2011

REPORTING QUALITY OF ONCOLOGY RCTS & COMPLIANCE WITH CONSORT

Ian H. Arra

REPORTING QUALITY OF ONCOLOGY RCTS & COMPLIANCE WITH CONSORT

(Spine title: Reporting Quality of RCTs)

(Thesis format: Integrated-Article)

By

Ian H. ARRA

Graduate Program in Epidemiology and Biostatistics

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science

The School of Graduate and Postdoctoral Studies
The University of Western Ontario
London, Ontario, Canada

THE UNIVERSITY OF WESTERN ONTARIO
SCHOOL OF GRADUATE AND POSTDOCTORAL STUDIES

CERTIFICATE OF EXAMINATION

Supervisor

Dr. George Rodrigues

Supervisory Committee

Dr. Brian Rotenberg

Dr. Tracy Sexton

Examining Board

Dr. Richard Malthaner

Dr. Jeffery Mahon

Dr. Michael Lock

The thesis by

Ian ARRA

entitled:

REPORTING QUALITY OF ONCOLOGY RCTS & COMPLIANCE WITH CONSORT

is accepted in partial fulfillment of the
requirements for the degree of
Master of Science

Date

Chair of the Thesis Examination Board

ABSTRACT

Introduction: The Consolidated Standards for Reporting Trials (CONSORT) checklist has been formulated to improve the reporting of randomized controlled trials (RCTs). The purpose of this investigation is to determine predictors of CONSORT checklist compliance in the oncology literature over the past two decades.

Methods: Eight-hundred and fifty parallel two-arm RCTs assessing oncological interventions in adult breast, prostate, colorectal, and lung cancer between 1992-2010 were identified by a systematic search of the medical literature. Exclusion criteria included investigations reporting interim/secondary/long-term update analyses, pilot or phase 2 studies, and studies not employing a parallel design. After full article review, 408 articles were eligible for inclusion into the CONSORT audit database. RCT descriptive variables including number of authors/study patients, 2009 journal impact factor/journal classification, type of cancer and therapeutic intervention, publication year, primary study country, and cooperative group involvement were captured for all trials. CONSORT guideline compliance was assessed by two qualified auditors in order to generate average and difference CONSORT checklist scores.

Results: Mean average CONSORT score was 16.6 (SD 3, max 25) and median difference score was 2 (interquartile range 1-3). Kappa agreement for each individual CONSORT checklist item ranged from (0.02-0.92) with an overall two-way intraclass correlation coefficient of 0.71 (95%CI 0.61-0.78) for comparison of overall CONSORT score between raters. Recent year of publication, increasing author number, and higher impact factor were associated with higher average CONSORT scores ($p<0.0001$). Recent year of publication was the only factor associated with a decrease in the CONSORT difference score.

Conclusions: In this large reported CONSORT compliance audit in the medical literature, improvements in RCT reporting have been observed over time in the cancer literature. Further work in the assessment of the inter-observer reliability of individual CONSORT items is warranted given the observed kappa agreement heterogeneity.

Keywords: Randomized Controlled Trial, Reporting, Quality, Breast, Colorectal, Prostate, Lung, Cancer.

DEDICATION

This thesis is dedicated to my Mom and Dad

ACKNOWLEDGEMENTS

I would like to thank my thesis advisors, Dr. George Rodrigues, Dr. Brian Rotenberg, and Dr. Sexton for their support and assistance during the development, execution, data analysis and writing of the thesis.

Dr. Vikram Velker for database construction and data extraction.

Mr. Gabe Boldt for his support in the literature review, database search, and data extraction.

My heartfelt thanks to Mary Lu Lacasse for the outstanding secretarial support.

My thanks to the good friend, Dr. Muhammad Ali, for his encouragement and his general advice on the thesis.

Finally, I would like to thank Andrea and Emily Campbell for their love, support, and encouragement.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

CABG	Coronary Artery Bypass Grafting
CCS	American Canadian Cancer Society
CINAHL	Cumulative Index to Nursing and Allied Health Literature
CONSORT	Consolidated Standards of Reporting Trials
EMBASE	Excerpt Medical Database
ICMJE	International Committee of Medical Journal Editors
IF	Impact Factor
HealthSTAR	Health Services Technology, Administration, and Research
HL	Hodgkin's Lymphoma
MedicReS	Medical Research Support Foundation
NHL	Non-Hodgkin's Lymphoma
QA	Quality Assurance
RCT	Randomized Controlled Trials
WAME	World Association of Medical Editors

1.0 Introduction

The Canadian Cancer Society publishes general cancer statistics on an annual basis. In the current 2011 release, the main focus of this publication was on the four most common cancers which include breast, colorectal, lung, and prostate cancer).[1] In 2011, the Canadian Cancer Society estimated that 177,800 new cases of cancer and 75,000 deaths will occur in Canada. Approximately 84,800 women and 93,000 men will be diagnosed with cancer, and of these numbers approximately 35,100 women and 39,900 men will die of cancer in Canada. On average, 487 Canadians will be diagnosed with cancer, and 205 Canadians will die of cancer every day. The Society also stated that lung, prostate, breast and colorectal cancer account for 50% of all cancer deaths, and 50% of new cancer cases, with lung cancer accounting for over a quarter (27%) of all cancer deaths; breast cancer accounting for over a quarter (28%) of new cancer cases in women; and prostate cancer accounting for over a quarter (27%) of new cancer cases in men (Figures 1 and 2).

Research into the cause, prevention, and treatment of cancer is a highly complex enterprise as documented by the following:

1. Cancer research includes numerous areas and generally involves extensive collaboration among several disciplines such as Molecular and Physical Cancer Research, Cancer Imaging, Basic Cancer Biology, Basic Cancer Physics, and Clinical Cancer Research.

2. The conduct of cancer research is expensive. Even a basic science laboratory that is operated by a single scientist might require funding between CAN $50,000 to $500,000 per year to cover equipment and salaries. Cancer cells required for just one experiment cost between

CAN$300 to $500.[1] Clinical research can be the most expensive. For example, a review that included 28 Randomized Controlled Trials (RCTs) reported total expenses of approximately US$ 335 million with a mean of about US$ 12 million per RCT.[2]

3. Besides financial costs, cancer research heavily consumes other resources such as time and human effort. It can take several years and the involvement of many investigators and participants in several centers to complete a single study. For example, over the past 40 years a single clinical trial group, the NCIC Clinical Trials Group, has enrolled about 45000 patients on various cancer clinical trials.[3]

The Randomized Controlled Trial (RCT) is an indispensible tool in to investigate and assess various cancer treatments. Metaphorically, RCTs are the bridge to transfer basic science knowledge to clinical practice. They are frequently used to investigate a broad range of therapies as listed by the Canadian Cancer Society:[3]

"1. new anti-cancer drugs, including chemotherapy, hormonal therapy, biological therapy and immunotherapy agents
2. new approaches to cancer prevention, screening, surgery and radiation therapy
3. new combinations of treatments
4. new ways of using standard treatments
5. complementary and alternative cancer therapies
6. supportive care to reduce the impact of cancer on emotions and behaviour"

RCTs are used to validate new drugs and surgical procedures that are potentially more effective than existing standard of care treatments for specific types of cancer. Many of the cancer treatments that are used today were tested and developed in clinical trials. Even the most promising scientific findings must first be proven to be safe and effective in clinical trials before they can be routinely utilized in clinical practice and be made available to the public.

The reporting of successfully completed RCTs is the final stage of clinical research and is extremely important in the translation of acquired knowledge to potential users of such information. A RCT literature report summarizes the trial rationale, objectives, methodology, results, discussion and conclusion. Therefore, hypotheses generated from preclinical and early clinical studies are tested in more definitive RCTs, and then the knowledge gained from the findings are transferred to clinical practice ideally through, in part, clear reporting in the medical literature. Furthermore, studies have shown that RCTs with poor reporting quality are associated with biased findings.[4, 5] Awareness concerning the quality of RCTs reporting is growing. Inadequately conducted trials are viewed as a waste of time, effort, and resources. Similarly, well-conducted trials with inadequate reporting quality can represent a waste of these same resources.

Various guidelines have been previously created to alleviate problems arising from inadequate RCTs reporting.[6] These various guidelines are currently encapsulated by the "CONSORT - Consolidated Standards of Reporting Trials" clinical trial reporting guideline that have been developed by the CONSORT Group. The main product of CONSORT is the CONSORT Statement, first published in 1996 with subsequent revisions in 2001 and 2010. The statement is "an evidence-based, minimum set of recommendations for reporting RCTs. It offers a standard way for authors to prepare reports of trial findings, facilitating

their complete and transparent reporting, and aiding their critical appraisal and interpretation."[6] A considerable number of journals worldwide and many editorial bodies have approved and recommended the CONSORT Statement including the International Committee of Medical Journal Editors, the World Association of Medical Editors, and the Council of Science Editors.

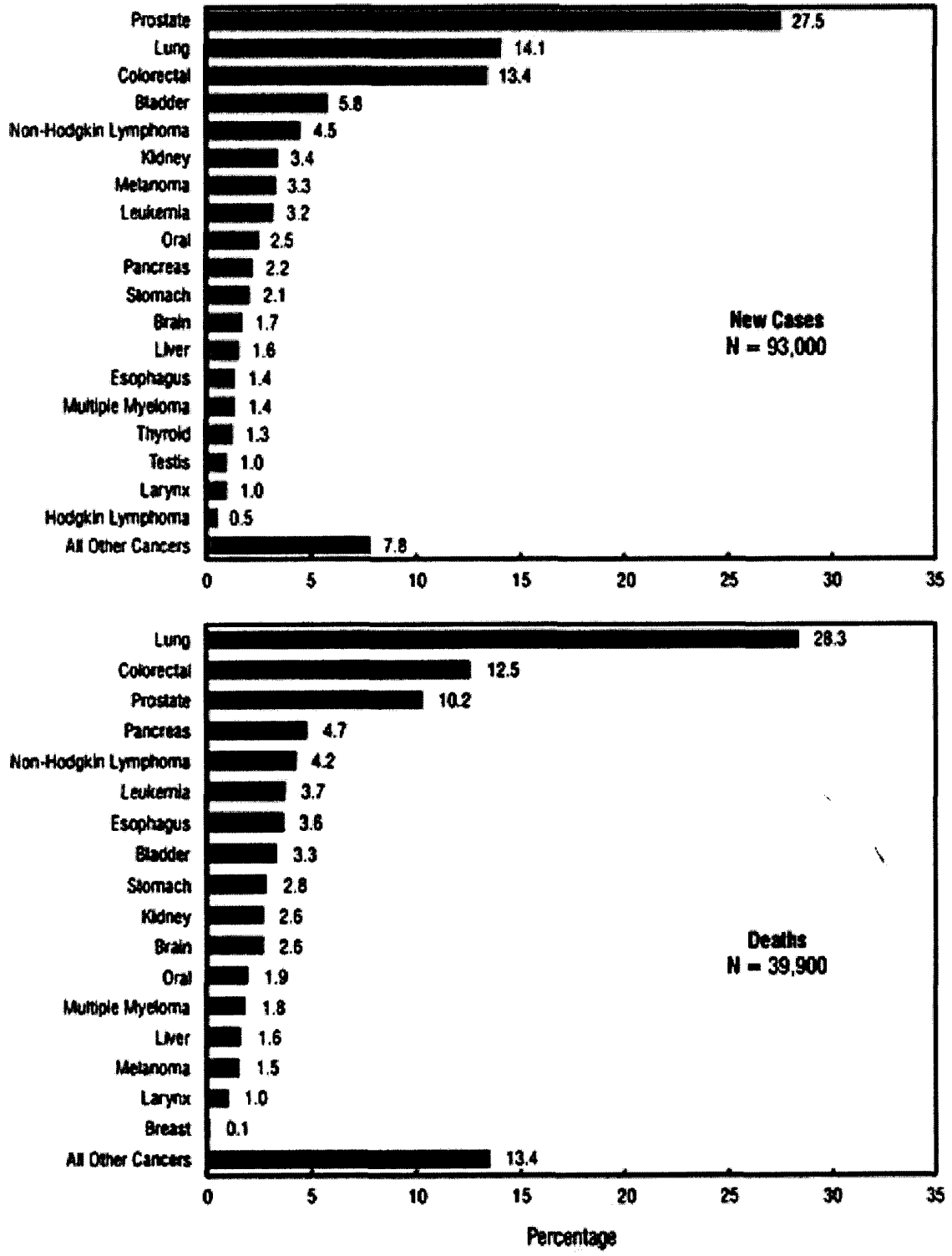
Although there have been numerous recommendations over the past 30 years to enhance RCTs reporting, and more intensely over the past 15 years for many medical literature publishers to attempt to adhere to the CONSORT Statement; reviews in several fields of medicine have repeatedly shown that the reporting quality is still problematic with poor adherence to the statement.[7,8] To our knowledge, an assessment of the quality of RCTs reporting and adherence to the CONSORT statement in cancer research has not been previously attempted or documented. Given the reliance on RCTs on progress in the fight against various cancers over the last two decades as well as the proliferation of RCTs in cancer research (and medicine in general), an audit of RCT reporting quality was felt to be warranted by the study investigators.

The primary objective of this investigation is to determine predictors of CONSORT checklist compliance in the oncology literature over the past two decades. Some of the potential predictors to be assessed are Journal Name, Type of Cancer, Publication Year, Number of Authors, Number of Patients, Intervention, Trial Site, One VS Multiple Countries, Primary Country, Cooperative Group, Impact Factor, Oncology VS Non-oncology Journal.

In addition, we plan for this investigation to shed light on the reliability of the CONSORT checklist items in the cancer RCT context. It is hypothesized that some of these items have clear working definitions, while others do not. This investigation will examine whether the quality of reports of randomized trials has improved over the past 20 years. We hypothesized that the quality of RCT

reporting has improved over the past 20 years given the increased utilization and knowledge regarding RCT reporting quality statements such as CONSORT.

