symptoms	Fever >37.8 and two more symptoms from list	Zanamivir	17-29	Not declared	Participants in very good physical health	28 days
WV15730, Australia	F. Hoffman-La Roche	Time to alleviation of symptoms		Oseltamivir	18-65	Not declared	None	

4.3.1 Statistical Analysis

From the provided data, differences in median time to alleviation of symptoms were calculated for the intention-to-treat population and the efficacy population. A meta-analysis could not be performed because individual study mean and standard deviation of time to recovery could not be obtained. Unfortunately only the group medians were reported in the literature without mention of the proportion of patients recovered by certain points in time. No formal attempt was made to obtain the quantitative information on means and standard

deviations for the primary outcome. Performing a meta-analysis was not the objective of this review and given time and resource constrictions the information was not pursued.

Unfortunately, not all of the data required for a comparative analysis were reported in the relevant trials. Six trials failed to present the intention-to-treat (ITT) treatment effect between drug and placebo or even the numbers needed to calculate the ITT effect (41, 47, and 49, 50-52). Therefore the ITTI analysis was left with no comparator. Careful and prolonged searches were made which involved contacting authors and funding organizations. Despite these efforts, these six trials with incomplete quantitative data were excluded from the analysis.

Table 4.3.2 Quantitative Data from Included Studies

Author, Year, Country	Drug	# Randomized	Reported ITT treatment effect	p-value	# Infected	Reported ITTI treatment effect	p-value	# excluded
LI, 2004, China	Oseltamivir	451	0.18	0.0355	273	0.14	0.0466	178
Nicholson, 2000, Canada	Oseltamivir	476	0.77	0.05	319	1.21	0.02	157
Treanor, 2000, United States	Oseltamivir	419	0.86	0.004	253	1.33	0.000	166
Whitley, 2001, United States	Oseltamivir	695	0.88	0.0002	452	1.5	0.000	243
Hayden, 1997 United States	Zanamivir	276	-0.7	0.04	174	-0.8	0.05	102
Hedrick, 2000, United States	Zanamivir	469	0.5	0.011	346	1.25	0.001	125
Makela, 2000, Europe	Zanamivir	356	2.5	0.001	277	2.5	0.0001	79
MIST group, 1998, Australia	Zanamivir	455	1.5	0.011	321	1.5	0.004	134
Monto, 1999, United States	Zanamivir	841	1	0.012	481	1.5	0.11	361
Murphy, 2000, United States	Zanamivir	525	1.0	0.123	312	1.5	0.009	213
Puhakka, 2003, Finland	Zanamivir	588	0.5	0.166	435	0.33	0.08	153

Overall, four oseltamivir trials and seven zanamivir trials were included in the analysis as illustrated in Table 4.3.2. All oseltamivir trials and five zanamivir

trials revealed a significant benefit of study medication over placebo among the ITT population. Greater effects were seen in the ITTI population (efficacy population) where all but one trial showed a significant benefit of study drug over placebo. All trials excluded substantial proportions of their study populations.

A formal meta-analysis could not be performed because in all cases, the investigators chose median time to alleviation as their primary outcome. Median times cannot be pooled due to the lack of standard deviations and confidence intervals unless individual patient data is provided. Since individual data was not available, the median times were plotted against one another for each trial in Figure 4.3.2 in order to illustrate the differences between the intention-to-treat population and its infected counterpart.

A summary of the randomized, included and excluded patients is provided in summary Table 4.3.3.

Table 4.3.3 Summary of Randomized and Infected Patients

Trial	# Randomized	# Infected	% Excluded
LI, 2004, China	451	273	40%
Nicholson, 2000, Canada	476	319	33%
Treanor, 2000, United States	419	253	40%
Whitley, 2001, United States	695	452	35%
Hayden, 1997 United States	276	174	37%
Hedrick, 2000, United States	469	344	27%
Makela, 2000, Europe	356	277	22%
MIST group, 1998, Australia	455	321	30%
Monto, 1999, United States	841	480	43%
Murphy, 2000, United States	525	312	41%
Puhakka, 2003, Finland	588	435	26%

Figure 4.3.2 Differences in Median Time to Alleviation of Symptoms between the Entire Randomized Sample (ITT), and the Efficacy 'Actual Infected' Sample (ITTI)

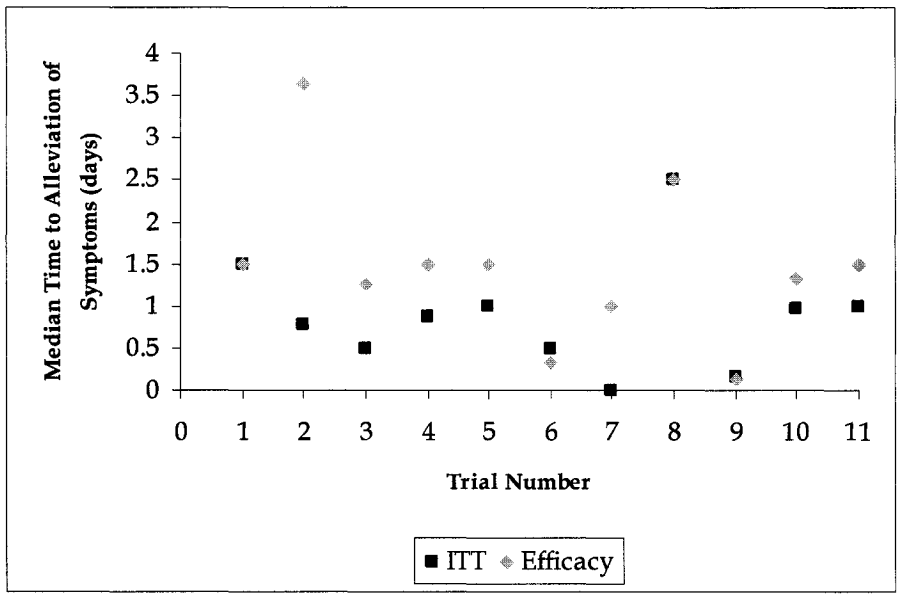


Figure 4.3.2 suggests a trend toward a 0.5 day difference between the intention-to-treat population and the efficacy population of infected patients only. One oseltamivir and two zanamivir trials show no difference in the treatment effect of the antiviral medication between the two comparison groups. The trial by Puhakka et al. (Trail Number 11 on graph) reveals that the ITT population yielded a larger treatment effect than did the infected population. When the median time to alleviation with no use of relief medication was evaluated, this treatment effect disappeared (0.83 ITTI vs. 0.67 ITT) (2). This effect is likely due in part to the very short duration of illness in both groups (2.33 d placebo, 2.0 d zanamivir). Moreover, the study population was composed of young military conscripts with a mean age of 19 years and only 5% overall reported severe symptoms at study entry (2).

Table 4.3.4 Adverse Events in Intention-to-Treat Population

Trial	Drug	Nausea	Vomiting	Pneumonia	Asthma exacerbations	Nose irritation/ nasal symptoms	Increased cough/ bronchitis
LI et al.	Oseltamivir	Reported more frequently in treatment group but not significant.	Reported more frequently in treatment group but not significant	0.5% placebo 5.2% drug	Not reported	Not reported	Not reported
Nicholson et al.	Oseltamivir	Placebo 4% Drug 2%	Placebo 4% Drug 2%	Not reported	Not reported	Not reported	Not reported
Treanor et al.	Oseltamivir	Placebo 7.4% Drug 17%	Placebo 3.4% Drug 13.1%	Placebo 0.5% Drug 0%	Not reported	Not reported	Not reported
Whitley et al.	Oseltamivir	Not reported	Not reported	Placebo 0.3% Drug 0.5%	Not reported	Not reported	Not reported
Hayden et al.	Zanamivir	Not reported	Not reported	Not reported	Not reported	5-6% in both groups	
Hedrick et al.	Zanamivir	Placebo 2% Drug <1%	Placebo 3% Drug 3%	Not reported	Placebo 1% Drug <1%	Not reported	Not reported
Makela et al.	Zanamivir	Placebo 3% Drug 2%	Placebo 3% Drug 2%	Placebo 3% Drug 1%	Not reported	Not reported	Placebo 5% Drug 2%
MIST group	Zanamivir	Placebo 4% Drug 2%	Placebo 4% Drug 2%	4% in high-risk patients	Not reported	Not reported	Cough Placebo 6% Drug 4% Bronchitis Placebo 7% Drug 3%
Monto et al.	Zanamivir	Placebo 3% Drug 3%	Placebo 3% Drug 3%	Not reported	Not reported	Placebo 3% Drug 2%	Placebo 3% Drug 1%
Murphy et al.	Zanamivir	Placebo 5% Drug 2%	Placebo 2% Drug 0%	Not reported	Placebo 7% Drug 7%	Not reported	Placebo 10% Drug 3%
Puhakka et al.	Zanamivir	Not reported	Not reported	Placebo 1% Drug 3%	Not reported	Placebo 2% Drug 2%	Placebo 2% Drug 1%

Table 4.3.4 indicates that most trials reported adverse events, but emphasizes the lack of a consistent set of reportable conditions. Nausea and vomiting were the most commonly reported adverse events for oseltamivir. Adverse events in zanamivir trials centered around increased coughing and in a few cases, the development of bronchitis. However, some trials mentioned that the adverse events were likely related to influenza-like-illness and not caused by the antiviral medication (2, 39, and 53).

4.4 Discussion

Between November and April each year, an estimated 10-15% of Canadians

may become infected with influenza and of these, 4,000 to 8,000 people, mostly seniors, die from pneumonia or other related flu complications (54). When the substantial burden influenza places on the health care system is coupled with the emergence of new anti-viral therapies as well as the imminent danger of influenza pandemic, the medical community demands rigorous trials and the highest degree of accountability and transparency in reporting. Therefore, conducting clinical trials and subsequently excluding 22- 43% of the patients from the analysis can be misleading to the readers, and fails to address the safety, economic, and practical implications of conducting trials in such a way.

Safety Implications

The influenza trials evaluated here consistently reported the adverse events for the entire randomized population. However, most trials do not consider the ethical dilemma that accompanies administering medication to uninfected patients who have the same risk of experiencing an adverse event as those infected, but no potential to benefit from the therapy. Particularly since the 2000 revision of the Declaration of Helsinki, placebo-controlled trials have recently seen a decline due to ethical issues with denying effective treatment to individuals in the control group while providing potentially effective treatment to those in the experimental group (55). As a result, most trials require head-to-head trials with conventional therapy versus a new experimental treatment unless the condition is such that it only causes discomfort and is so mild that some people will forego therapy (55). Consequently,

for many conditions, such as bronchial asthma, investigators have difficulty justifying a placebo-controlled trial despite numerous solid arguments for their valuable contribution to scientific literature (55). Yet, despite these arguments, institutional review boards still exercise extreme caution when considering the prospect of exposing even informed and willing patients to a treatment with no potential to benefit the patients' condition (56). Consider then how ethics review boards should judge trials where the patients are not only exposed to a therapy that can not benefit them, but can cause all of the adverse events that accompany the therapy. The antiviral trials did not reveal significant differences in adverse events between the patients in the treatment and control groups in the intention-to-treat population, except for the possibility of increased nausea and vomiting in the oseltamivir-treated group. Unfortunately, the trials did not elaborate on whether the infected patients experienced greater or fewer adverse events than the uninfected patients in the experimental treatment groups.

On the other hand, in some surgical trials the only way to obtain valid research data is to perform sham surgeries on patients creating similar adverse events as experimental treatment so as to maintain blinding and reduce placebo effects (56). Although interesting for infected placebo patients, this argument does not apply to the randomized ineligible influenza patients as these uninfected patients should not even be candidates for treatment.

Economic Implications

A comprehensive literature search on this topic was beyond the scope of this

chapter however some studies investigating the cost-effectiveness of anti-viral drugs given various prevalence rates of influenza were examined in detail. As mentioned earlier, illness from influenza can result in losses in worker productivity, and quality of life, not to mention the impact on direct health care costs (36). O'Brien et al. reported that using a trial-based prevalence rate of 69% influenza infection among those presenting with influenza-like-illness, the cost per influenza day averted using oseltamivir is A$48.52 (95% CI $30.72-$107.32) and the cost per quality-adjusted life year (QALY) was A $57,863 (95% CI $48,919 - $70,149) (57). This analysis was reported to be sensitive to two key model parameters; prevalence of influenza among patients presenting with ILI and the percentage of patients presenting "late" (>48hrs) who receive the drug inappropriately. Mauskopf and colleagues performed a cost analysis on zanamivir among high-risk patients. She found that the incremental cost per day of symptoms avoided was $14.20 Australian dollars and the incremental cost per QALY gained using the drug was A$11, 715 (35). Mauskopf went on to state that empirical prescribing without a diagnostic test is likely to be as cost-effective as prescribing based on a diagnostic test when influenza is known to be circulating in the community and a high proportion of the people presenting to the physician with influenza-like illness, will actually have influenza. They found that when prevalence rates dropped below 70% it was more cost-effective to perform a diagnostic test than prescribe based on clinical symptoms (35). However, Blitz et al. found Mauskopf's cost estimate for the diagnostic test to be 15% less than their own (37). They also stated that Mauskopf et al. assumed 100%

sensitivity and specificity for their diagnostic test, which is certainly not the case (58). This is an area of the rapid diagnostic test that requires further and immediate attention. Another point to consider is the ineffectiveness of the neuraminidase inhibitors against other respiratory viruses. These therapies should therefore be directed toward patients who are able to benefit from treatment (37). It follows then, that a publicly funded health care system should ensure that the decision to diagnose or prescribe treatment empirically is directly dependent on the probability of influenza, the cost and accuracy of testing, and the cost of treatment (37). According to Blitz et al, the average wholesale price for a standard 5-day course of zanamivir and oseltamivir in 2002 was US$48.02 and US$59.54, respectively (37). And yet, this cost of treatment could be offset "if future studies demonstrate that NI [neuraminidase inhibitor] use leads to reductions in antibiotic use, clinician visits, and/or hospitalizations, or reductions in influenza-related mortality" thereby increasing the perceived value of the benefit of NI therapy (37).