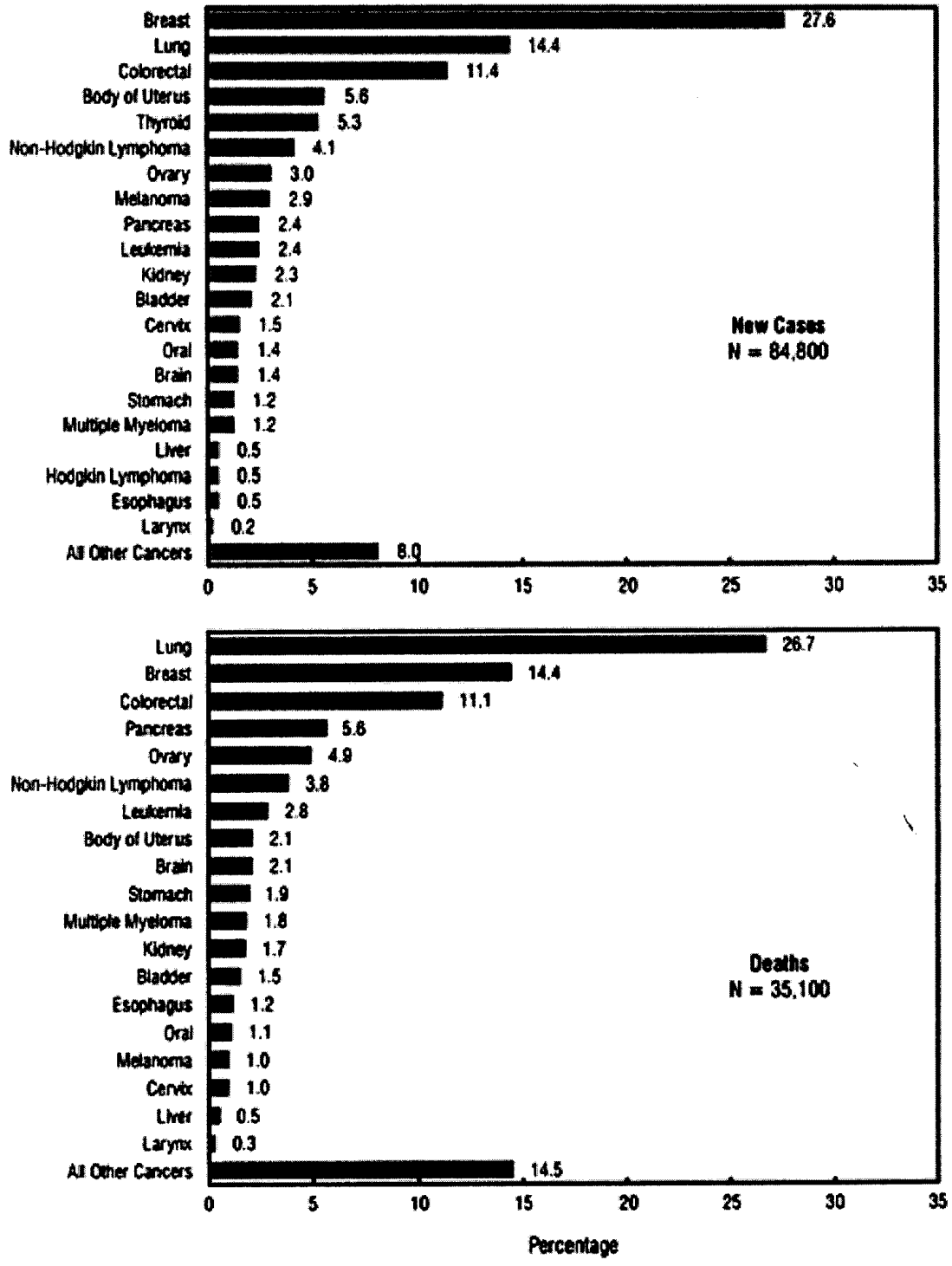
Figure 1: Percentage Distribution of Estimated New Cases and Deaths for Selected Cancers, Males, Canada, 2011



Note: New cases exclude an estimated 40,700 new cases of non-melanoma skin cancer (basal and squamous). The number of deaths for "All Other Cancers" includes about 170 deaths with underlying cause "other malignant neoplasms" of skin.

Analysis by: Chronic Disease Surveillance and Monitoring Division, CCDPC, Public Health Agency of Canada
Data sources: Canadian Cancer Registry and Canadian Vital Statistics Death databases at Statistics Canada

Figure 2: Percentage Distribution of Estimated New Cases and Deaths for Selected Cancers, Females, Canada, 2011



Note: New cases exclude an estimated 33,300 cases of non-melanoma skin cancer (basal and squamous). Deaths for "All Other Cancers" include about 100 deaths with underlying cause "other malignant neoplasms" of skin.

Analysis by: Chronic Disease Surveillance and Monitoring Division, CCDPC, Public Health Agency of Canada
Data sources: Canadian Cancer Registry and Canadian Vital Statistics Death databases at Statistics Canada

2. Randomized Controlled Trials (RCTs)

2.1 Historical Perspective

Around the year 600 B.C. Daniel of Judah conducted what might be the first controlled clinical trial. This trial was a comparison of the health effects of a vegetarian diet (vegetable and water) with those of a royal diet (meat and wine) over a period of ten days. At the end of the trial, the first group looked healthier and seemed better nourished than the second group. By current research standards this trial might have many drawbacks such as selection bias, confounding, measurement bias, and others, yet it still stands as one of the earliest clinical trials assessing different interventions for an important endpoint (in the opinion of the observer).[9]

The 19th century witnessed chief developments in clinical trials. For example, in the year 1836 the French trial studying the effect of blood-letting in management of pneumonia by P. C. A. Louis[10] had a very important influence on medical science. This importance was captured clearly in a statement published in the American Journal of Medical Sciences "one of the most important medical works of the present century, marking the start of a new era of science ... the first formal exposition of the results of the only true method of investigation in regard to the therapeutic value of remedial agents."

The 20th century witnessed the development of RCTs in a manner consistent with modern RCT practices. A trial entitled "Streptomycin treatment of pulmonary tuberculosis" published in 1948 was considered the first modern RCT.[11] Sir Austin Bradford Hill, one of this study's authors, was given credit for this achievement.[9] However, in 2007, Forsetlund investigated when random

allocation was first used. His investigation revealed that RCTs had been used in social and educational studies as early as 1928.[12] In the second half of the 20th century, RCTs were rightly regarded as a landmark principle that guided a new era of rational evidence-based medicine. The methodology of the RCT has been increasingly acknowledged, and its use has become almost universal to provide a high level when it comes to quality of evidence. RCTs are considered the "gold standard" for evidence-based medicine.[13] As of 2004, more than 150,000 RCTs have been cataloged in the Cochrane Library.[9]

2.2 Classifications of RCTs

The following are four common methods for classifying RCTs. Classification by phase (Phase 1, 2, 3, and 4 trials), by type of design, by type of hypothesis (superiority vs. noninferiority), and by clinical trial aim (explanatory v. pragmatic) are routinely utilized in the literature and in practice.

The first method of classifying RCTs is by phase (Phase 1, 2, 3, and 4 trials).[9] These four phases are used to describe the different potentially sequential steps in the process of investigating a new intervention, usually a new drug. In general, this system of clinical trials is used to increase the efficiency of data collection by testing interventions with increasing numbers of patients. Additionally, the clinical trials move from safety endpoints in the early stages to clinically relevant endpoints in later phases of clinical trial therapeutic intervention development. It is important to note that not all phases are necessarily used for all new therapeutic interventions.

As a first step after preclinical testing, a phase 1 trial is used to investigate an intervention once the intervention is deemed safe in pre-clinical (animal) research. Its purpose is to assess the safety of the intervention in humans, and

to establish the therapeutic dose and maximum tolerated dose. Participants are usually a small number of healthy volunteers, or patients who have failed all conventional treatment and have no other standard options. Then, a phase 2 trial is conducted once the intervention passes phase 1. Typically, the intervention is given to a small group of patients that may benefit from treatment. Its purpose is to establish the efficacy of different modes and doses of the interventions, although data on safety is still collected and examined in order to establish the therapeutic ratio of the intervention. Phase 1 and 2 trials are often combined into one trial for efficiency of data collection. Once the safety and efficacy of the intervention has been documented in a phase 2 trial, a definitive phase 3 trial is conducted. Its purpose is to establish the effectiveness of the proposed intervention(s) against a control group (placebo or current standard of care). Participants are real patients generally numbering in the hundreds or thousands based on an a-priori sample size calculation based on effect size, power, variance, and alpha (usually 0.05). As a final optional step, phase 4 trial is usually conducted after the marketing of the intervention. Its purpose is to discover possible rare or late-occurring side effects not observed in phase 3 RCT due to the limited follow-up and number of participants.

The second method of classifying RCTs is by the type of design. In a comparative study of 616 RCTs indexed in PubMed in 2006, Hopewell *et al* found that over three quarters (78%) of these reports were of parallel group trials, 16% were crossover trials, and the remaining 6% were cluster, factorial, or split-body trials.[14]

In parallel group trials, individual participants are randomly allocated to one of two (two-arm trial) or more intervention groups. All subjects in one group are given the same a-priori defined intervention or placebo intervention if no standard of care exists for the situation being studied. The group that receives the investigated intervention is called the experimental group. The other group

receives a standard of care intervention or no intervention and is routinely called the control group.[15] This is the most common trial design (over three quarters of RCTs)[14] as it is a straightforward design and its analysis is usually more straightforward than the other designs.

In a cross-over trial, the participants in one group receive an initial intervention then will be exposure to the other groups at a pre-determined point. Thus, each participant in the trial will receive both interventions one after the other but the order of exposure will be randomly assigned.[15] A crossover design has two potential advantages over non-crossover designs. First, the effect of potential confounding is decreased because each participant serves as his or her own control. Second, fewer participants are generally required in this type of study.

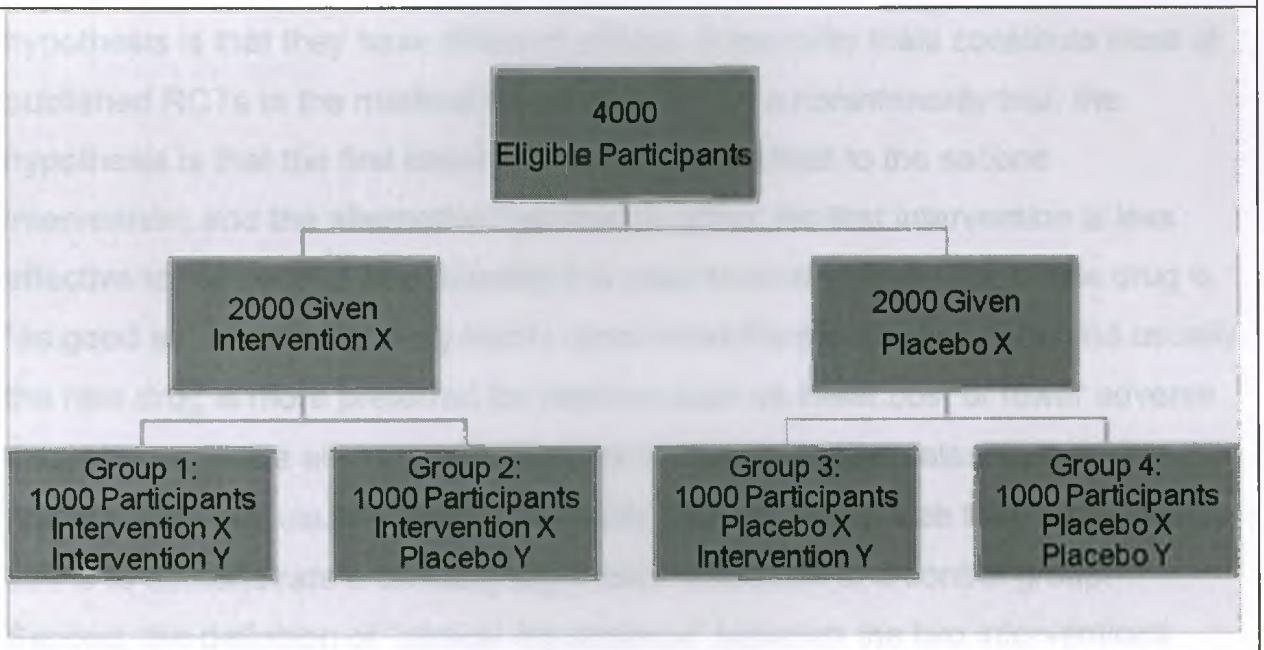
In cluster trials predefined groups or populations of people rather than independent participants, are randomly allocated to intervention groups. For example: schools, towns, and cities can randomly be assigned to be given or not given a specific intervention. Cluster randomized controlled trials have two main advantages over individually-randomized controlled trials. First, this design allows for the ability to control for the contamination effect among participants (e.g., change in one participant's behaviours may influence another participant's behaviours). Second, the ability to study interventions that cannot be directed toward selected individuals (e.g. the use of a radio show directed to change lifestyle targeted towards a population).[16]

In split-body trials, detached parts of an individual participant rather than independent participants are randomly allocated to intervention groups. For example: separate lesions on the skin, or the left and right eyes are randomly given or not give an intervention.[15] This design reduces in-between subjects variation and bias. One of the disadvantages of this clinical trial design is the

potential lack of independence among body parts in the same individual, which creates special methodological challenges in the design, conduct, and analysis of such trials. This disadvantage is a fundamental issue frequently discussed in the published literature of Otology and Ophthalmology.[17]

In factorial trials, participants are allocated to groups of separate interventions, combined interventions, and no intervention. For instance, a 2x2 factorial trial randomly allocates participants to four groups using two steps as shown in Figure (3). In the first step they are randomly assigned to one of two groups. One group is given intervention X and the other group is given placebo X (Sometimes it is a different intervention not necessary placebo). In the second step participants in each group are again randomly assigned to one of two groups. One group is given intervention Y and the other group is given placebo Y. The effect of intervention X can be obtained from comparing groups 1 & 3 vs. 2 & 4, while the effect of intervention Y can be obtained from comparing groups 2 & 3 vs. 1 & 4. The effect of intervention X and Y can be compared against each other, combination of X and Y, and placebo.[15] A factorial design provides efficiency in answering multiple therapeutic questions as well as the assessment of therapeutic interactions (additive and synergistic effects) compared to a non-factorial one (e.g., it enables studying two interventions in one sample instead of having to conduct two separate trials).

Figure 3: Trial with 2 x 2 Factorial Design



The third method of classifying RCTs is by the type of hypothesis being assessed in the trial (e.g. superiority vs. noninferiority). In a superiority trial, the null hypothesis is that the interventions have equal effects and the alternative hypothesis is that they have different effects. Superiority trials constitute most of published RCTs in the medical literature.[18,19,20] In a noninferiority trial, the hypothesis is that the first intervention has equal effect to the second intervention, and the alternative hypothesis is that the first intervention is less effective to the second one. Usually it is used to demonstrate that a new drug is "as good as" an existing drug that is considered the standard of care, and usually the new drug is more preferred for reasons such as lower cost or fewer adverse side-effects. There are two requirements for noninferiority trials. First, a noninferiority trial usually requires a much greater sample size than a trial whose aim is to demonstrate a clinically significant difference to a control group. Second, the definition of "clinical equivalence" between the two interventions must be specified in advance as readers can and should be suspicious of studies that apply 1-sided analyses without previous planning and reporting.[19,20]

The fourth method of classifying RCTs is by the fundamental trial aim and approach. RCTs can be classified as explanatory (efficacy) or pragmatic (effectiveness) clinical trials.[9,21] Explanatory (efficacy) trials are conducted in ideal setting, and designed to answer the question "Can the intervention work?". In these studies, the inclusion/exclusion criteria are very strict; poorly compliant participants and those with conditions which might dilute the effect are often excluded, the intervention is strictly enforced and adherence is monitored closely, and the outcomes are often short-term surrogates or measures of effect.[21] Pragmatic (effectiveness) trials are conducted in real life setting, and designed to answer the question "Does the intervention work when used in normal practice?". In these studies, inclusion criteria are generally less strict and resemble the target patient population of interest. The intervention may be applied flexibly as it

would be in real world practice. Reported outcomes are directly relevant to participants, funders, communities and healthcare providers.[21]

2.3 Randomization

Randomization is defined by Donner as a process where "participants allocated to one of the groups (study, control) by a random mechanism which assures that each of the participants has an equal chance of being assigned to any group."[22] Randomization consists of two temporal steps. The first step is the randomization procedure, in which investigators (or their delegates) generate a random sequence to assign participants to trial groups. The second step is allocation concealment, which is a set of strict preventative measures taken by the investigators (or delegates) to make certain that the assignment sequence is kept unknown and unpredictable until each participant has been officially allocated to an intervention.

Proper randomization in RCTs facilitates blinding of the identity of interventions from participants, investigators, and assessors. It allows the use of probability theory (the likelihood that any difference in outcome between treatment groups merely indicates chance.) and it enables comparability between trial groups on factors (whether known or unknown) that may influence outcomes.[23,24]

2.3.1 Randomization Procedures:

Randomization procedures can be classified as simple (unrestricted), restricted, or adaptive.