Practical Implications

Although rapid diagnostic tests are available for influenza, empiric diagnosis remains the clinical standard. Despite their availability, administering a test each time a patient presents with an influenza-like illness may not be cost-effective, even if it is the most accurate way of ensuring uninfected individuals do not take medication unnecessarily. Conversely, in the event an influenza pandemic sweeps across Canada, rapid diagnostic tests should become standard procedure. It is well known that despite the stockpiling of zanamivir and oseltamivir, not enough anti-

viral medication is available to protect the nation. In a public health emergency the Trade-Related Aspects of Intellectual Property Rights Agreement (TRIPS, Ar. 31) allow countries to grant compulsory licenses which permit other countries to produce a product without authorization from the patent holder (59). Hoffman-La Roche Inc, the patent hold for oseltamivir, has plainly stated that the global demand had largely surpassed their production capacity. And although the company holds the patent until 2016, they are reluctant to embrace to compulsory licensing for various reasons such as scarcity of raw materials, complex manufacturing processes and the necessity of patent protection in order to create incentives (60). Consequently, countries have been faced with the prospect of being unable to adequately protect their vulnerable populations, particularly health care workers, elderly, and the immunocompromised. The Canadian Pandemic Influenza Plan (61), followed closely by many provincial pandemic plans (62-64), has designated priority groups of people for influenza vaccination and treatment with antiviral therapy. Obviously, it is assumed that precious and expensive doses of antiviral medication will not be wasted on uninfected patients, yet this assumption is not reflected in the way clinical trials are currently conducted. Many of the antiviral trials address pandemic influenza, but none address the need for more efficient diagnostic testing, nor do they recognize the need to conduct trials that reflect clinical reality in a crisis. If rapid diagnostic testing is not a feasible, cost-effective option, than not only should influenza trials reflect this clinical reality, but this issue should also be mirrored in the pandemic plans. In addition to rapid diagnostic testing, little work

has been done to educate the public on the circumstance in which antiviral medication is effective. If providers inappropriately prescribe neuraminidase inhibitors to patients who present with symptoms more than 48 hours after onset, or patients who are afebrile, the drugs are being misused and their effectiveness is substantially affected (37).

Future Recommendations

Recommendations to mitigate the number of excluded patients in influenza trials are three-fold. First and foremost, rapid diagnostic testing should be employed whenever appropriate and cost-effective. However, if it is not realistic for this diagnostic tool to be used in the clinical setting, then the next logical step is to do an in-depth comparative analysis of the information used to empirically diagnose the infected and uninfected trial participants. In this way, physicians and trialists can determine the distinguishing factors of eligible patients, such as fever and cough (37, 38), which may help to clarify which patients are truly suffering from influenza and which are not. This second point segues into the third recommendation; stricter inclusion and exclusion criteria. The only way to increase the prevalence of infected patients in influenza trials is to improve our ability to identify these patients and enrol them.

Limitations

Several limitations exist in this systematic review. The language restriction to

trials printed in English may have implications for the generalizability of our findings, however according to the work of Moher (9) and Juni (10), the contribution of these articles may not significantly affect the review results. Second, data abstraction was completed by one reviewer, which could potentially lead to errors in abstraction. Finally, all eligible trials could not be included because of missing or unavailable information. Consequently, valuable information and insight into the issue of post-randomization exclusions was lost. The omitted trials may have illustrated a more marked difference or similarity between the two treatment populations.

Conclusions

Overall, this evaluation of influenza trials illustrated the relative strengths and weaknesses of excluding randomized ineligible patients. The efficacy results presented in the influenza trials showed a much stronger effect of antiviral therapy than did the intention-to-treat results. Although the uniform exclusion of uninfected patients from each intervention arm did not bias the trials' results, the weaknesses of this approach were manifested by a number of safety, economic, and practical implications. For future trials, investigators should look closely at the cost-effectiveness and practicality of rapid diagnostic testing, as well as devise more accurate and strict eligibility criteria for influenza trials.

Chapter 5: Final Conclusions

Patients deemed ineligible post-randomization present a formidable challenge to any investigator. This holds true regardless of whether the ineligibility stems from a change in disease-state or an incorrect diagnosis. Chapter 2 illustrated that the approaches available to deal with these patients are few and each approach possesses strengths and weaknesses that affect the external validity, and power of the trial. Most statistical approaches that serve to conserve study power and include as many patients as possible are not appropriate for post-randomization ineligibles. Many of these approaches are designed to account for non-compliance with treatment and other treatment-related changes. Unfortunately, there is a paucity of information on statistical approaches to dealing with patients with post-randomization change in disease status.

Chapter 3 illustrated the difficulty of applying methods designed for other analytic issues such as misclassification and non-compliance, to a trial with the issue of post-randomized ineligible patients. For the MCBT dataset, this resulted in estimates of treatment effect that were less informative than the ITT estimate, or were potentially biased due to patient exclusion.

Both the systematic review (Chapter 2 of this thesis) and systematic review of the influenza trials (Chapter 4 of this thesis) provide evidence that whenever possible the potential for ineligible patients to be discovered after randomization should be recognized prior to trial initiation and the course of action for dealing with such patients outlined in the trial protocol. As suggested in Chapter 2,

validation studies could provide useful information regarding expected rates of post-randomized ineligible patients.

Consideration should be given to the patients involved in the trial. If the possibility of being excluded from analysis after randomization exists, then patients should be made aware of this. It is likely that patients do not realize their information will be discarded if they fail to meet inclusion criteria after they have given their time to and been inconvenienced by a clinical trial (28).

Therefore, before an approach for dealing with these patients is selected, much thought should be given to whether this approach will best represent clinical reality, what the repercussions will be on the external validity of the trial and comparability both within trial groups as well as with other trials of the same nature, and the interpretability of the trial results. Even if a new approach appears to work theoretically, an un-interpretable result is no better than a biased result when it comes to informing treatment policy. Therefore, care must be taken to ensure the reader understands how the data was analyzed, the rationale behind the analytic approach and the limitations of the analytic approach used. The reader should be aware of the possible explanations for both significant and non-significant results as well as the clinical implications of these results.

Finally, as described in both the systematic review results and in the influenza trial analysis, trial outcomes can have significant economic impacts. The results of most trials will not negatively affect the economy of a country or even a province or state. However, the influenza trials serve as an example of the

important influence that clinical trials can have on public health policy and planning as well as the economic implications of such planning.

In conclusion, the dearth of information on the subject of post-randomization ineligible patients represents a lack of recognition of this methodological issue. Such an issue requires attention if the rigorousness of conduct of clinical trials is to be maintained and improved.

Bibliography

(1) Efron B. Foreword Limburg Compliance Symposium. Statistics in Medicine 1998;17:249-250.

(2) Puhakka T, Lehti H, Vainionpaa R, Jormanainen V, Pulkkinen M, Sharp S, et al. Zanamivir: a significant reduction in viral load during treatment in military conscripts with influenza. Scand.J.Infect.Dis. 2003;35(1):52-58.

(3) Whitley RJ, Hayden FG, Reisinger KS, Young N, Dutkowski R, Ipe D, et al. Oral oseltamivir treatment of influenza in children.[erratum appears in Pediatr Infect Dis J 2001 Apr;20(4):421]. Pediatr.Infect.Dis.J. 2001 Feb;20(2):127-133.

(4) Li L, Cai B, Wang M, Zhu Y. A double-blind, randomized, placebo-controlled multicenter study of oseltamivir phosphate for treatment of influenza infection in China. Chin.Med.J. 2003 Jan;116(1):44-48.

(5) Aaron SD, Vandemheen KL, Ferris W, Fergusson D, Tullis E, Haase D, et al. Combination antibiotic susceptibility testing to treat exacerbations of cystic fibrosis associated with multiresistant bacteria: a randomised, double-blind, controlled clinical trial.[see comment]. Lancet 2005 Aug 6-12;366(9484):463-471.

(6) Fergusson D, Aaron SD, Guyatt G, Hebert P. Post-randomisation exclusions: the intention to treat principle and excluding patients from analysis. Br.Med.J. 2002;325(7365):652-654.

(7) Hughes DA, Bagust A, Haycox A, Walley T. Accounting for noncompliance in pharmacoeconomic evaluations. [Review] [47 refs]. Pharmacoeconomics 2001;19(12):1185-1197.

(8) Hedrick JA, Barzilai A, Behre U, Henderson FW, Hammond J, Reilly L, et al. Zanamivir for treatment of symptomatic influenza A and B infection in children five to twelve years of age: a randomized controlled trial. Pediatr.Infect.Dis.J. 2000 May;19(5):410-417.

(9) Moher D, Pham B, Klassen TP, Schulz KF, Berlin JA, Jadad AR, et al. What contributions do languages other than English make on the results of meta-analyses? J.Clin.Epidemiol. 2000 Sep;53(9):964-972.

(10) Juni P, Holenstein F, Sterne J, Bartlett C, Egger M. Direction and impact of language bias in meta-analyses of controlled trials: empirical study.[see comment]. Int.J.Epidemiol. 2002 Feb;31(1):115-123.

(11) Berlin JA. Does blinding of readers affect the results of meta-analyses? University of Pennsylvania Meta-analysis Blinding Study Group.[see comment].

Lancet 1997 Jul 19;350(9072):185-186.

(12) Cuzick J, Edwards R, Segnan N. Adjusting for non-compliance and contamination in randomized clinical trials. Stat.Med. 1997;16(9):1017-1029.

(13) Nagelkerke N, Fidler V, Bernsen R, Borgdorff M. Estimating treatment effects in randomized clinical trials in the presence of non-compliance. [erratum appears in Stat Med 2001 Mar 30;20(6):982.]. Stat.Med. 2000;19(14):1849-1864.

(14) Peduzzi P, Wittes J, Detre K, Holford T. Analysis as-randomized and the problem of non-adherence: an example from the Veterans Affairs Randomized Trial of Coronary Artery Bypass Surgery. Stat.Med. 1993;12(13):1185-1195.

(15) Robins JM, Finkelstein DM. Correcting for noncompliance and dependent censoring in an AIDS Clinical Trial with inverse probability of censoring weighted (IPCW) log-rank tests. Biometrics 2000;56(3):779-788.

(16) Sommer A, Zeger SL. On estimating efficacy from clinical trials.[erratum appears in Stat Med 1994 Sep 30;13(18):1897]. Stat.Med. 1991 Jan;10(1):45-52.

(17) Brenner H. How independent are multiple 'independent' diagnostic classifications?[see comment]. Stat.Med. 1996 Jul 15;15(13):1377-1386.

(18) Greenland S. Variance estimation for epidemiologic effect estimates under misclassification. Stat.Med. 1988 Jul;7(7):745-757.

(19) Kim MY, Goldberg JD. The effects of outcome misclassification and measurement error on the design and analysis of therapeutic equivalence trials. Stat.Med. 2001;20(14):2065-2078.

(20) Lin HM, Lyles RH, Williamson JM. Bias in a placebo-controlled study due to mismeasurement of disease status and the regression effect. Control.Clin.Trials 2002 Oct;23(5):497-501.

(21) Hurst KS, Byar DP. An analysis of the effects of changes from the assigned treatment in a clinical trial of treatment for prostatic cancer. J.Chronic Dis. 1973;26(5):311-324.

(22) May GS, DeMets DL, Friedman LM, Furberg C, Passamani E. The randomized clinical trial: bias in analysis. Circulation 1981;64(4):669-673.

(23) Brittain E, Lin D. A comparison of intent-to-treat and per-protocol results in antibiotic non-inferiority trials. Statist. Med 2005;24:1-10.

(24) Lachin JM. Statistical considerations in the intent-to-treat principle.[comment][erratum appears in Control Clin Trials 2000 Oct;21(5):526].

Control.Clin.Trials 2000;21(3):167-189.

(25) Lee YJ, Ellenberg JH, Hirtz DG, Nelson KB. Analysis of clinical trials by treatment actually received: is it really an option? Stat.Med. 1991 Oct;10(10):1595-1605.

(26) Sheiner LB, Rubin DB. Intention-to-treat analysis and the goals of clinical trials. [Review] [18 refs]. Clinical Pharmacology & Therapeutics 1995 Jan;57(1):6-15.

(27) Gent M, Sackett DL. The qualification and disqualification of patients and events in long-term cardiovascular clinical trials. [Review] [15 refs]. Thrombosis & Haemostasis 1979 Feb 28;41(1):123-134.

(28) Senn S. Intention to Treat. Statistical Issues in Drug Development. 1st ed. : Wiley; 1997. p. 1-13.

(29) Aschengrau A, Seage III, George R. Overview of Epidemiologic Study Designs. Essentials of Epidemiology in Public HealthSudbury, MA: Jones and Bartlett Publishers; 2003. p. 139-140.

(30) White I. Uses and limitations of randomization-based efficacy estimators. Statistical Methods in Medical Research 2005;14:1-21.

(31) Hesterberg T, Moore DS, Monaghan S, Clipson A, Epstein R. Bootstrap Methods and Permutation Tests. Available at: http://bcs.whfreeman.com/ips5e/content/cat_080/pdf/moore14.pdf. Accessed 01/26, 2006.

(32) Efron B, Tibshirani RJ. An Introduction to the Bootstrap. New York: Chapman & Hall/CRC; 1993.

(33) World Health Organization. Cumulative Number of Confirmed Human Cases of Avian Influenza A/(H5N1) Reported to WHO. 2006; Available at: http://www.who.int/csr/disease/avian_influenza/country/cases_table_2006_03_24/en/print.html. Accessed 03/24, 2006.

(34) Cooper NJ, Sutton AJ, Abrams KR, Wailoo A, Turner D, Nicholson KG. Effectiveness of neuraminidase inhibitors in treatment and prevention of influenza A and B: systematic review and meta-analyses of randomised controlled trials.[see comment]. [Review] [15 refs]. BMJ 2003 Jun 7;326(7401):1235.

(35) Mauskopf JA, Cates SC, Griffin AD, Neighbors DM, Lamb SC, Rutherford C. Cost effectiveness of zanamivir for the treatment of influenza in a high risk population in Australia. Pharmacoeconomics 2000 Jun;17(6):611-620.