Unrestricted Randomization

This is also referred to as simple randomization or complete randomization, and is the most common type of randomization seen in the literature. Examples of this type of randomization are repeated coin-tossing and computer based randomization. The main disadvantage of unrestricted randomization is the possibility of generating unequal trial groups (by chance) if the trial sample size is small. Therefore it is usually recommended in trials with a sample size above 200 participants.[25]

Restricted randomization

Restricted randomization is also known as block randomization. The main benefit of using this type of randomization is to generate equal trial groups usually in the conduct of smaller trials. [25] The number of participants in one group versus the number of participants in the other group(s) is pre-specified according to a ratio that is called "allocation ratio". To maintain the allocation ratio during recruiting, participants are randomized within blocks according to the allocation ratio.[24] For example, if the allocation ratio were specified as 1:3, a block size of 8 would force a random assignment of 2 participants to the first group and 6 participants to the second group. Block size can be fixed or variable so long as participants in each block are randomized according to the same allocation ratio. Varying the block size can reduce the predictability of the sequence.[26]

Adaptive Randomizations

This type of randomization is less commonly used compared to the previous two types. This type can be classified into two main sub-types:

1. Minimization, also known as covariate-adaptive randomization: the probability of being assigned to a group decreases or increases to minimize potential imbalance between predictive factors (known factors).[25] The first participant's assignment is performed randomly, but the rest of the assignment is adaptive based on previous assignments. Although this procedure is quite robust in minimizing bias on known factors, the lack of true randomization renders it less than optimal in minimizing bias on unknown factors. [27] Therefore, its adequacy as a robust randomization procedure is considered controversial by many investigators.[24]

2. Outcome-adaptive randomization: The chance of assigning a participant to a group is directly related to the percentage of previous patients with favorable outcome in that group. For example, if 80% of the subjects in group A and only 50% of the subjects in group B have favourable responses, the next recruited subject will be assigned to group A as it has higher chances of favourable response at this point of the trial. This type of randomization is usually used when investigating a serious disease like AIDS where favourable patient outcomes are crucial.[25]

2.3.2 Allocation Concealment

Allocation concealment is defined as "the procedure for protecting the randomization process so that the treatment to be allocated is not known before the patient is entered into the study".[28] The integrity of effective randomization rests on appropriate allocation concealment. Although allocation concealment is sought-after in RCTs,[29] it is not always logistically followed in real-life practice. Clinical investigators are not always neutral when it comes to their own patients care. There have been incidences where clinical researchers have held up sealed envelopes to lights to find out the allocation sequence in order to control their next patient's assignment.[24] Such action voids the key benefits of

randomization, namely minimizing selection bias and confounding. [24] Fortunately, this breach could be largely prevented by various measures to prevent tampering and in the varying of the block size if stratified randomization is utilized.

Intervention-group assignment can be performed at a central study office by trained staff whose main responsibility is to preserve randomization validity. It also can be performed by other means such as sequentially numbered sealed opaque envelopes or sequentially numbered pre-randomized medication containers.[24] Due to the central role allocation concealment systems play in the validity of an RCT, it is strongly recommended to report allocation concealment systems in detail in the RCT protocol, as well as in the final published RCT report. Unfortunately, most published RCTs have vague allocation concealment in their protocols, in their reports, or in both.[30]

2.4 Blinding

Blinding (sometimes known as masking) in RCTs is a set of procedures that ensures that individuals involved in the study do not know which study participants are in which group (e.g. intervention vs. control).[31] It is a methodologic approach that can be employed to decrease potential observation bias and the ascertainment of outcomes bias. Furthermore, it preserves the integrity of the randomization by preventing switching of participants from one group to another by trial personnel.

RCTs can be classified into four types according to the level of blinding:[32]

1. Open (Unblinded) RCT: All persons involved in the study are aware which participant is receiving which intervention.[33,34] If the intervention is a drug treatment, the RCT is referred to as an open-label clinical trial.[34]

2. Single-blinded RCT: Either all participants or all investigators involved in the study are unaware which participant is receiving which intervention.[36,37]
3. Double-blinded RCT: Both participants and investigators involved in the study (usually patients and clinicians) are unaware which participant is given which intervention.[36,37]
4. Triple-blinded and quadruple-blinded RCTs: A triple-blinded RCT consists of blinding of participants, investigators, and study evaluators and a quadruple-blinded RCT consists of blinding of participants, investigators, study evaluators, and statisticians.

In well-designed trials, other people such as patient caregivers and proxies who do not play a role in treatment, measurement, or analysis might be considered for blinding as well. The rationale for this kind of blinding stems from the fact that caregivers' actions potentially could influence treatment results, and be influenced by knowing which treatment is given to the patient.[38]

In 2001 and 2006, two papers demonstrated that the terms "single-blind," "double-blind," or "triple-blind" have different meanings for different people. [36,37] The 2010 Consolidated Standards of Reporting Trials (CONSORT) Statement (which will be described in detail in chapter 3) recommended that authors should be more transparent when reporting on blinding by specifying "who was blinded after assignment to interventions (for example, participants, care providers, those assessing outcomes) and how, If relevant, description of the similarity of interventions".[27]

Unlike allocation concealment, blinding is not always feasible, practical or even possible. For example, most surgical studies are open-label because blinding is usually impossible, impractical, and/or unethical with surgical interventions. Another example, if the intervention requires the participant to

perform an active role such as physical therapy. In this case, by definition, blinding is impossible.[31]

2.5 Analysis of Data from RCTs

The characteristics of data and design dictate the type of analyses used in RCTs. Analysis in RCTs could be simply classified as main analysis, subgroup analysis, as well as missing data analysis.

1. Main analysis: Statistical methods vary widely depending upon the type of outcome data. To analyze binary (dichotomous) outcome data, logistic regression and other methods can be used. To test the effects of predictor variables for continuous outcome data, analysis of covariance can be used. To examine time-to-event outcome data that may be censored (for example, time to cancer death after receipt of chemotherapy in women with breast cancer) survival analysis (e.g., Kaplan-Meier estimators and Cox Proportional Hazards Models) can be used. Vittinghoff *et al* presented, in table (1), different types of data, an example for each type, and a proper method of analysis for it.

Table 1: Type of outcome variable determines choice of multivariable regression model (Vittinghoff *et al.*, 2005)[39]		
Type of Data	Example	Regression Model
Continuous	Birth weight (grams)	Linear Regression (ANOVA)
Dichotomous	Low birth weight? (< 2500 grams)	Logistic Regression
Ordinal	Birth weight (Very low, low, normal)	Ordinal Logistic Regression
Nominal	Cause of death	Polytomous Logistic Regression
Counts	Incidence rate	Poisson Regression
Time to event	Time to death	Cox Model

2. Subgroup analysis: Analysis of subgroups could be utilized in some RCTs.[27] In the case of multiple statistical tests, the chance of finding at least one test statistically significant due to multiple comparisons, and to incorrectly declare a difference can quickly increase with the number of such tests (a type I error). Multiple comparison correction methods such as Bonferroni correction are used in this case to make the outcome analysis more stringent and less likely to produce a type I error.

3. Missing data analysis: Missing data in RCTs could be adjusted for by many methods depending on the type of the data loss and its magnitude. Options include analyzing only cases with known outcomes and using imputed data.[27]

Regardless of the type of analysis, the following two considerations may also apply with respect to clinical trial analysis:

1. Interim Analyses: It is recommended that an RCT design include a pre-specified series of analyses as the data is being collected. The result of these analyses might suffice stopping the trial before the intended sample size is reached. For instance, participants in one group experience a "larger than expected benefit or harm," or if "investigators find evidence of no important difference between experimental and control interventions."[27]

2. Intent-to-Treat Analysis: In this type of analysis "data are analyzed in the way patients were randomized, regardless of whether or not they received the intended intervention."[27] A "pure" intention-to-treat analysis is "possible only when complete outcome data are available" for all randomized subjects.[33] For non-inferiority trials, intent-to-treat analyses may not lead to the most conservative findings; therefore, per-protocol

analyses where patients are analyzed depending on the actual treatment received may be more appropriate.

2.6 Disadvantages of RCTs

Although RCTs provide the gold standard of evidence, they do have disadvantages. Some of the most common explored shortcomings are listed below:[40]

2.6.1 Limited Generalizability (External validity)

Generalizability is the applicability of study findings to real life practice. Generalizability in RCTs could be restricted by many factors: [40]

1. Quality of intervention: RCTs conducted in advanced research institutions might use superior procedures compared to procedures used in real life practice.
2. Expertise of investigators: The expertise of the RCT medical team might be different from the expertise of medical teams in real life practice.
3. Setting and location: Findings of an RCT that was conducted on hospital patients might not be applicable to patients seen in clinics.
4. Participants' Characteristics: Which can include age, sex, severity of the medical condition, etc.
5. Rare side effects: Side effects might be too rare to manifest in the sample size of typical clinical trials.

2.6.2 Expenses

Although RCTs are usually expensive,[41] they have been shown repeatedly to be cost effective from a societal point of view. For example, Johnston *et al* [1] studied the cost and effect of 28 RCTs that were funded by the National Institute of Neurological Disorders and Stroke. Their total expense was around US$ 335 million, and their return produced a net benefit to society at 10 years of 46 times their total expense.

2.6.3 Pro-industry findings in industry-funded RCTs

The two main sources of most research funding are corporations (including pharmaceutical companies) and government (including universities and specialized government agencies). A small portion comes from charitable foundations that usually deal with a specific disease such as Cancer, AIDS, and Multiple Sclerosis.

Research has shown that results of RCTs supported by pharmaceutical industry are more likely to be influenced compared to results of RCTs supported by other sources of funding. This influence systematically favors positive findings for the products associated with the study sponsor.[42] A systematic review of 30 RCTs conducted by Lexchin *et al* supported this conclusion (odds ratio 4.05, 95% C.L. 3, 5.5).[43]

It is difficult to pinpoint the exact percentage of published RCTs that did not report funding, yet literature reviews have shown that a sizable portion of them failed to do so. For example, a survey of 370 drug RCTs showed a 29% failure to report rate.[44] Another survey of 519 RCTs published in December 2000 in the medical literature showed a failure to report rate of 34%.[45]

The influence of a funder on the result of an RCT might manifest in several ways and to variable extent. It could be at the level of design, method, analysis, and reporting. Consequently, it is crucial that the role of funders is described in sufficient detail in the published report. Similarly, if the funder plays no role in the conduct or reporting of a trial, a clear statement describing this should be made in the report.[46]

2.6.4 Conditions for Use of RCTs

Another disadvantage in conjunction with RCTs is the requirement of specific conditions for use of RCTs. To explain, the study exposure must be changeable for a RCT to be utilized to gain knowledge. Factors such as genetic traits, blood type, and family history are not modifiable, thus observation studies are more appropriate design methodology in this situation. Although some exposures such as smoking and marital status are changeable in principle, it is impractical to alter them by the investigators for the purpose of research. For example, all the available evidence on the negative health effects of smoking was obtained through observational studies.

There should be true lack of certainty concerning which intervention strategy is more beneficial. If there is existing evidence that drug A is superior to drug B, it would be unethical to give a group of people the drug B. Additionally, the primary endpoint is relatively common and prompt. The power of an RCT to detect a statistically significant effect is directly dependent on the number of endpoint events. Both rare and tardy endpoints require large sample sizes, which is usually impractical and/or not feasible. Therefore, RCTs are not the preferred tools when investigating interventions with rare or delayed outcomes.[40]

2.6.5 Statistical Error

Statistical analysis in RCTs is subject to two types of potential error:

Type I Error (Alpha Error): The error of rejecting a null hypothesis when it is actually true.[47] In other words, it occurs when an RCT falsely concludes difference between two interventions when a difference truly does not exist.

Type II Error (Beta Error): The error of failing to reject a null hypothesis when in fact it is false. In other words, it occurs when an RCT falsely concludes equality between two interventions when a difference truly does exist.

On the subject of Type II Error, a 1978 study stated that many published RCTs, which failed to reject their null hypotheses, did not actually have a large enough sample size to definitively support its conclusions. After three decades, this unfortunate situation can still be observed in the literature. Several studies documented in their review that a significant percentage of published RCTs still had erroneous or less than optimal sample size calculations, which are the basis for study power.[48, 49]

2.7 The Ethics of RCTs

Examining RCT from the ethical point of the view will be incomplete without understanding the principle of clinical equipoise which is defined by Freedman *et al* as "genuine uncertainty within the expert medical community... about the preferred treatment", [14] and referred to by Stolberg as "state of knowledge in which no evidence exists that shows that any intervention in the trial is better than another and that any intervention is better than those in the trial."[9]

This concept basically stems from the fine balance existing between possessing adequate proof that a new intervention is beneficial for a specific condition and hoping that a clear-cut proof of this benefit will be proved. If there is a proof that participants in one group in an RCT are more likely to benefit than participants in the other group, the design of this RCT is unethical. It is only ethical if RCTs are designed in areas of uncertainty, and should be conducted so long as the uncertainty exists.

Special considerations might arise from the potential conflict between clinical equipoise and benefit as perceived by patients, the public, and healthcare professionals.[18] For in-depth information on ethical concerns unique to RCTs Hellman [50] provide an excellent discussion on the subject.

Even though an informed consent (permission given by a competent patient based on understanding of all relevant facts) is almost universal in RCTs, studies have showed that many participants are under the impression that the treatment they received is favourable for their specific condition.[51] Additionally, the incorporation of RCTs in clinical research brought to existence cultural considerations that yet to be examined and understood.[52]

3.0 Reporting of RCT / Guidelines

3.1 Importance of RCT Reporting

It is important to differentiate between examining the quality of an RCT and the quality of its reporting. The quality of RCT as defined by Moher *et al* is "the confidence that the trial design, conduct, and analysis has minimized or avoided biases in its treatment comparisons." This main focus of this definition is the quality of the methodology. On the other hand, Moher's definition of the quality of an RCT reporting, which is our primary interest here, was "providing information about the design, conduct, and analysis of the trial."[53] It is important to keep in mind that combinations of a poorly designed RCT that is well reported or a well designed RCT that is poorly reported can occur in the medical literature.

Awareness concerning the quality of reporting randomized controlled trials is growing. Inadequately conducted trials are viewed as a waste of time, effort, and funds by funders and knowledge users alike. Similarly, well-conducted trials with inadequate reporting quality can represent a waste of the same resources due to impaired knowledge translation. Interestingly, poor reporting quality is paradoxically associated with an increased estimate of benefit for the intervention.[5]

A chief obstacle hindering the assessment of RCT quality is that the quality of reporting is often used as a proxy measure for methodologic quality. While research readers in most cases must rely on the information presented in the written report to judge a trial and make inferences, essential methodologic details may be omitted from these written reports. Low-quality reporting may hide important deficiencies in methodologic quality, and it may hide strength in well-conducted trials.[54] Devereaux *et al* stated "health care providers depend

upon authors and editors to report essential methodological factors in randomized controlled trials (RCTs) to allow determination of trial validity (i.e., likelihood that the trials' results are unbiased)."[55]

Studies have proven that there is an association between poor reporting and poor methodology in RCTs.[54] Schulz *et al* investigated this association; his conclusion was that "this study provides empirical evidence that inadequate methodological approaches in controlled trials, particularly those representing poor allocation concealment, are associated with bias. Readers of trial reports should be wary of these pitfalls, and investigators must improve their design, execution, and reporting of trials."[4] For example, trials that reported no exclusions are more likely to have impaired concealed allocation.[56]

Some studies have considered the poor quality of statistical analyses as the main reason for low quality reporting.[57,58] "Trials should have a clearer predefined policy for data analysis and reporting. The overuse of arbitrary significance levels (for example, $p < 0.05$) is detrimental to good scientific reporting, and more emphasis should be given to the magnitude of treatment differences and to estimation methods such as confidence intervals." [58]

The importance of reporting is paramount due to the sizable amount of taxpayer money, private funding, and charity fundraising invested year after year in cancer research. The average cost of journal subscription paid by a university library is $20,000 per year.[59] Scientific publishing has become a multi billion industry. For example, Elsevier Journals made £1 billion in pre-tax profit in 2003.[60] Beyond these significant financial considerations, the well documented incidents of research fraud play an equal and certainly more alarming reason to examine and improve reporting quality.[61] The "publish or perish" culture prevalent in academic medicine, together with the lack of consultation with statisticians, has been attributed to what has been dubbed "the scandal of poor

medical research".[57] The highest possible standards should be sought in the performance and reporting of medical research, especially in regards to RCTs.