(36) Demicheli V, Jefferson T, Rivetti D, Deeks J. Prevention and early treatment of influenza in healthy adults. [Review] [37 refs]. Vaccine 2000 Jan 6;18(11-12):957-1030.

(37) Blitz SG, Cram P, Chernew ME, Monto AS, Fendrick AM. Diagnostic testing or empirical neuraminidase inhibitor therapy for patients with influenza-like illness: what a difference a day makes. Am.J.Manag.Care 2002 Mar;8(3):221-227.

(38) Zambon M, Hays J, Webster A, Newman R, Keene O. Diagnosis of influenza in the community: relationship of clinical diagnosis to confirmed virological, serologic, or molecular detection of influenza. Arch.Intern.Med. 2001 Sep 24;161(17):2116-2122.

(39) Hayden FG, Osterhaus AD, Treanor JJ, Fleming DM, Aoki FY, Nicholson KG, et al. Efficacy and safety of the neuraminidase inhibitor zanamivir in the treatment of influenzavirus infections. GG167 Influenza Study Group.[see comment]. N.Engl.J.Med. 1997 Sep 25;337(13):874-880.

(40) Boivin G, Goyette N, Hardy I, Aoki F, Wagner A, Trottier S. Rapid antiviral effect of inhaled zanamivir in the treatment of naturally occurring influenza in otherwise healthy adults. J.Infect.Dis. 2000 Apr;181(4):1471-1474.

(41) Hayden FG, Gubareva LV, Monto AS, Klein TC, Elliot MJ, Hammond JM, et al. Inhaled zanamivir for the prevention of influenza in families. Zanamivir Family Study Group.[see comment]. N.Engl.J.Med. 2000 Nov 2;343(18):1282-1289.

(42) Makela MJ, Pauksens K, Rostila T, Fleming DM, Man CY, Keene ON, et al. Clinical efficacy and safety of the orally inhaled neuraminidase inhibitor zanamivir in the treatment of influenza: a randomized, double-blind, placebo-controlled European study. J.Infect. 2000 Jan;40(1):42-48.

(43) Randomised trial of efficacy and safety of inhaled zanamivir in treatment of influenza A and B virus infections. The MIST (Management of Influenza in the Southern Hemisphere Trialists) Study Group.[see comment][erratum appears in Lancet 1999 Feb 6;353(9151):504]. Lancet 1998 Dec 12;352(9144):1877-1881.

(44) Monto AS, Robinson DP, Herlocher ML, Hinson JM,Jr, Elliott MJ, Crisp A. Zanamivir in the prevention of influenza among healthy adults: a randomized controlled trial.[see comment]. JAMA 1999 Jul 7;282(1):31-35.

(45) Treanor JJ, Hayden FG, Vrooman PS, Barbarash R, Bettis R, Riff D, et al. Efficacy and safety of the oral neuraminidase inhibitor oseltamivir in treating acute influenza: a randomized controlled trial. US Oral Neuraminidase Study Group.[see comment]. JAMA 2000 Feb 23;283(8):1016-1024.

(46) Nicholson KG, Aoki FY, Osterhaus AD, Trottier S, Carewicz O, Mercier CH, et al. Efficacy and safety of oseltamivir in treatment of acute influenza: a randomised

controlled trial. Neuraminidase Inhibitor Flu Treatment Investigator Group.[erratum appears in Lancet 2000 Nov 25;356(9244):1856]. Lancet 2000 May 27;355(9218):1845-1850.

(47) Johnston SL, Ferrero F, Garcia ML, Dutkowski R. Oral oseltamivir improves pulmonary function and reduces exacerbation frequency for influenza-infected children with asthma. Pediatr.Infect.Dis.J. 2005 Mar;24(3):225-232.

(48) Murphy K, Eivindson A, Pauksens K, Stein W, Tellier G, Watts R, et al. Efficacy and safety of inhaled zanamivir for the treatment of influenza in patients with asthma or chronic obstructive pulmonary disease: a double-blind, randomised, placebo-controlled, multicentre trial. Clin Drug Invest 2000;20(5):337-349.

(49) Boivin G, Osterhaus AD, Gaudreau A, Jackson HC, Groen J, Ward P. Role of picornaviruses in flu-like illnesses of adults enrolled in an oseltamivir treatment study who had no evidence of influenza virus infection. J.Clin.Microbiol. 2002 Feb;40(2):330-334.

(50) NAIA 3002. Available from GlaxoSmithKline at: http://ctr.gsk.co.uk/Summary/zanamivir/studylist.asp

(51) NAIB 2007. Burls,A., Clark,W., Stewart,T., Preston,C., Bryan,S., Jefferson,T., Fry-Smith,A. Zanamivir for the treatment of influenza in adults: a systematic review and economic evaluation. Health Technol.Assess. 2002 6(9): 1-87.

(52) WV15730. Available by contacting F. Hoffman-LaRoche.

(53) Monto AS, Fleming DM, Henry D, de Groot R, Makela M, Klein T, et al. Efficacy and safety of the neuraminidase inhibitor zanamivir in the treatment of influenza A and B virus infections. J.Infect.Dis. 1999 Aug;180(2):254-261.

(54) Health Canada, Public Health Agency of Canada. Influenza (The "Flu"). November, 2005; Available at: http://www.hc-sc.gc.ca/iyh-vsv/alt_formats/cmcd-dcmc/pdf/flu-grippe_e.pdf. Accessed 03/23, 2006.

(55) Onder RF. The ethics of placebo-controlled trials: the case of asthma. Journal of Allergy & Clinical Immunology 2005 Jun;115(6):1228-1234.

(56) Edward SJ, Stevens AJ, Braunholtz DA, Lilford RJ, Swift T. The ethics of placebo-controlled trials: a comparison of inert and active placebo controls. World J.Surg. 2005 May;29(5):610-614.

(57) O'Brien BJ, Goeree R, Blackhouse G, Smieja M, Loeb M. Oseltamivir for treatment of influenza in healthy adults: pooled trial evidence and cost-effectiveness model for Canada. Value in Health 2003 Mar-Apr;6(2):116-125.

(58) Cram P, Blitz SG, Monto A, Fendrick AM. Diagnostic testing for influenza: review of current status and implications of newer treatment options. Am.J.Manag.Care 1999 quiz 1562-3; Dec;5(12):1555-1561.

(59) World Trade Organization. Agreement on trade-related aspects of intellectual property rights, article 31. Available at: http://www.wto.org/english/docs_e/legal_e/27-trips.doc. Accessed 03/24, 2006.

(60) Zamiska N, Dean J. Generics Challenge Roche's Tamiflu Claims. Wall Street Journal 2005 11/03;B1.

(61) Public Health Agency of Canada. Canadian Pandemic Influenza Plan: Annex E. 2005; Available at: http://www.phac-aspc.gc.ca/cpip-pclcpi/pdf-cpip-03/cpip-appendix-e.pdf. Accessed 03/28, 2006.

(62) Government of Alberta. Alberta's Plan for Pandemic Influenza. 2006; Available at: http://www.health.gov.ab.ca/influenza/Pandemic_plan.pdf. Accessed 03/28, 2006.

(63) Ministry of Health and Long-Term Care. Ontario's Pandemic Influenza Plan. 2006; Available at: http://www.health.gov.on.ca/english/providers/program/emu/pan_flu/ohpip/plan_vaccine.pdf. Accessed 03/28, 2006.

(64) BC Centre of Disease Control. British Colombia Pandemic Influenza Preparedness Plan. 2005; Available at: http://www.bccdc.org/content.php?item=150. Accessed 03/28, 2006.

Appendix A: Systematic Review

Appendix Ai: Medline Search Strategy

1. RANDOMIZED CONTROLLED TRIAL.pt.
2. CONTROLLED CLINICAL TRIAL.pt.
3. RANDOMIZED CONTROLLED TRIALS.sh.
4. RANDOM ALLOCATION.sh.
5. DOUBLE BLIND METHOD.sh.
6. SINGLE-BLIND METHOD.sh.
7. or/1-6
8. (ANIMALS not HUMAN).sh.
9. 7 not 8
10. CLINICAL TRIAL.pt.
11. exp CLINICAL TRIALS/
12. (clin$ adj25 trial$).ti,ab.
13. ((singl$ or doubl$ or trebl$ or tripl$) adj25 (blind$ or mask$)).ti,ab.
14. PLACEBOS.sh.
15. placebo$.ti,ab.
16. random$.ti,ab.
17. RESEARCH DESIGN.sh.
18. or/10-17
19. 18 not 8
20. 19 not 9
21. COMPARATIVE STUDY.sh.
22. exp EVALUATION STUDIES/
23. FOLLOW UP STUDIES.sh.
24. PROSPECTIVE STUDIES.sh.
25. (control$ or prospectiv$ or volunteer$).ti,ab.
26. or/21-25
27. 26 not 8
28. 27 not (9 or 20)
29. 9 or 20 or 28
30. 29 and adjust$.mp. [mp=title, original title, abstract, name of substance word, subject heading word]
31. 29 and correct$.mp. [mp=title, original title, abstract, name of substance word, subject heading word]
32. 29 and complia$.mp. [mp=title, original title, abstract, name of substance word, subject heading word]
33. 29 and adhere$.mp. [mp=title, original title, abstract, name of substance word, subject heading word]
34. 29 and analy$.mp. [mp=title, original title, abstract, name of substance word, subject heading word]
35. 29 and misclassif$.mp. [mp=title, original title, abstract, name of substance word, subject heading word]
36. 35 and analy$.mp. [mp=title, original title, abstract, name of substance word, subject heading word]
37. 32 and analy$.mp. [mp=title, original title, abstract, name of substance word, subject heading word]
38. 37 and adjust$.mp. [mp=title, original title, abstract, name of substance word, subject heading word]
39. 29 and diagnos$.mp. [mp=title, original title, abstract, name of substance word, subject heading word]
40. 39 and misclassif$.mp. [mp=title, original title, abstract, name of substance word, subject heading word]
41. 29 and eligib$.mp. [mp=title, original title, abstract, name of substance word, subject heading word]
42. 41 and exclu$.mp. [mp=title, original title, abstract, name of substance word, subject heading word]
43. 41 and misclassif$.mp. [mp=title, original title, abstract, name of substance word, subject heading word]
44. 33 and analy$.mp. [mp=title, original title, abstract, name of substance word, subject heading word]
45. 44 and adjust$.mp. [mp=title, original title, abstract, name of substance word, subject heading word]
46. 45 or 43 or 40 or 38

Appendix Aii: Cochrane Methodology Register Hand-searched Journals and Conference Proceedings

Journal handsearching

Methodological articles are published in a wide range of journals. Handsearching of these to identify methodological studies is difficult and is not part of the routine handsearching being done within The Cochrane Collaboration. The Methodology Review Group handsearches specific journals that are expected to yield a high proportion of relevant methodological articles. With support from the National Health Service Research and Development Methodology Programme the following journals have been handsearched:

- *American Journal of Epidemiology* (1948-2004)
- *Annals of Epidemiology* (1991-2004)
- *Biometrics* (1948-2000)
- *Biometrika* (1948-2001)
- *BMC Health Services Research* (2001-2004)
- *BMC Journal of Negative Results in Biomedicine* (2002-2003)
- *BMC Medical Informatics and Decision Making* (2001-2003)
- *BMC Medical Research Methodology* (2001-2004)
- *Bulletin of the Medical Library Association* (1953-2001) (handsearching supported by the US Medical Libraries Association)
- *Chinese Journal of Evidence-based Medicine* (2001-2004)
- *Computer Methods & Programmes in Biomedicine* (1987-2003)
- *Controlled Clinical Trials* (1980-2003)
- *Epidemiologic Reviews* (1979-1993)
- *Epidemiology* (1990-2001)
- *Health Information and Libraries Journal* (2001-2002) (handsearching supported by the UK Library Association Health Libraries Group)
- *Health Libraries Review* (1984-2000) (handsearching supported by the UK Library Association Health Libraries Group)

- *International Journal of Epidemiology* (1972-2002)
- *International Journal of Technology Assessment in Health Care* (1985-2003)
- *Journal of Clinical Epidemiology* (1955-2003)
- *Journal of Epidemiology and Biostatistics* (1996-1999)
- *Journal of Epidemiology & Community Health* (1953-2003)
- *Journal of Information Science* (1979-2002)
- *Journal of the Medical Library Association* (2001-2002) (handsearching supported by the US Medical Libraries Association)
- *Medical Decision Making* (1996-2002)
- *Medical Informatics & the Internet in Medicine* (1976-2002)
- *Statistical Methods in Medical Research* (1992-2002)
- *Statistics in Medicine* (1982-2003)

The following special issues of journals have also been handsearched:

- *Evaluation & the Health Professions - The Cochrane Collaboration* **25** (1) 2002
- *International Journal of Epidemiology - Systematic reviews & meta-analysis* **31** (1) 2002
- *JAMA - Peer Review Congress IV* **287** (21) 2002
- *Statistics in Medicine - 3rd Symposium on Systematic Review Methodology* **21** (11) 2002

The reference lists of the following UK Health Technology Assessment Methodology Reviews have also been searched:

- Whiting P, Rutjes AWS, Dinnes J, Reitsma JB, Bossuyt PMM, Kleijnen J. Development and validation of methods for assessing the quality of diagnostic accuracy studies. *Health Technology Assessment* 2004;8(25):1-248
- Oliver S, Clarke-Jones L, Rees R, Milne R, Buchanan P, Gabbay J, Gyte G, Oakley A, Stein K. Involving consumers in research and development agenda setting for the NHS: developing an evidence-based approach. *Health Technology Assessment* 2004;8(15):1-148
- Moher D, Pham B, Lawson ML, Klassen TP. The inclusion of reports of randomized trials published in languages other than English in systematic

reviews. *Health Technology Assessment* 2003; **7** (41): 1-90

- Royle P, Waugh N. Literature searching for clinical and cost-effectiveness studies used in health technology assessment reports carried out for the National Institute for Clinical Excellence appraisal system. *Health Technology Assessment* 2003; **7** (34): 1-64