The ability to evaluate the methodologic quality of RCTs is central to the appraisal of individual trials, the conduct of unbiased systematic reviews, and the performance of evidence-based health care. Whatever the outcome of a trial is, poor reporting might lead to misinterpretation of the trial's findings by the average reader, health care providers in this case. Based on such misinterpretation unfavourable changes to the clinical practice might occur, negatively affecting patients' care. It was said best by Devereaux *et al* "Until these inadequacies are resolved health-care providers will remain limited in their ability to make informed inferences about the validity of the studies upon which they base their clinical practice."[55]

3.2 RCT Reporting Guidelines / CONSORT

The broad goals of any reporting guidance are to improve the transparency and reporting of the specific clinical trial design. The CONSORT statement, which stands for CONsolidated Standards of Reporting Trials statement, is broadly considered the current standard in RCT reporting as evidenced by its adoption by important medical journal editorial groups. The CONSORT group described the statement as "an evidence-based, minimum set of recommendations for reporting RCTs. It offers a standard way for authors to prepare reports of trial findings, facilitating their complete and transparent reporting, and aiding their critical appraisal and interpretation."[62]

The statement encompasses various initiatives developed by the CONSORT group to alleviate the problems arising from inadequate reporting of

randomized controlled trials (RCTs). It is viewed as a part of the broader effort to improve the quality of research used in decision-making in healthcare.

3.2.1 CONSORT Statement

The CONSORT Statement consists of a 25-item checklist and a flow diagram along with some brief explanatory text. The statement also produced an elaboration document with more detailed descriptions and examples. "The CONSORT statement is intended to improve the reporting of a randomized controlled trial (RCT), enabling readers to understand a trial's design, conduct, analysis, and interpretation, and to assess the validity of its results. It emphasizes that this can only be achieved through complete transparency from authors."[6,63,64,65,66,67,68]

The CONSORT Statement was designed to improve reporting of the 'standard' two-group parallel RCT design, but was not designed to address reporting of the other types of RCTs, such as factorial and cluster RCT design. To help improve the reporting of these alternative trial designs, extensions and modifications of the statement have been underway by the CONSORT group, who collectively considers the statement to be an evolving and improving document. Therefore, it is subject to periodic changes as new evidence emerges. The most up-to-date revision of the CONSORT Statement is the 2010 revision, which can be freely viewed and downloaded from the CONSORT website (http://www.consort-statement.org). There are current efforts to develop and update extensions of the CONSORT Statement to address reporting quality for other types of RCT designs.

3.2.1.1 CONSORT Checklist

The checklist consists of 25 items that focus on the reporting of how the trial was designed, analyzed, and interpreted. Its main components are divided into "Title", "Abstract", "Introduction", "Methods", "Results", "Discussion", and "Other Information". These items were included in the checklist because "empirical evidence indicates that not reporting the information is associated with biased estimates of treatment effect, or because the information is essential to judge the reliability or relevance of the findings."[6,63,64,65,66]

A copy of the checklist is below (Figure 4). In depth Details of these items can be found in the CONSORT 2010 Explanation and Elaboration document.[27,69]

3.2.1.2 CONSORT Flow Diagram

The flow diagram (Figure 5) is intended to illustrate the passage of participants through the four phases of a parallel RCT of two groups. These four stages are enrollment, intervention allocation, follow-up, and analysis.[70,71] The diagram explicitly shows the number of participants, for each intervention group, included in the primary data analysis. The main function of the diagram is to provide enough transparency to the reader to judge whether the investigators have performed an intention-to-treat analysis.

3.2.2 CONSORT Explanation and Elaboration Document

The CONSORT group has endorsed a strong recommendation that CONSORT Statement be used in conjunction with the CONSORT Explanation and Elaboration Document. "This document is intended to enhance the use,

understanding and dissemination of the CONSORT Statement. Through examples and explanations, the meaning and rationale for each checklist item are presented."[27,69]

3.2.3 CONSORT Endorsers

By the end of 2010, CONSORT had been endorsed by 435 medical journals. To put this into context, it has been approved and promoted by over 50% of the core medical journals listed in the Abridged Index Medicus on PubMed.[72]

CONSORT has also been endorsed by Medical Research Support Foundation (MedicReS), and many editorial groups such as Council of Science Editors, International Committee of Medical Journal Editors (ICMJE), and World Association of Medical Editors (WAME). An up-to-date list of these journals and organizations that have endorsed this statement can be found on the CONSORT website (http://www.consort-statement.org).

Figure 4: CONSORT 2010 checklist (Items highlighted yellow are the optional items that were not considered in our study score)

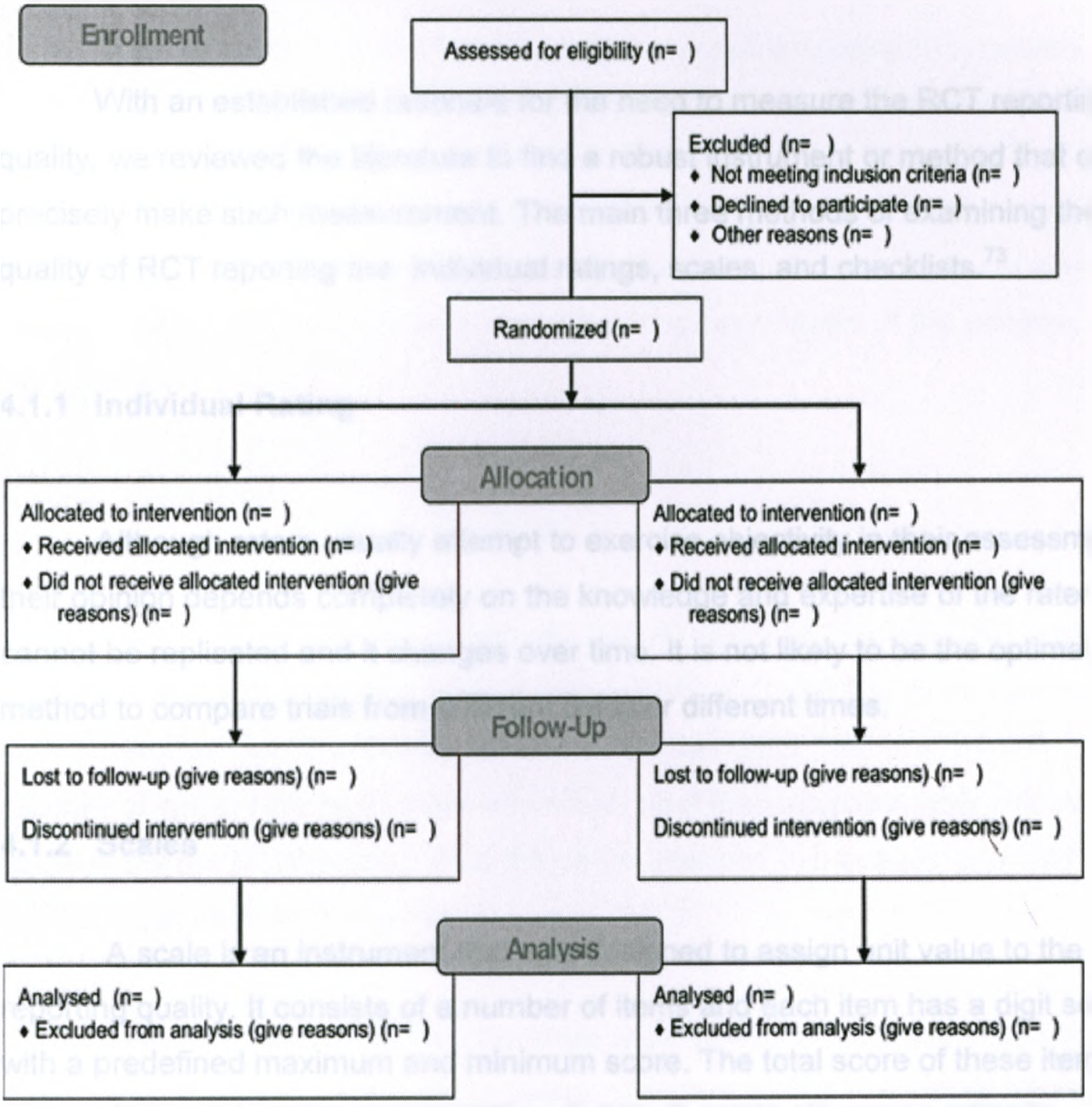
Section/ Topic	Item No	Checklist item	Reported on page No
Title and abstract			
	1a	Identification as a randomised trial in the title	
	1b	Structured summary of trial design, methods, results, and conclusions (for specific guidance see CONSORT for abstracts)	
Introduction			
Background and objectives	2a	Scientific background and explanation of rationale	
	2b	Specific objectives or hypotheses	
Methods			
Trial design	3a	Description of trial design (such as parallel, factorial) including allocation ratio	
	3b	Important changes to methods after trial commencement (such as eligibility criteria), with reasons	
Participants	4a	Eligibility criteria for participants	
	4b	Settings and locations where the data were collected	
Interventions	5	The interventions for each group with sufficient details to allow replication, including how and when they were actually administered	
Outcomes	6a	Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed	
	6b	Any changes to trial outcomes after the trial commenced, with reasons	
Sample size	7a	How sample size was determined	
	7b	When applicable, explanation of any interim analyses and stopping guidelines	
Randomisation:			
Sequence generation	8a	Method used to generate the random allocation sequence	
	8b	Type of randomisation; details of any restriction (such as blocking and block size)	
Allocation concealment mechanism	9	Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned	
Implementation	10	Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions	
Blinding	11a	If done, who was blinded after assignment to interventions (for example, participants, care providers, those assessing outcomes) and how	
	11b	If relevant, description of the similarity of interventions	

Statistical methods	12a	Statistical methods used to compare groups for primary and secondary outcomes	
	12b	Methods for additional analyses, such as subgroup analyses and adjusted analyses	
Results			
Participant flow (a diagram is strongly recommended)	13a	For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analysed for the primary outcome	
	13b	For each group, losses and exclusions after randomisation, together with reasons	
Recruitment	14a	Dates defining the periods of recruitment and follow-up	
	14b	Why the trial ended or was stopped	
Baseline data	15	A table showing baseline demographic and clinical characteristics for each group	
Numbers analysed	16	For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups	
Outcomes and estimation	17a	For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval)	
	17b	For binary outcomes, presentation of both absolute and relative effect sizes is recommended	
Ancillary analyses	18	Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing pre-specified from exploratory	
Harms	19	All important harms or unintended effects in each group (for specific guidance see CONSORT for harms)	
Discussion			
Limitations	20	Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses	
Generalizability	21	Generalizability (external validity, applicability) of the trial findings	
Interpretation	22	Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence	
Other information			
Registration	23	Registration number and name of trial registry	
Protocol	24	Where the full trial protocol can be accessed, if available	
Funding	25	Sources of funding and other support (such as supply of drugs), role of funders	

*We strongly recommend reading this statement in conjunction with the CONSORT 2010 Explanation and Elaboration for important clarifications on all the items. If relevant, we also recommend reading CONSORT extensions for cluster randomised trials, non-inferiority and equivalence trials, non-pharmacological treatments, herbal interventions, and pragmatic trials. Additional extensions are forthcoming: for those and for up to date references relevant to this checklist, see www.consort-statement.org.

CONSORT 2010 checklist of information to include when reporting a randomised trial*

Figure 5: CONSORT 2010 Flow Diagram



4.0 Measuring the Quality of RCT Reporting

4.1 Methods of Measuring the Quality of RCT Reporting

With an established rationale for the need to measure the RCT reporting quality, we reviewed the literature to find a robust instrument or method that can precisely make such measurement. The main three methods of examining the quality of RCT reporting are: individual ratings, scales, and checklists.[73]

4.1.1 Individual Rating

Although raters usually attempt to exercise objectivity in their assessment, their opinion depends completely on the knowledge and expertise of the rater. It cannot be replicated and it changes over time. It is not likely to be the optimal method to compare trials from different fields or different times.

4.1.2 Scales

A scale is an instrument that was designed to assign unit value to the reporting quality. It consists of a number of items and each item has a digit score with a predefined maximum and minimum score. The total score of these items reflects the quality of the trial reporting. Scales theoretically are superior to individual rating and checklists because they offer a quantitative value of the reporting quality.

Scales differ from one another in the number of items, how items were arrived at, what they measure, how reliable these items are in measuring what they were intended to measure, approximate time to complete scoring a trial, the range of the score, and how much weight each item is given. Using a specific

scale to measure reporting quality might introduce certain biases. For example, a scale that is constructed to give higher weight to blinding by definition penalizes surgical RCTs in which blinding might be impractical or unethical.[53]

Between 1981 and 1993 twenty-five scales were developed to assess RCTs.[53] Of these only three were designed to assess reporting quality. We have listed these three scales with some of their characteristics in Table (2). The Jadad scale is the only one that was validated using established methodologic procedures.[73] Although the Jadad scale was validated, its low number of items (n=3) renders it of limited value in evaluating the many facets of the complex RCT design.

4.1.3 Checklists

A checklist is an instrument that was initially designed in a stepwise method to guide authors in producing good quality reporting. Checklists could also be used to assess reporting quality as well. Checklists differ from one another in the number of items, how items were arrived at, what they measure, how reliable these items are in measuring what they were intended to measure, and how much weight each item is given.

Between 1961 and 1993 ten checklists were developed to assess RCTs. Of these, only three were designed to assess reporting quality.[53] All three addressed reporting of blinding, patient assignment and statistical analysis. None of the three took into account reporting of follow-up. We have listed these three checklists with some of their description parameters in Table (3).

All of the scales and checklists examined above have major weaknesses, shortcomings that were to be addressed by the development of the CONSORT statement and checklist. In summary, these weaknesses include poor

instructions on how to score, giving greater weight for some items, containing too many or too few items, and being specific to a certain design of RCTs or to a certain specialty.[53]

4.2 Using the CONSORT Checklist to Measure the Quality of RCT Reporting

To examine the use of the CONSORT Checklist as measurement of the quality of RCT reporting, we reviewed the literature for factors that should be considered to help us make this judgment. We found a credible list of such factors summarized in one of the most important works in the field of RCT reporting by Moher *et al.* [53] We measured the CONSORT Checklist according to these factors and summarize the assessment in the following four points:

1. Comprehensiveness of items: As demonstrated in Chapter 3 of this document, the CONSORT Checklist is very comprehensive. Furthermore, the checklist comprehensiveness could be established by two facts. First, it includes all items that could be found in all the checklists and scales presented earlier in this chapter. Second, it consists of 37 items, while the number of items in the presented checklists and scales ranged from 3 to 34.[53]

2. Ease of instructions: The checklist instructions were designed to be short, to the point, and straightforward.

3. Definition of the study population: The checklist was originally designed to improve reporting quality of the "standard' two-group parallel RCT design and was developed and adopted by many core medical journals.