- Deeks JJ, Dinnes J, D'Amico R, Sowden AJ, Sakarovitch C, Song F, Petticrew M, Altman DG. Evaluating non-randomized intervention studies. *Health Technology Assessment* 2003; **7** (27): 1-186

- Egger M, Jüni P, Bartlett C, Holenstein F, Sterne J. How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study. *Health Technology Assessment* 2003; **7** (1): 1-76

- MacLehose, Reeves BC, Harvey IM, Sheldon TA, Russell IT, Black AMS. A systematic review of comparisons of effect sizes derived from randomized and non-randomized studies. *Health Technology Assessment* 2000; **4** (34): 1-154

- Song F, Eastwood AJ, Gilbody S, Duley L, Sutton AJ. Publication and Related Biases. *Health Technology Assessment* 2000; **4** (10): 1-115

- Prescott RJ, Counsell CE, Gillespie WJ, Grant AM, Russell IT, Kiauka S, Colthart IR, Ross S, Shepherd SM, Russell D. Factors that limit the quality, number and progress of randomized controlled trials. *Health Technology Assessment* 1999; **3** (20): 1-143

- Sutton AJ, Abrams KR, Jones DR, Sheldon TA, Song F. Systematic Reviews of Trials and Other Studies. *Health Technology Assessment* 1998; **2** (19): 1-276

Appendix Aiii: Content Expert Summary and Information

In order to ensure all pertinent information regarding post-randomization exclusions was represented in this review, ten content experts were contacted initially by formal letter and subsequently followed up by email until a response was provided. The full listing of those contacted and the letter and email templates used for correspondence are available in Appendices Aiv and Av, respectively.

Although not all experts provided additional references or authors, 8 of the 10 contacted did respond with dialogue on the topic. In particular, Drs Altman, Lavori, and Shapiro provided some insightful comments.

Dr. Altman referred me to the pivotal paper written by Schwartz and Lellouch in 1967 which discussed pragmatic and explanatory approaches to analyzing trials. This paper speaks to the importance of knowing what the trial objective is prior to analysis. For instance, if the trialist would like to know the effect of a drug on people who have a given illness, then those who are found to be suffering from something else should not be included in the analysis. However, this explanatory approach is subject to various other sources of bias as discussed in the systematic review discussion. Alternately, if the trialist would like to know the effect of a drug on those who appear to be suffering from a given illness, then all of the patients randomized should be analyzed. This pragmatic approach may lead to some issues as well, particularly with detecting an effect, as also discussed in the discussion. Dr. Shapiro felt "the approach to this problem as being driven by the underlying clinical reality" (Dr. Shapiro, personal communication, Jan 15, 2006). He believed the problem to be a design issue and not one to be dealt with statistically, if at all possible. Essentially, if it is possible to delay randomization and treatment until confirmation of eligibility, then do so,

but if this is not clinically feasible, then randomized ineligible patients who will not benefit from treatment are a clinical reality and should be reflected in the study. (Shapiro). Dr. Lavori suggested that if there was no way to obtain the needed diagnostic information pre-randomization, then the trial design should accommodate an adaptive design. Using the MCBT trial as an example, doctors could intend to give drug A, but if bacterial profile comes back unsuitable for treatment, give drug B. He stressed the point of trying to convert the profile results into a pre-randomization covariate and if this was not possible, then the altering profiles were actually part of the treatment strategy (Lavori, personal communication, January 23, 2006).

Appendix Aiv: Contacted Content Experts

Doug Altman

Professor of Statistics in Medicine
Centre for Statistics in Medicine
Wolfson College Annexe
Linton Road
Oxford OX2 6UD

Dave DeMets
Biostatistics & Medical Informatics
University of Wisconsin
600 Highland Avenue
Box 4675, K6/446 CSC
Madison, WI 53792-4675

Phil Lavori

Stuart Pocock
Medical Statistics Unit
London School of Hygiene and Tropical Medicine
Keppel Street
London WC1E 7HT

James Robins

Stephen Senn

Stan Shapiro

Louis Sheiner

Ian White
MRC Biostatistics Unit
Institute of Public Health
Robinson Way
Cambridge CB2 2SR

Scott Zeger

Appendix Av: Formal Letter and Email Templates

Formal Letter Template

Hello Dr.,

My name is Darcie Gibson and I am a Master's student in Epidemiology at the University of Ottawa. For my thesis, I am working on a systematic review of different statistical approaches to supplement intention-to-treat (ITT) analysis in randomized clinical trials when faced with patients that have deviated from trial protocol.

Specifically, for trials where treatment has to be administered immediately, and the administration is based on pre-randomization diagnostic information collected weeks or months before administration, there is a probability of randomizing patients that will never benefit from treatment due to a change in their disease-status . Arguably, this could be referred to as a form of involuntary non-compliance; although for my purposes I refer to it as a change in disease-state ascertainment. Some may argue that these patients are randomized ineligibles, since they no longer appear to have the disease state under investigation.

If you have any information on articles or abstracts pertaining to this methodological issue, it would be deeply appreciated.

Thank you in advance for your cooperation.

Sincerely,

Darcie Gibson
darcie.gibson@gmail.com

Email Template

Hello Dr. ,

I contacted you by letter last October and am just following up to see if you have any information for me regarding this specific issue of ITT. I am concerned with trials that deal with patients who effectively encounter a change in disease-status sometime between pre and post-randomization. In my circumstance, it is a change in bacterial profile which rendered the antibiotic ineffective in the patients in the treatment group that were to receive it. I am conducting a systematic review of the literature which proposes different statistical approaches to dealing with a problem such as this one, and the various effects the approaches have on power, generalizability, and comparability of treatment groups. Any information you can provide me regarding this issue would be very useful.

Thank you so much for your time.
Darcie Gibson

Appendix Avi: Data Abstraction Form for Systematic Review of Methods

Data Collection Form

Article Title:

Authors:

Year of Publication:

Journal/Vol./Issue:

Pages:

General description of approach (name, name of person who created it):

Type of outcome (binary, continuous, time-to-event):

Conditions under which approach is best applied:

Assumptions (Explicit and Implicit):

Strengths/Advantages of this approach:

Weakness/Disadvantages of this approach:

Details of analytic approach (formulae and example):

Additional points of interest or importance:

Additional References:

Appendix B: MCBT Exploration

Appendix Bi: Bootstrap Macro

```
Sub Bootstrap()
'
' Bootstrap Macro
' Macro recorded 17/02/2004 by Doug Coyle
'
' Keyboard Shortcut: Ctrl+Shift+M


   Index = 0
   Do
   Sheets("Bootstrap RESULTS").Select
   Range("b3:ao3").Select
   Range("b3").Activate
   Selection.Copy
   Range("b5").Select
   ActiveCell.Offset(Index, 0).Range("A1").Select
   Selection.PasteSpecial Paste:=xlValues, Operation:=xlNone,
SkipBlanks:= _
      False, Transpose:=False
   Index = Index + 1
   Loop While Index < 3000
```