4. Number of citations (studies that used the checklist as a scale): we can see in the coming chapter, almost all audits of reporting quality after the publication of the CONSORT Statement 1996 have used the CONOSRT checklist as a scale, whether in whole or in part. By the end of the year

2010 the number of such studies as reported in the CONSORT website was 95.[74]

Table 2: Description of three scales constructed to measure the quality of randomized controlled trials reporting

Name of Scale's Author with PubMed ID #	Generic vs. specific	Quality defined	# of items	Inter-rater agreement *	Average time to complete (minutes) **	# of citation ***	clear instructions on how to score	Reporting Blinding	Reporting type of assignment	Reporting type of statistical analysis	Reporting follow up
Andrew [75]	S	No	11	.95	10	N/A	Yes	Yes	Yes	Yes	No
Annals:[76]	G	Yes	34	.12	15	193	Yes	Yes	Yes	Yes	No
Jadad:[73]	G	Yes	3	.66 - .77	< 10	122	Yes	Yes	Yes	Yes	No

*** Inter-rater agreement as reported by authors**
**** Average time to complete scoring one study as reported by authors, or as estimated by :[53]**
***** Number of articles that cited the scale in PubMed**

Table 3: Description of three checklists constructed to measure the quality of randomized controlled trials reporting

Name of Checklist's Author with PubMed ID #	Generic vs specific	Quality defined	# of items	Inter-rater agreement *	Average time to complete (minutes) **	# of citation ***	clear instructions on how to score	Reporting Blinding	Reporting type of assignment	Reporting type of statistical analysis	Reporting follow up
DerSimonian: [77]	N/A	No	11		15	263	N/A	Yes	Yes	Yes	Yes
Grant :[78]	N/A	No	28		20	N/A	N/A	Yes	Yes	Yes	Yes
Mahon :[79]	N/A	No	4		10	101	N/A	Yes	Yes	Yes	No

*** Inter-rater agreement as reported by authors**
**** Average time to complete scoring one study as reported by authors, or as estimated by :[53]**
***** Number of articles that cited the scale in PubMed**

5.0 Previous Works in RCT Reporting Quality in the Literature

5.1 RCT Reporting in Several Medical Fields

Almost all studies investigating the reporting quality of RCTs have concluded that improvements in RCT reporting have been observed after the introduction of the CONSORT Statement, yet the overall quality of the trials remains unsatisfactory.[80] The following paragraphs summarize some of the important studies in this field.

In the Pediatric literature, Al-Namankany *et al* published a study in 2009 that assessed published RCTs in Paediatric dental journals between 1985 and 2006. The inclusion criteria as stated by the author were "(i) the trial was a randomized clinical trial; (ii) the trial was published between 1985 and 2006, and in English; (iii) the trial participants were infants and children, aged 18 years or under; and (iv) the article had been published in one of the paediatric dental journals specified." The two main aims were to determine "(i) whether quality of reporting allows readers to assess the validity of trials; and (ii) whether quality of reporting has improved since the introduction of the CONSORT guidelines". The report included 173 articles. The authors concluded that "the quality of reporting of clinical trials is poor, and often not adequate to allow readers to assess trial validity. Overall quality of reporting has not substantially improved since the publication of CONSORT". The authors sent letters to the journals included in the study, to ascertain their status regarding the adoption of the CONSORT guidelines. Only two of the five journals responded to the letter, stating that they have not yet adopted CONSORT statement.[81]

In Cardiology literature, Ethgen *et al* published a study in 2009 evaluating the quality of published reports of RCTs assessing stents for percutaneous

coronary interventions. Their sample size was 132 RCTs, and their conclusion was in line with the previous study in that the current reporting needs to be improved to allow readers to judge the risk of bias and the applicability of the results. One possible explanation provided by the authors for the poor quality was in reporting primary outcomes "in about half of the reports, the main outcomes relied on angiographic evaluation such as coronary restenosis or late lumen loss. These outcomes are surrogates of clinical events, ... and may prove challenging for the interpretation of results,"[82]

In a plastic surgery study published in 2008, Taghinia *et al* analyzed RCTs with respect to reporting standards, methodologic quality, and impact on the specialty as RCTs in plastic surgery have not been analyzed comprehensively before that date. Their analysis included 163 RCTs published from 1986 to 2006 in three major plastic surgery journals. They used the CONSORT checklist to score these RCTs. Their conclusion read, "there were deficiencies in the reporting of parameters that influence bias and statistical significance. The reporting and methodologic standards of randomized controlled trials in plastic surgery need improvement". The main areas with poor reporting quality identified by the study were statistical analysis, sample size determination, blinding, randomization, sequence generation, and allocation concealment.[83]

In the Dermatology literature, Adetugbo *et al* published a survey study in 2000 in which they examined the reporting quality of all published parallel group RCTs in Clinical and Experimental Dermatology from its inception in 1976 through 1997. As measures of reporting quality the authors examined the adequacy of randomization, trial sample size, baseline comparisons, and intention-to-treat analysis. A total of 68 RCTs were included in their analysis. Of these trials, only 1% reported the method of random sequence generation, 1% reported sample size and statistical power considerations and had an a priori main hypothesis, 7% reported adequate concealment of allocation. Among 38

trials that used simple randomization, the sample sizes in the comparison groups were identical in 22 occasions; raising the possibility that simple randomization might not have been adequately generated or concealed. Their final statement read "there is the need for higher methodological quality in clinical trial reporting in dermatology journals. The adoption of the CONSORT (Consolidated Standards of Reporting Trials) statement and checklist for the reporting of trials should enhance the validity of and strengthen the evidence from clinical trials reports".[84]

In the Urology literature we found a study published in 2007 by Scales *et al.* Their sample included 152 published RCTs from 1996 to 2004, and their assessment of reporting quality was based on the CONSORT checklist. The authors had two main conclusions: "reporting in the urology literature has improved since the publication of the Consolidated Standards of Reporting Trials statement in 1996...certain areas, such as reporting of trial methods, continue to meet Consolidated Standards of Reporting Trials criteria in fewer than half of publications". The areas with poor reporting quality identified by the study included: calculation to justify sample size, randomization method specified, allocation concealment, who generated allocation sequence, who enrolled participants, who assigned participants to groups, participants blinded, Intervention personnel blinded, assessors blinded, flow diagram, and intent to treat analysis.[85]

In Occupational Therapy, a study by Moberg-Mogren *et al* investigated the quality of reporting in published RCTs by using a modified CONSORT checklist. The kappa statistics computed on individual items ranged from .58 indicating high levels of agreement for most items to .40 indicating low levels of agreement for some of the items. The study concluded that a few of the CONSORT items are impossible to comply with in most occupational therapy research, such item

11 "Blinding". However, most of the items are possible to report as recommended by CONSORT.[86]

A systematic review by Plint *et al* analyzed the results of 8 studies that were conducted to determine whether the adoption of the CONSORT checklist is associated with improvement in the quality of reporting of randomized controlled trials (RCTs). Their unit of analysis was published reviews on CONSORT adherence. Their search was able to identify 8 eligible studies published between 1996 and 2005. Their results proved an association between journal adoption of CONSORT Statement and improved reporting of RCTs.[87]

Pat *et al* published a study in 2008 that examined adherence to the CONSORT statement in RCTs with information on symptom control and quality of life during chemotherapy for advanced non-small cell lung cancer. The CONSORT Checklist was used as a scale for adherence to the statement (a proxy for reporting quality). On the contrary to the findings in the previous studies (above), the overall adherence of RCTs to CONSORT in this study was found to be acceptable with no clear sign of change over time.[88]

Although most of the studies above have shown improvements in RCT reporting quality after the introduction of the CONSORT Statement with unsatisfactory overall quality, others have shown contradicting findings. This uncertainty in the literature warrants further research.

5.2 RCT Reporting in Cancer Research

Very few studies have investigated the reporting quality of RCTs published in Oncology journals.[49] Our literature search was only able to identify

six studies, none of which directly examined RCTs reporting in breast, prostate, colorectal, or lung cancer. Following is a brief description of these studies.

One of them is the study mentioned above published by Pat *et al* in 2008. Another example is a study published in 2008 by Mathoulin-Pelissier *et al.* This study limited its evaluation to reporting of survival endpoints in 104 phase III trials. Their initial electronic search and review of abstracts identified 274 cancer RCTs while their full article review revealed that only 104 of these articles were indeed RCTs. Their main conclusion was stated as "A majority of articles failed to provide a complete reporting of survival endpoints, thus adding another source of uncontrolled variability".[89]

Ziogas *et al* published an article in 2009 in which they evaluated the reporting quality of published RCTs concerning myeloid hematologic malignancies. Quality of reporting was assessed using a 24-item questionnaire based on the CONSORT checklist. Their search identified 261 eligible RCTs. Their findings were summarized, as "Quality of reporting in RCTs focusing on myeloid malignancies remains unsatisfactory. Further improvement of reporting is necessary to assess the validity of clinical research".[90]

Kober *et al* published a study in 2006 that examined reporting quality as assessed by adherence to the CONSORT statement in published RCTs with information on patients with Hodgkin lymphoma. The sample size was 242 RCTs and the quality of reporting was assessed using a 14-item questionnaire based on the CONSORT checklist. Reporting was evaluated in two pre-CONSORT periods (1966-1988 and 1989-1995) and one post-CONSORT period (1996-2002). Their main conclusion was "Despite recent improvements, reporting levels of CONSORT items in RCTs involving patients with Hodgkin lymphoma remain unsatisfactory".[91]

In 2009, Bekelman *et al* published a study in which they hypothesized that radiotherapy RT reporting may be inadequate in Hodgkin's lymphoma (HL) and non-Hodgkin's lymphoma (NHL) phase III RCTs. They searched PubMed and the Cochrane registry for relevant RCTs published between 1998 and 2007. Their initial abstract review identified 133 RCTs while their full article review revealed that only 61 of these articles were indeed RCTs. Their assessment depended on the presence of six quality measures: target volume, radiation dose, fractionation, radiation prescription, quality assurance (QA) process use, and adherence to QA. Their main conclusion was stated as "Reporting of RT in HL and NHL RCTs is deficient."[49]

Bentzen *et al* published a study in 1998 to assess the quality of the design, analysis, and reporting in RCTs with information on radiation oncology. The authors were motivated by special concerns in relation to the reporting of radiotherapy RCTs. One of their conclusions was "we need to improve the quality of RCTs in terms of their design, conduct, analysis and reporting."[48]

Although the reporting quality of RCTs has been investigated adequately in RCTs published in different fields of medicine, such an investigation seems to be lacking in RCTs published in oncology journals, which in the opinion of our research team warrants further research.

5.3 Predictors of Reporting Quality

Although several studies have evaluated the quality of RCTs published in medical journals, our review revealed that very few studies tried to determine the predictors of CONSORT checklist compliance.

The first example is a study that was conducted by Lai *et al* and published in 2006. It found that impact factor, publication after 1995, and sample size more than 280 were significant factors associated with better overall reporting quality. It also concluded that the reporting quality of RCTs in the primary treatment of brain tumors is suboptimal.[80] The second example is a study that was conducted by Barbui *et al* to assess whether the impact factor is a proxy measure of the reporting quality. The sample included 132 RCTs and the results revealed that the impact factor is not a valid measure of reporting quality.[92]

Our review revealed that studies identifying predictors of CONSORT checklist compliance are very few. This may be due, in part, to the relatively recent adoption of the checklist, which did not allow enough time for thorough understanding of the predictors and the relationships among them. The importance of determining predictors of CONSORT checklist compliance stems from the fact that identifying and overcoming the obstacles of improving RCT reporting quality may help health-care practitioners and research consumers to make informed inferences about the validity of the studies upon which they base their decisions.

5.4 Reliability of Individual CONSORT Items

Again, although several studies have evaluated the quality of RCTs published in medical journals, few studies tried to investigate the reliability of the CONSORT Checklist items (the following are two examples of such studies), and none were conducted with a focus on cancer research.

In 2007, Farrokhyar *et al* published a study examining the quality of reports of RCTs in coronary artery bypass grafting (CABG) surgery when

comparing off- and on-pump techniques. Their data came from electronic searches of MEDLINE, the Cochrane Library, CINAHL, HealthSTAR and EMBASE, and they used the CONSORT Checklist to score the quality. Two interesting points were mentioned in this publication: "The kappa value was greater than 0.6 for 73 of 105 (70%) items", and "the quality of the publications' reporting methods, results and discussion sections was suboptimal".[93] Another study by Moberg-Mogren *et al* demonstrated similar findings. The kappa statistics computed on individual items of the CONSORT checklist ranged from 0.58 to 1.00 indicating high levels of agreement for most items. However, some item kappas fell below the moderate level of agreement 0.40.[86] This demonstrates that there might be real weakness in the working definitions of the checklist items. If such results are confirmed, improving the reporting quality of RCTs may be performed by identifying the unclear items and improving their reliability in future versions of the CONSORT checklist.

The reliability of the CONSORT checklist items is crucial for improving the reporting quality of RCTs. It enhances the use, understanding, and dissemination of the CONSORT statement. Because of this importance, the CONSORT group has a strong recommendation that the statement be used in conjunction with the CONSORT Explanation and Elaboration Documents. These documents present the meaning and rationale for each checklist item.[27, 69]

As we stated above, very few studies tried to investigate the reliability of the CONSORT Checklist items, and none of them did that in RCTs published in Oncology journals. Further research inquiry in this area is warranted.

5.5 Clinical Articles Inaccurately Presented as RCTs

It was interesting to find that the literature contains many clinical articles that were inaccurately presented as RCTs. For example, Mills *et al* conducted a study to examine the CONSORT compliance in clinical pharmacology journals. Two points were reported. First, of the 482 clinical trials included in the initial search, only 193 were considered to be RCTs after study review. Second, the use of the certain CONSORT items was questionable in these journals, possibly because many items may not be relevant to clinical pharmacology research.[94] Another example is one of the cancer studies mentioned above.[89] The initial electronic search and review of abstracts in this study identified 274 cancer RCTs while their full article review revealed that only 104 of these articles were indeed RCTs.

5.6 Summary of the Literature Review

The main points of our review can be summarized in the fact that to date there has been no study that assessed the reporting quality of published RCTs involving common cancers. Research on determining predictors of CONSORT checklist compliance and on examining the reliability of the CONSORT Checklist items is lacking in the literature and especially in the oncology literature. Also, scientific data are lacking with regards the reliability of the CONSORT checklist items.

Based on the limited reliable information on the reporting quality of published RCTs in oncology and on the importance of cancer as a disease entity associated with significant morbidity and mortality, we decided to study the predictors of CONSORT checklist compliance in the oncology literature. We also decided to examine the reliability of the CONSORT Checklist items, and to

examine whether the reporting quality of RCTs in the oncology literature has improved over the past two decades.

6.0 Objectives and Hypotheses

6.1 Study Objectives

1. To determine predictors of CONSORT checklist compliance in the oncology literature over the past two decades.

2. To examine the reliability of the CONSORT checklist items.

3. To examine whether the quality of reports of randomized trials has improved over the past 20 years.

As we demonstrated in the Chapter 5, these three questions have been previously unanswered. Scientific data are lacking with regards to the first two questions, and evidence is inconsistent concerning the third one.

6.2 Study Hypotheses

Null Hypothesis: There is no association between CONSORT statement compliance/accuracy and the following variables Journal Name, Type of Cancer, Publication Year, Number of Authors, Number of Patients, Intervention, Trial Site, One vs. Multiple Countries, Primary Country, Cooperative Group, Impact Factor, Oncology vs. Non-oncology Journal.

Alternate Hypothesis: An association exists between CONSORT statement compliance/accuracy and the following variables Journal Name, Type of Cancer, Publication Year, Number of Authors, Number of Patients, Intervention, Trial Site, One vs. Multiple Countries, Primary Country, Cooperative Group, Impact Factor, Oncology vs. Non-oncology Journal.

7.0 Methods

7.1 Study Design

The study design is a cross-sectional CONSORT compliance audit of published parallel two-arm RCTs assessing oncological interventions in adult breast, prostate, colorectal, and lung cancer between 1992-2010.

7.2 Setting and Relevant Dates

This study was conducted in collaboration between the Departments of Oncology and Epidemiology & Biostatistics, University of Western Ontario, London, Ontario, Canada. The selection of the study sample was conducted between May and June 2010. Data collection was performed between June and November 2010. Data analyses were performed between November 2010 and January 2011.

7.3 Study Population

7.3.1 Inclusion/Exclusion Criteria

In choosing our inclusion and exclusion criteria, we desired to select RCTs that the majority of oncologists may be exposed to during their years of practice.

These criteria are described in the following two subsections.

7.3.1.1 Inclusion Criteria:

1. Published phase III RCTs from 1992 to 2010: To meet this criterion, a trial must have been a prospective study that assessed healthcare interventions in human participants who were randomly allocated to study groups.
2. English language
3. Involving adults: A specific age range is not specified as many of the RCTs included in the study did not report such ranges.
4. Cancer type: Breast, Prostate, Colon, and Lung (four most common adult solid tumors)
5. Parallel group design: constitutes the majority of published RCTs
6. Studies published in journals that published >4 RCT studies in the 20 years period: Practically, journals that publishes less than 5 RCTs on one of the four cancer types in a period of about 20 years are not ideally considered journals that routinely publish on oncology topics.

7.3.1.2 Exclusion Criteria

1. Non-English language reports
2. Investigations reporting interim analysis that did not result in stopping the trial
3. Secondary and long-term update (primary report available) analyses
4. Pilot/phase 2 studies
5. Trial that did not employ a parallel design such as cross-over, factorial, cluster, split-body and multiple arm trial. (For definitions please see chapter 2)
6. Duplicate reports

7. Cost effectiveness and economic studies
8. Trials studying benign tumours or pre-cancerous lesions
9. Trials studying cancers other than the four mentioned in the inclusion criteria, or a combination of two or more of these four cancers.

7.3.2 Selection Methods

Selection of reports was performed in three stages:

7.3.2.1 Database Search (Stage 1)

A professional librarian at the London Regional Cancer Program conducted a search of PubMed database for RCTs in compliance with the study inclusion criteria. SEARCH STRATEGY: randomized controlled trials as topic[mh] AND (quality control[mh] OR guideline adherence[mh] OR guidelines as topic[mh] OR publishing/standards[mh] OR publication/standards[mh]). Eight-hundred and fifty parallel two-arm RCTs assessing oncological interventions in adult breast, prostate, colorectal, and lung cancer between 1992-2010 were identified.

7.3.2.2 Titles and Abstract Review (Stage 2)

One reviewer screened the titles and abstracts of the 850 retrieved reports to exclude any obvious reports of non-eligible trials. Of these, 515 RCTs were deemed eligible for inclusion in a full article review.

7.3.2.3 Full Article Review (Stage 3)

A copy of the full article was obtained for each of the 515 included reports with any additional material about the article such as an appendix on the journal that published the article. Two qualified reviewers conducted a full article review of the 515 non-excluded reports. One of the two reviewers is a Canadian Medical Graduate involved in this work during his Oncology residency training and the other is an International Medical Graduate (with Canadian medical qualifications) involved in this work as a requirement to complete a Clinical Epidemiology Master's program. This review had three goals, first to exclude any reports of non-eligible trials, second, to score the included reports using the CONSORT checklist, and third, to collect data on specific variables for further analyses. Of the 515 RCTs, 408 RCTs were deemed eligible for inclusion into the RCT oncology database. Each one of these RCTs was given a number from 1 to 408 termed as ID Number in order to provide a straightforward unique clinical trial identifier for data management.

7.4 Variables

7.4.1 Primary and Secondary Outcomes

Primary Outcome Measure: The average of two "CONSORT Scores" (see scoring procedure below) for each RCT is termed the CONSORT average score. This average is considered a measure of quality as measured by two independent reviewers.

Secondary Outcome Measure: The difference between these two scores is termed the CONSORT difference of scores. This difference reflects disagreement between the two raters, and is considered an estimate of reliability - i.e. the higher the difference score, the less reliable the checklist is for that specific clinical trial.

7.4.2 Variables (Predictors)

1. Journal Name: The name of the journal in which the trial was published.
2. Type of Cancer: The type of cancer under investigation in the trial. Categorized as a nominal variable.
3. Publication Year: The year in which the trial was published. We categorized this continuous variable into three groups: (1992-1996), (1997-2001), and (2002-2010).
4. Number of Authors: The number of authors of the trial. In the main effects model, we used this variable as a continuous variable. In the descriptive analysis, we categorized this continuous variable into groups of three authors. We considered groups of three authors as a reasonable and meaningful number to present these data for histogram purposes. The number of authors in our sample varied between 1 – 36 authors. Categorizing the variable into groups of three authors reflects an increase of about 10% that is traditionally considered the rule of thumb in epidemiology if standard categories did not exist in the literature.
5. Number of Patients: This is equivalent to the trial sample size. In the main effects model, we used this variable as a continuous variable. In the descriptive analysis, we categorized this continuous variable into groups of 250 patients. We considered groups of 250 patients a reasonable and meaningful number to present these data. The number of patients in our sample varied between 30- 5187 patients. Categorizing the variable into groups of 500 (according to the 10%) will cause most of the RCTs to fall in one group as most of them in our study have a sample size of less than 500.
6. Intervention: The intervention under investigation in the trial. Nominal variable categorized into four groups: Radiation, Chemotherapy, Surgical,

and Multiple Therapy (Any combination of Radiation, Chemotherapy, and Surgical).

7. Trial Site: The number of sites where the trial was conducted. Binary variable, 0 if the trial was conducted in one site, and 1 in two sites or more. For example, if a trial was conducted in four hospitals, the Trial Site value is 1.
8. One vs. Multiple Countries: The number of countries where the trial was conducted. Binary variable, 0 if the trial was conducted in one country, and 1 in two countries or more.
9. Primary Country: The name of the country where the trial was originated. Nominal variable.
10. Cooperative Group: Whether the trial was conducted by a cooperative group or not. Binary variable (1 for cooperative group, and 0 for non-cooperative group).
11. Oncology vs. Non-oncology: Whether the trial was published in an Oncology or Non-oncology journal. Binary variable (1 for Oncology, and 0 for Non-oncology).
12. Impact Factor: The impact factor (IF) is a measure reflecting the average number of citations to articles published in science and social science journals. [95] We dichotomized this ordinal variable into two groups: Low ≤ 10, and High >10. After completion of data collection, we found that these cut-off points can be written as Low ≤ 7.667, and High ≥ 14.069 as our sample did not include trials from journals with impact factors in between 7.667 -14.069. The cut-off point of 10 was chosen because the literature suggests that "good" journals are generally going to have an IF greater than 10.[96]

7.4.3 Potential Confounders

Both the association between CONSORT Average Scores and the predictors, and the association between CONSORT Difference of Scores and the predictors might be confounded by several variables. Here is a list of these potential confounders:

1- Journal Name: Potential confounder for both Impact Factor and Oncology vs. Non-oncology.
2- Trial Site: Potential confounder for both Number of Patients and Number of Authors.
3- Cooperative Group: Potential confounder for both Number of Patients and Number of Authors.
4- Type of Cancer: Potential confounder for intervention.
5- One vs. Multiple Countries: Potential confounder for trial site.

These potential confounders were suspected through insight in the relationship between the variables. To identify confounders we used the classical criteria of confounding "A variable is a confounder if it is associated with exposure and causally related to the outcome." For example, the variable "one vs. multiple countries" is suspected to be a potential confounder because it fulfills the first condition (It is logically associated with exposure "trial site" as a trial conducted in multiple countries, by definition, has more than one trial site), and may fulfill the second one (may be causally related to the outcome "reporting quality"). Disproving one of these two conditions is sufficient to rule out the confounding effect. This could be done through the main effects model by showing whether associations between variables (potential confounders) and the outcome exist.

7.5 Data Sources and Measurement (Scoring)

Compliance with the CONSORT checklist guidelines was assessed by two qualified reviewers in order to generate average and difference CONSORT checklist scores. We used the CONSORT Checklist as an extraction form. We included only 25 items of the checklist 35 items. The excluded items were the items defined by the checklist as optional (highlighted yellow in checklist, and the last three items in the checklist Other Information items Figure (4). Our rationale for these two exclusions was as follows:

1. Inclusion of optional items might penalize trials in which these items do not apply. For example, inclusion of Blinding, an optional item, penalizes most of the surgical trials in which Blinding is impossible or impractical.
2. The last three items in the checklist were added to the 2010 revision of the checklist. Inclusion of these items will not measure adherence to CONSORT Guidelines published before 2010.

Each RCT was given a score out of 25, reflecting how many of the 25 extraction form items were complied with (with each item being given equal weighting), this score was termed the "CONSORT score". Before data collection started, and to ensure similar understanding of the scoring process by the two reviewers, a sub-sample of 10 articles was randomly selected from the sample of articles included in Stage 3. The two reviewers discussed the interpretation of the different items. Differences primarily lend themselves to differing interpretation of the data extraction form items. In the event of disagreement, discussion took place with the senior investigator (Dr. George Rodrigues) until concordance was reached.

7.6 Efforts to Address Potential Sources of Bias

To reduce potential measurement bias in:

1. Stage 3: The optional items were excluded when constructing the extraction form.

2. Stage 3: The two reviewers independently scored the articles
3. Analysis: The small number of predictors relative to the large sample size (408 articles) ensures minimal biased estimates of the regression coefficients.

To reduce potential inter-observer error in Stage 3:

1. To reduce the number of data entry steps, data were directly entered into the database (Excel Sheet).
2. To ensure similar understanding of the scoring process by the two reviewers, a subsample of 10 articles were randomly selected for establishing definition, and to assess intra-observer agreement respectively.
3. Quality assurance on all steps of data collection, entry, and management was performed. Twenty percent of the overall sample was randomly selected and evaluated again by each one of the reviewers to double-check the scores for inter-observer error.

To reduce potential selection bias:

1. In Stage 2: Of the 850 articles included through Stage 1, the senior investigator and a qualified reviewer initially screened 200 articles. Discussion took place when screening each of these 200 articles to ensure complete understanding of the inclusion and exclusion criteria.
2. In Stage 3: The choice of the ten-article subsample was random. A digit between 0 and 9 was chosen randomly, and then articles with an ID number ending with this digit were included in the sub-samples.

7.7 Study Sample Size

A traditional sample size calculation is not possible as there is no known estimate of what clinical importance may be in the case of RCT reporting quality. Therefore we considered estimating the sample size as a function of the number of coefficients that can be safely included in the study analysis. The commonly used rule of thumb is ten observations per variable, a ratio of 10:1.[97]

Since we intended to analyze the association between two outcome measures and eleven predictors (13 variables or coefficients), according to the above rule the minimum sample size is 130 RCTs. Any ratio greater than 10/1, a sample size bigger than 130 RCTs, should generate regression coefficient estimates that are precise. The greater the ratio, the more precise regression coefficient estimates are likely to be. "The model which optimizes the bias-variance trade-off is by definition the model which minimizes prediction error."[98]

7.8 Quantitative Variables

Publication Year: We categorized this continuous variable into three groups. The cut-off points were chosen to allow us to examine the difference in reporting quality among three time periods: Pre-CONSORT (1992-1996), between the release of CONSORT and its first revision (1997-2001), and between the release of the first CONSORT revision and its second revision (2002-2010).

Impact Factor: We dichotomized this ordinal variable into two groups: Low ≤ 7.667, and High ≥ 14.069.

7.9 Statistical Methods

7.9.1 Statistical Analysis

We calculated Kappa statistics for each individual CONSORT item and for the total scores on the entire sample of 408 articles. Also, descriptive summary statistics were calculated for all variables. The main analysis was performed by constructing two main effects models. In the first main effects model we looked at the association between the predictors (Intervention, Year of Publication, Trial Site, Cooperative Group, Cooperative group, Oncology Journal type, Number of Authors, Number of Patients, and Impact Factor) and the CONSORT Average Score (predictors of quality). In the second, we looked at the association between the above predictors and the CONSORT Difference of Scores. Backward elimination analysis was used to identify all potential confounders. A p-value of <0.05 is considered statistically significant. All statistical analyses were conducted using SAS Software 9.2 (SAS Inc. North Carolina, USA).

7.9.2. Ethics Approval

This study used previously published data making it exempt from institutional review board approval.

8.0 Results

8.1 Participants

The PubMed search produced 850 potentially eligible publications. This group was examined for eligibility by screening the titles and abstracts. A total of 335 studies were excluded and 515 publications were found eligible in this stage. After a full article review of these 515 publications, a total of 107 studies were excluded and 408 publications were confirmed eligible for inclusion in the final analysis. Figure (6) presents a flowchart depicting the study screening process.

8.2 Descriptive Data

Results from the descriptive analyses carried out to assess demographic characteristics of the 408 RCTs included in the study are shown in Figure (7). Frequency by year of publication shows that the number of RCTs published in the three time periods (1992-1996), (1997-2001), and (2002-2010) were 51, 84, 273 RCTs respectively. Frequency by trial site shows that most of the trials were conducted in more than one site (377 RCTs), and much fewer were conducted in one site (31 RCTs). Frequency by number of countries shows that 156 RCTs were conducted in multiple countries and 252 RCTs were conducted in one country. Frequency by type of intervention shows that 13 RCTs investigated radiation therapy, 349 RCTs investigated chemotherapy, 1 RCT investigated surgical therapy, and 45 RCTs investigated a combination of the previous three therapies.

Frequency by type of cancer shows that 273 RCTs investigated lung cancer, 135 RCTs investigated breast cancer, 41 RCT investigated prostate

cancer, and 59 RCTs investigated a colorectal cancer. Frequency by journal shows that Journal of Clinical Oncology published a significant proportion of our sample (178 RCTs). Annals of Oncology published 52 RCTs. Each one of the other journals published < 21 RCTs. Frequency by primary country shows that 107 RCTs originated in the United States. Each one of the other countries published < 35 RCTs. Frequency by Oncology vs. Non-oncology shows that 374 RCTs were published in Oncology journals vs. 34 RCTs were published in non-oncology journals.

Figure (7) provides information on the predictors under investigation (Journal Name, Type of Cancer, Publication Year, Number of Authors, Number of Patients, Intervention, Trial Site, One vs. Multiple Countries, Primary Country, Cooperative Group, Oncology vs. Non-oncology, and Impact Factor). Category boundaries for quantitative data, for Sample Size, Number of Authors, and Impact Factor are displayed in Figure (7).

Our data did not find any significant confounding effect of the suspected potential confounders. This study used previously published data; therefore, there was no missing data for any of the variables of interest.

8.3 Outcome Data

Four hundred and eight articles were approached for descriptive analysis. Our primary outcome was the CONSORT Average Scores; the mean average CONSORT score was 16.6 (SD 3, max 25). Our secondary outcome was CONSORT Difference of Scores; the median difference score was 2 (interquartile range 1-3).

Four sets of descriptive statistics are presented in table (4) for all articles. Three of them are descriptive analysis of (mean, median, range): 1. Descriptive

statistics for the CONSORT Scores by the first reviewer, 2. Descriptive statistics for the CONSORT Scores by the second reviewer, 3. Descriptive statistics for the CONSORT Average Scores. The fourth set is a descriptive analysis (including interquartile range, median, interquartile deviation).

Figure (8) presents the distribution of the two outcome measures. The CONSORT Average Scores show a bell-shape distribution, while the CONSORT Difference of Scores shows a distribution that is skewed to the right. Four hundred and eight articles were approached for calculation of the one-way, two-way random, and two-way intraclass correlation coefficients (ICC) between the CONSORT Scores generated by the two reviewers. Good correlation was seen between the two raters; the overall two-way intraclass correlation coefficient was 0.71 (95%CI 0.61-0.78) for comparison of overall CONSORT score between raters. Figure (9) presents a scatter plot of this correlation. Table (5) presents the one-way, two-way Random, and two-way missed intraclass correlation coefficients (ICC).

Four hundred and eight RCTs (entire sample) were approached regarding the reliability in the final analysis. Kappa agreement for each individual CONSORT checklist item ranged from (0.02-0.92). Percent agreement for each individual CONSORT checklist item ranged from (30.9 - 97.8%). Table (6) presents simple Kappa statistic and percent agreement for each of the CONSORT checklist items.

Figure 6: Flowchart depicting the study screening process. The PubMed search produced 850 publications. From this group, 515 publications were eligible after screening the titles and abstracts. From this group, 408 publications were eligible for inclusion in the analysis after a full article review.

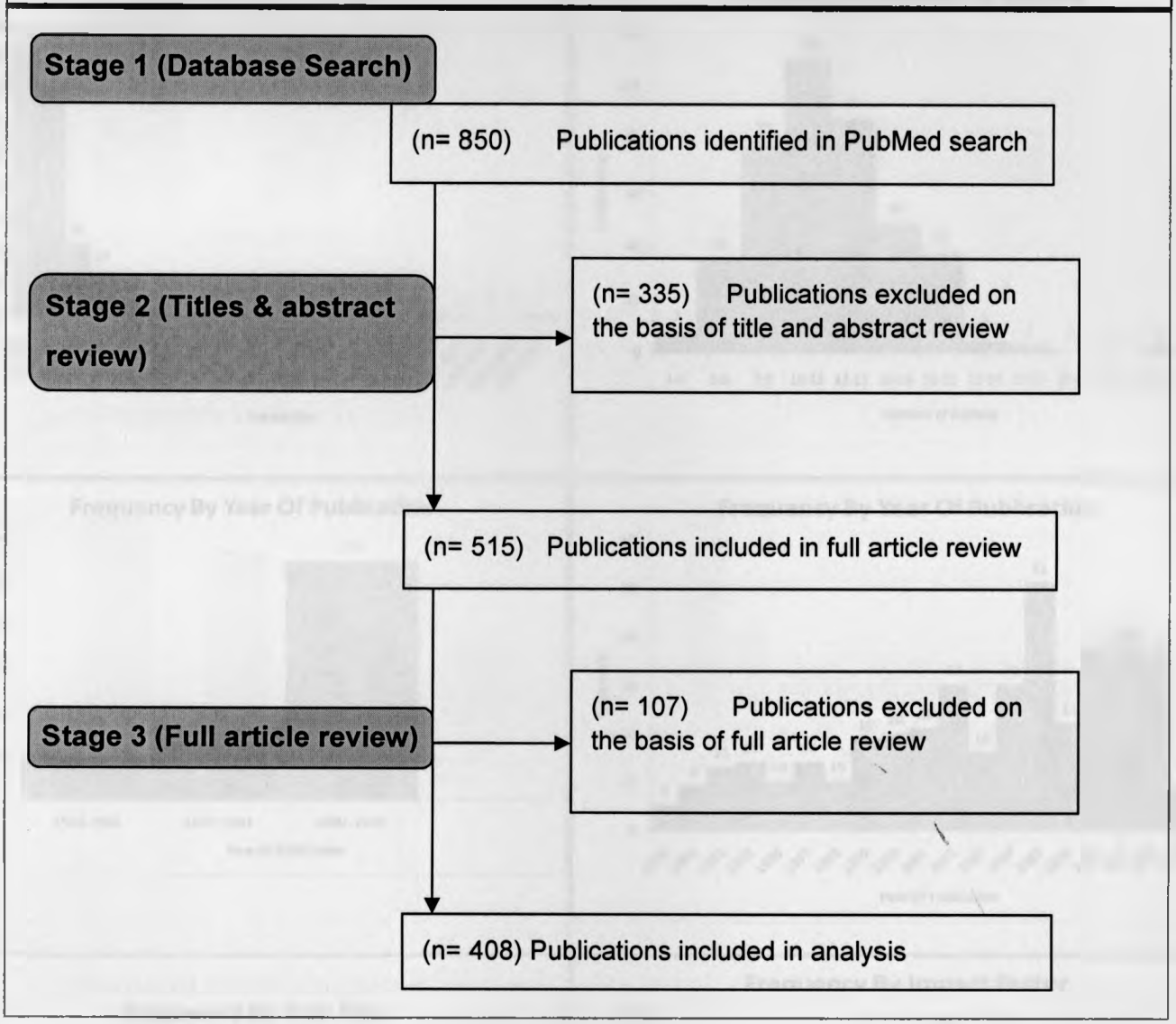
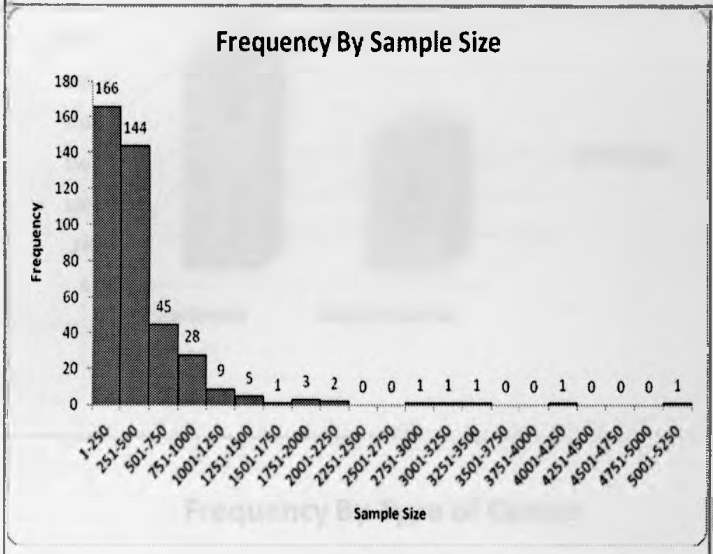
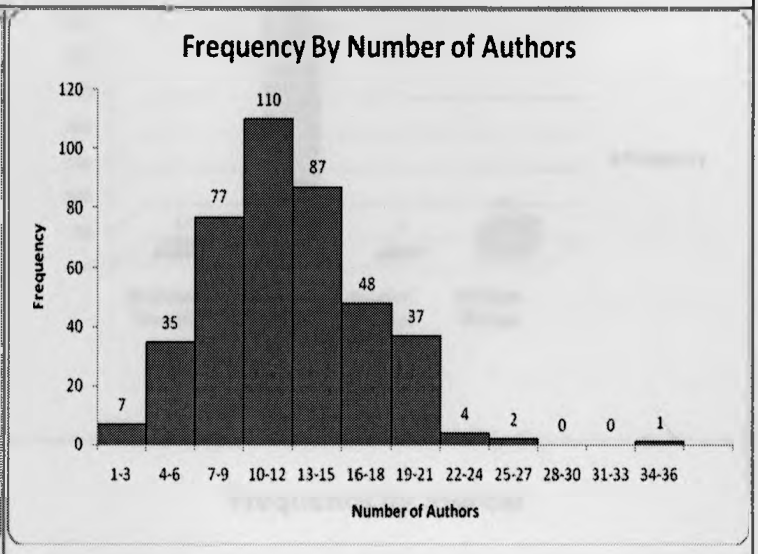
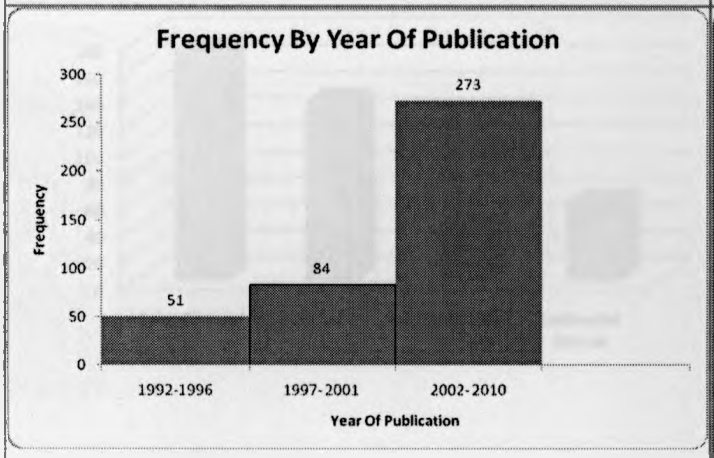
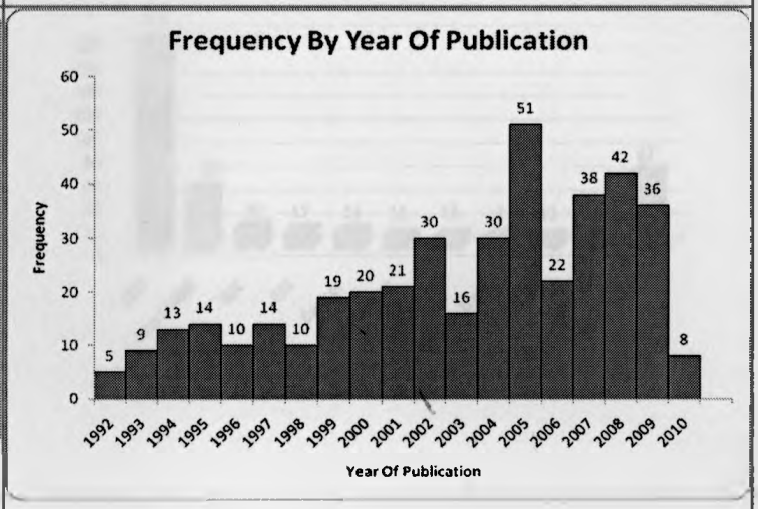


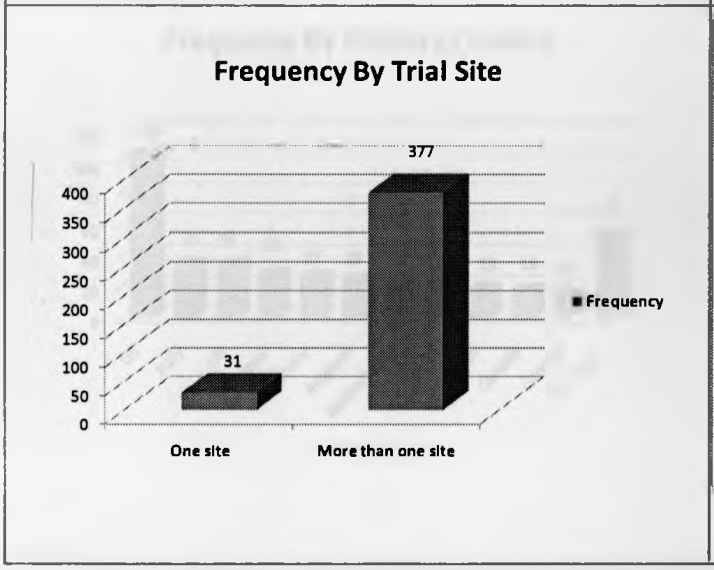
Figure 7: Characteristics of 408 RCTs and Information on Reporting Quality Predictors

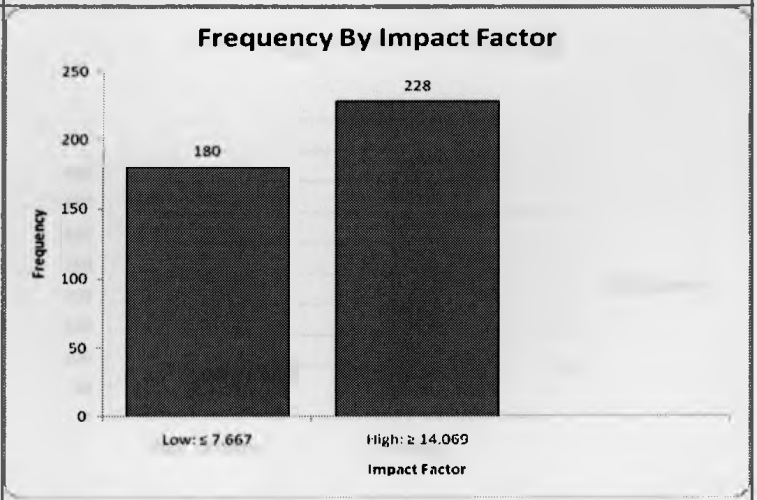




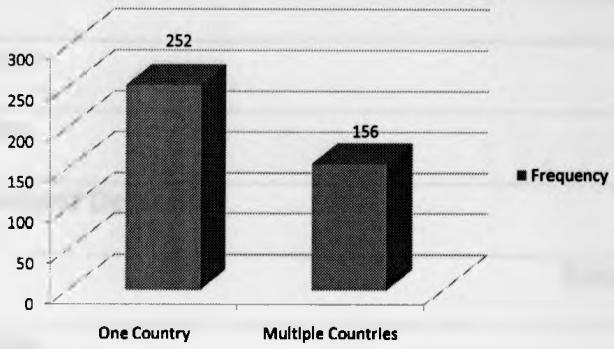




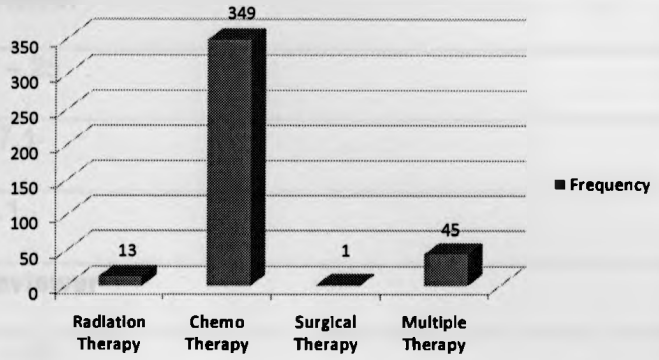




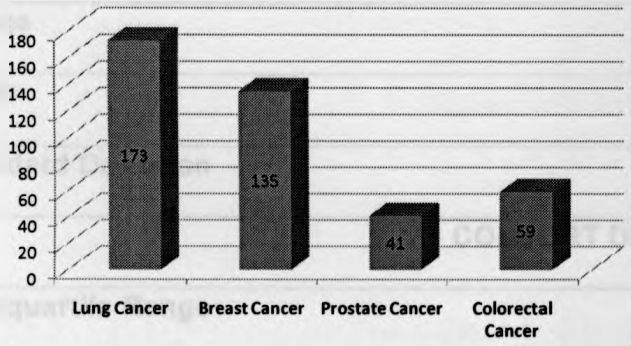
Frequency By Number Of Countries



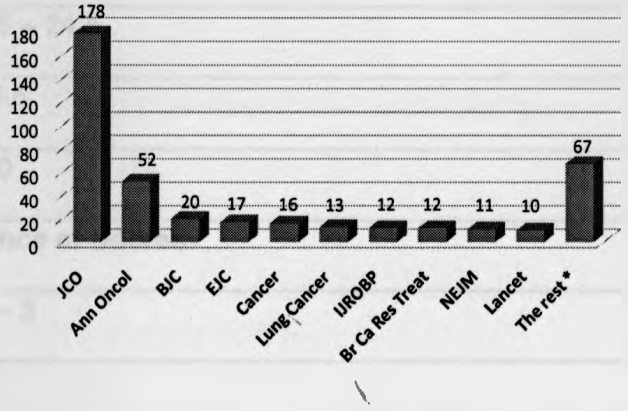
Frequency By Intervention



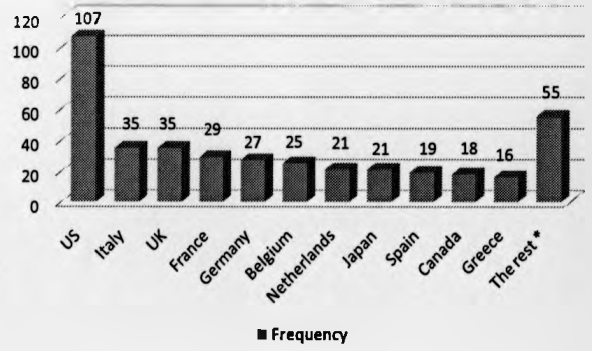
Frequency By Type of Cancer



Frequency By Journal



Frequency By Primary Country



Oncology VS Non-oncology

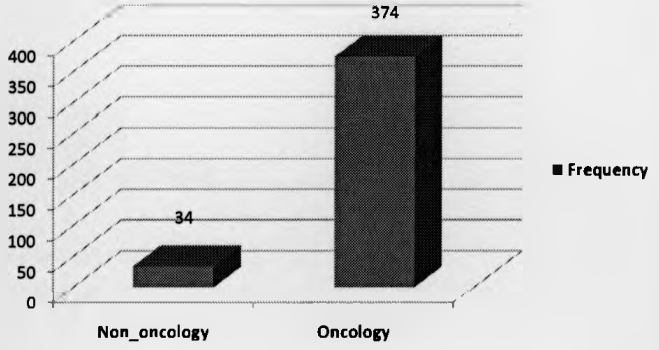
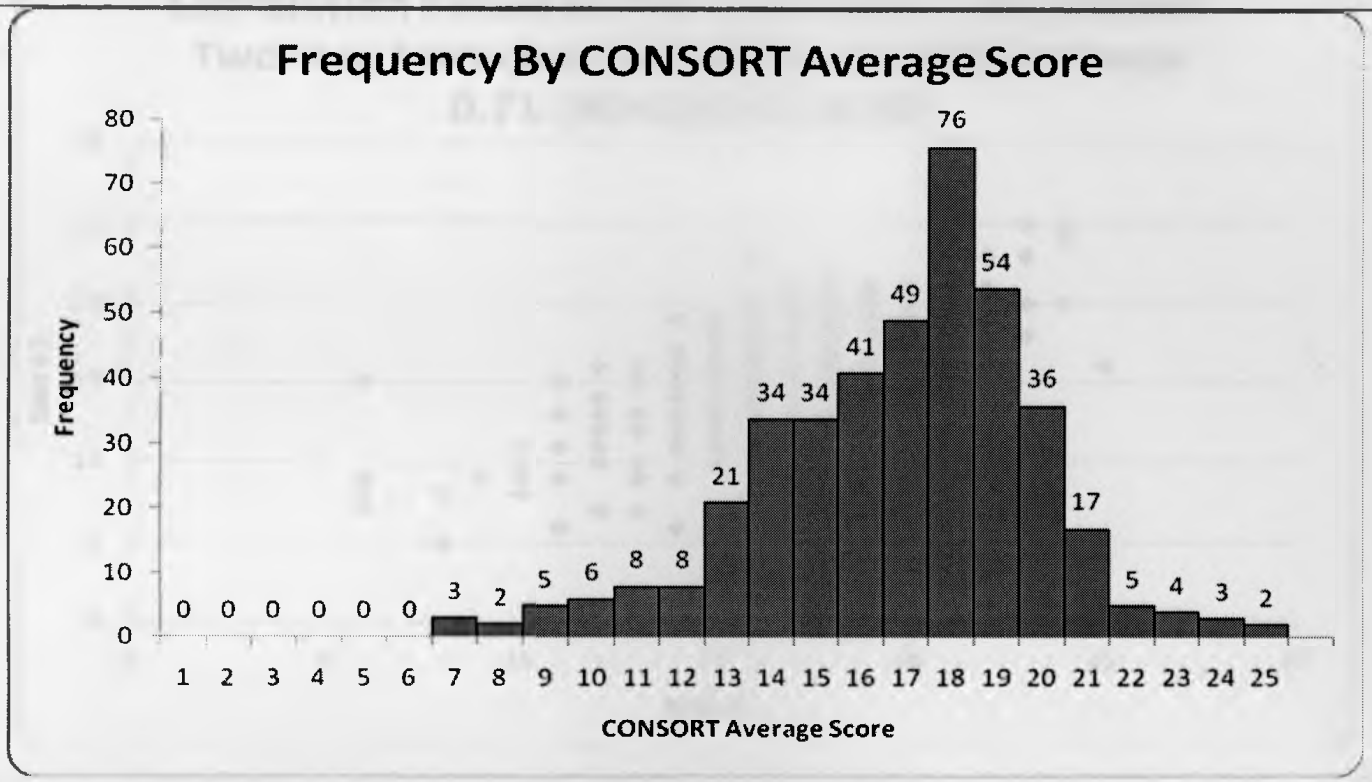


Table 4: Descriptive Analysis Data	
First Reviewer	
Range	6 – 25
Mean	17.1
Standard Deviation	3.1
Second Reviewer	
Range	5 – 25
Mean	16.1
Standard Deviation	3.4
CONSORT Average Scores	
Range	6.5 – 24.5
Mean	16.6
Standard Deviation	3.0
CONSORT Difference of Scores	
Interquartile Range	1 – 3
Median	2
Interquartile Deviation	1
Basic Descriptors of the CONSORT Score data from the 408 RCTs analyzed	

Figure 8: Distributions of CONSORT Average Scores CONSORT Difference of Scores; data from 408 RCTs.



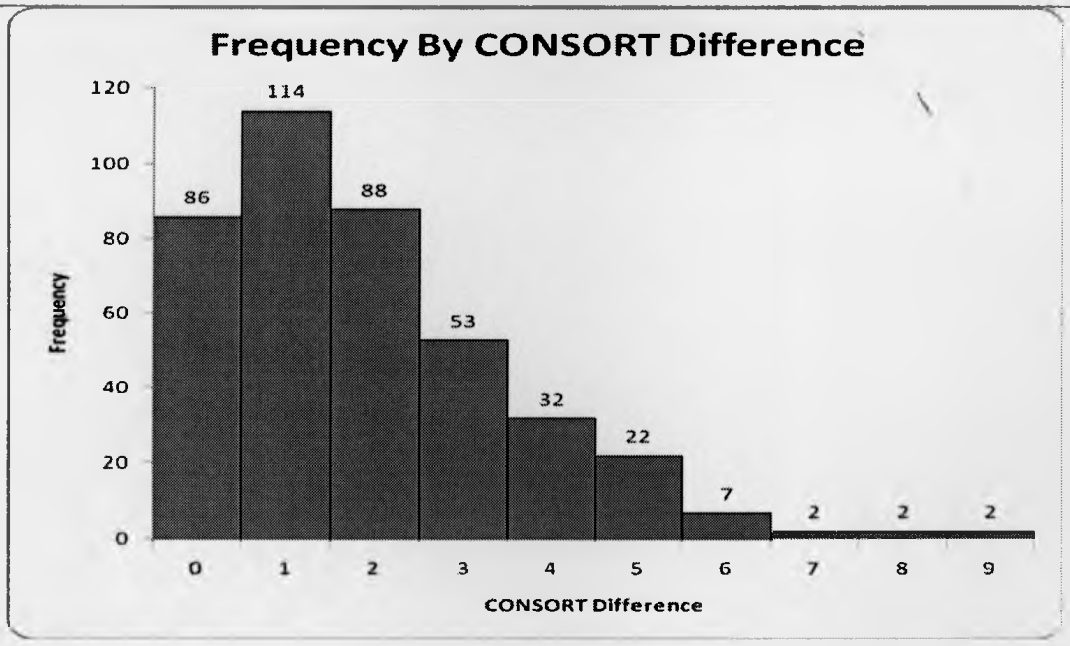


Figure 9: Scatter plot of the two-way missed intraclass correlation coefficients (ICC) between the CONSORT Scores generated by the two reviewers using data from 408 RCTs.

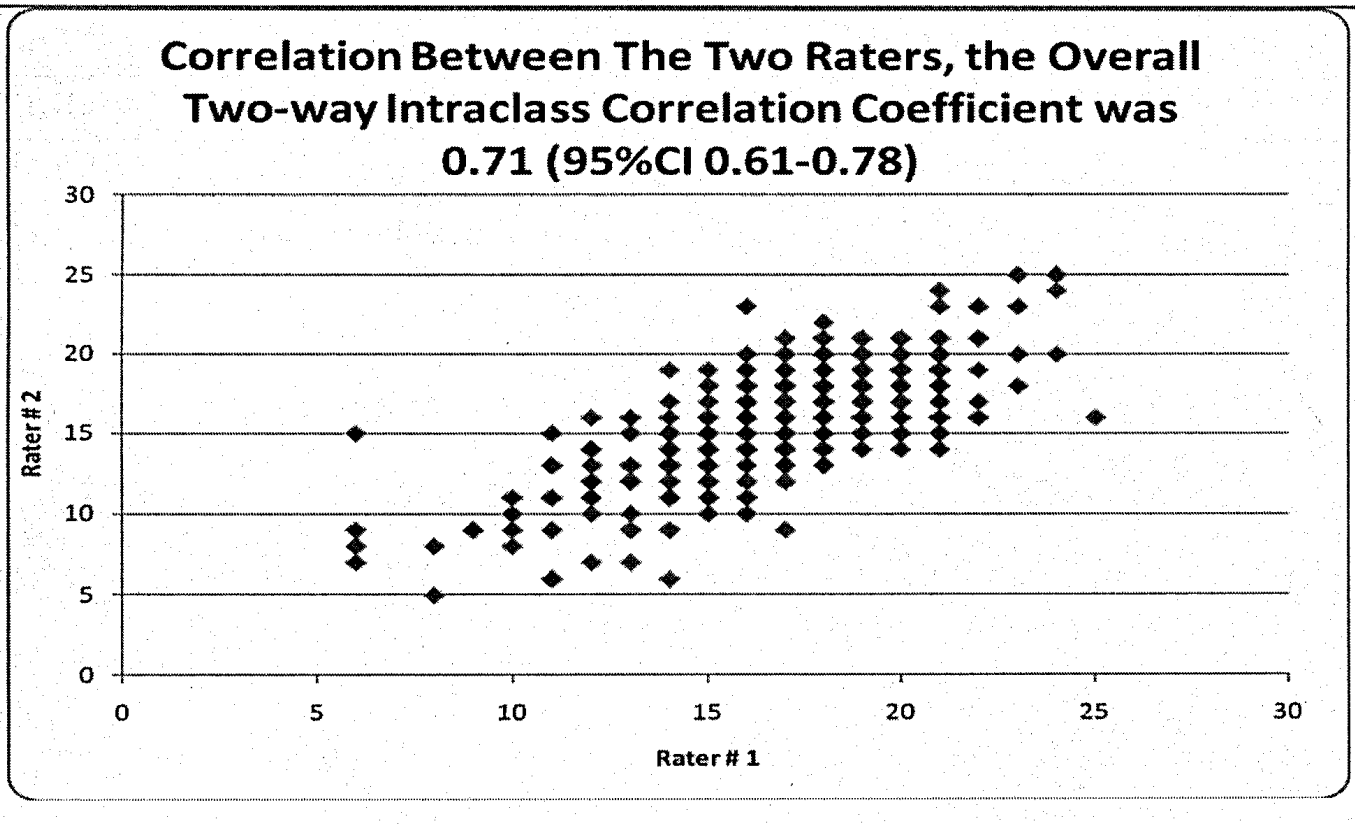


Table 5: Reliability Analysis Intra-rater correlation Coefficient (One-way, two-way Random, and two way Mixed), Data from 408 RCTs.

Form	Model	ICC (95% C.I.)
ICC(1,1)	One Way	0.70 (0.65, 0.75)
ICC(2,1)	Two Way Random (Raters Random)	0.71 (0.61, 0.78)
ICC(3,1)	Two Way Mixed (Raters Fixed)	0.74 (0.69, 0.78)

Table 6: Simple Kappa and Percent Agreement For Each of the CONSORT Checklist Items Using Data from 408 RCTs.		
Checklist Item	**Kappa Statistics**	**Percent Agreement**
Randomization (1a)	0.93	96.32
Design summary (1b)	0.88	96.57
Background (2a)	0.14 *	87.50
Objectives (2b)	0.16 *	87.75
Design (3a)	0.66 *	90.67
Participants Eligibility (4a)	0.30 *	97.79
Settings and Locations (4b)	0.55 *	87.25
Interventions (5)	0.37 *	94.12
Primary and Secondary (6a)	0.55 *	80.88
Sample Size (7a)	0.56 *	85.30
Sequence Generation (8a)	0.49 *	82.60
Sequence Generation (8b)	0.59 *	84.32
Allocation mechanism (9)	0.39 *	71.32 ^
Implementation (10)	0.10 **	76.71 ^
Blinding (11a)	0.52 *	82.00
Statistical Methods (12a)	0.56 *	96.80
Flow for Patients (13a)	0.24 **	71.32 ^
Flow for Loss to Follow up (13b)	0.27 **	61.03 ^
Recruitment Dates (14a)	0.88	95.10
Baseline data (15)	0.28 *	96.57
Number of patients An (16)	0.05 *	58.82 ^
Primary and Secondary (17a)	0.05 *	51.96 ^
Harms (19)	0.40 *	88.73

Limitations (20)	0.39 *	70.59 ^
Generalizability (21)	0.04 **	30.89 ^
Interpretation (22)	0.03 **	92.65
Registration (23)	0.81	97.30
Protocol (24)	0.18 **	94.37
Funding (25)	0.78 *	90.20

Kappa Statistic and Percent Agreement for each of the checklist items

*** Kappa Statistic < 0.80 Red < 0.30**

**** Kappa Statistic < 0.30**

^ Percent Agreement < 80%.

Coefficient interpretation (>.80 Almost perfect, .61 - .80 Substantial, .41 - .60 Moderate, .21 - .40 Fair, 0 - .20 Slight, < 0 Poor) Guidelines for interpretation of reliability. (After Landis and Koch, 1977)

8.4 Main Results

The result of the main effect model analysis of the 408 articles was as follows. Recent year of publication, increasing author number