Appendix Bii: SAS Code

```
data darcie.mcbtdruginfo;
set darcie.mcbtdruginfo;
run;

/*renaming study number variable to subnum in order to match up with
mcbtall2 set for merging*/
data darcie.mcbtdruginfo;
set darcie.mcbtdruginfo(rename=(study_number=subnum));
run;
/*renaming group assignemnt to treatcode2*/
data darcie.mcbtdruginfo;
set darcie.mcbtdruginfo (rename=(group_assignment=treatcode2));
run;
/*trying to convert text data into binary code*/

data darcie.mcbtdruginfo;
set darcie.mcbtdruginfo;
if (treatcode2='MCBT')then treatcode2=1;
if (treatcode2='MBCT')then treatcode2=1;
if (treatcode2='Conventional')then treatcode2=2;
run;

proc print data= darcie.mcbtdruginfo;
var treatcode2;
run;




/*This is the code to open the formatted mcbt data table*/
LIBNAME DARCIE 'C:\Documents and Settings\epid1\Desktop\Darcie';
OPTIONS fmtsearch=(Darcie.mcbtformats);
RUN;

/* to create a smaller dataset*/
DATA New;
SET darcie.mcbtall2 (keep=treatcode2 doexacer dob gender weight height
burkholderia pseudom stenotroph staphyl achroma otorgan otorgsp pr_prfev1
pr_psfev1 pr_prpred pr_pspredic pr_prfvc pr_psfvc pr_prfev pr_psfev
D0PRFEV1 D0PSFEV1 D0PRPRED D0PSPRE D0PRFVC D0PSFVC D0PRFEV D0PSFEV D0SAO2
CFDIAB PANCREAT CFLIVER pr_med1 pr_med3 pr_med4 pr_med2 DO2NDEXACERB);
RUN;

 Data darcie.newmcbt;
 set darcie.mcbtall2 (keep= subnum doexacer dob gender do2ndexacerb
burkholderia pseudom stenotroph staphyl achroma otorgan otorgsp);
run;


/* converting character data into binary data*/
data darcie.newdruginfo;
set darcie.mcbtdruginfo;
run;
```

```
/* sorting data by subject numbers*/
proc sort data= darcie.newdruginfo;
by subnum;
run;

/* merging files*/
data darcie.newmcbt;
merge darcie.newmcbt darcie.newdruginfo;
by subnum;
run;

/*deleting missing values*/
data darcie.newmcbt;
set darcie.newmcbt;
if treatcode2=. then delete;
run;

/*calculate time to info for data*/
data darcie.newmcbt;
set darcie.newmcbt;
timeto = DO2NDEXACERB-DOEXACER;
if timeto >0 then cens =0;
else do; timeto=mdy (2,9,2005)-DOEXACER;
cens=1;
end;
run;

/*Begin per-protocol analyses*/
/*ENDING UP WITH ONE OBSERVATION AND 4 VARIABLES*/
/* for per protocol, keeping only those who received their actual
treatment assignment*/
data darcie.perprotocol;
set darcie.newmcbt;
if treatcode3=actual;
run;

proc print data= darcie.perprotocol;
var timeto;
run;

/*proc phreg for per protocol*/
proc phreg data= darcie.perprotocol;
model timeto*cens(1)=treatcode3/RISKLIMITS;
RUN;

proc lifetest data=darcie.perprotocol plots=(s) outsurv=pp;
time timeto*cens(1);
strata treatcode3;
run;
/* calculations for ITT*/
/* for itt analysis, keep all variables*/

/*proc phreg*/
proc phreg data= darcie.newmcbt;
model timeto*cens(1)=treatcode3/RISKLIMITS;
RUN;
/*Kaplan Meier estimates*/
```

```
proc lifetest data=darcie.newmcbt plots=(s) outsurv=itt;
time timeto*cens(1);
strata treatcode3;
run;


/* for as-treated, move data around*/
proc phreg data= darcie.newmcbt;
model timeto*cens(1)=actual/RISKLIMITS;
RUN;
/*Kaplan Meier estimates*/
proc lifetest data=darcie.newmcbt plots=(s) outsurv=itt;
time timeto*cens(1);
strata actual;
run;
```

Appendix C: Influenza Trials

Appendix Ci: Medline Search Strategy

1. RANDOMIZED CONTROLLED TRIAL.pt.
2. CONTROLLED CLINICAL TRIAL.pt.
3. RANDOMIZED CONTROLLED TRIALS.sh.
4. RANDOM ALLOCATION.sh.
5. DOUBLE BLIND METHOD.sh.
6. SINGLE-BLIND METHOD.sh.
7. or/1-6
8. (ANIMALS not HUMAN).sh.
9. 7 not 8
10. CLINICAL TRIAL.pt.
11. exp CLINICAL TRIALS/
12. (clin$ adj25 trial$).ti,ab.
13. ((singl$ or doubl$ or trebl$ or tripl$) adj25 (blind$ or mask$)).ti,ab.
14. PLACEBOS.sh.
15. placebo$.ti,ab.
16. random$.ti,ab.
17. RESEARCH DESIGN.sh.
18. or/10-17
19. 18 not 8
20. 19 not 9
21. COMPARATIVE STUDY.sh.
22. exp EVALUATION STUDIES/
23. FOLLOW UP STUDIES.sh.
24. PROSPECTIVE STUDIES.sh.
25. (control$ or prospectiv$ or volunteer$).ti,ab.
26. or/21-25
27. 26 not 8
28. 27 not (9 or 20)
29. 9 or 20 or 28
30. 29 and influenza.mp. [mp=title, original title, abstract, name of substance word, subject heading word]
31. 30 and oseltamivir.mp. [mp=title, original title, abstract, name of substance word, subject heading word]
32. 30 and zanamivir.mp. [mp=title, original title, abstract, name of substance word, subject heading word]
33. 31 or 32
34. from 33 keep 1-202

Appendix Cii: Influenza Embase Search Strategy

1. Clinical trial/
2. Randomized controlled trial/
3. RANDOMIZATION/
4. single blind procedure/
5. double blind procedure/
6. crossover procedure/
7. placebo/
8. randomi?ed controlled trial$.tw.
9. Rct.tw.
10. random allocation.tw.
11. randomly allocated.tw.
12. allocated randomly.tw.
13. (allocated adj2 random).tw.
14. single blind$.tw.
15. double blind$.tw.
16. ((treble or triple) adj blind$).tw.
17. placebo$.tw.
18. prospective study/
19. or/1-18
20. case study/
21. case report.tw.
22. abstract report/ or letter/
23. or/20-22
24. 19 not 23
25. 24 and influenza.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer name]
26. 25 and oseltamivir.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer name]
27. 25 and zanamivir.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer name]
28. 26 and 27

Appendix Ciii: Data Collection Form for Influenza Trials

Influenza Data Abstraction

RW ID:
Title of Article:

Authors:

Journal of Publication:

Year:

Vol/Issue/Pages:

Country:

Source of Funding:

Study design:

Study Population (community or hospital based):

Patient characteristics (*was vaccination permitted):

Study Characteristics (drug, dosing instructions)

Primary Outcome:

Blinding:

Method of Randomization:

Number Randomized in each arm:

Number positively infected in each arm:

Reported Efficacy Effect:

Reported ITT effect:

Definition of ILI:

Study Duration:

References: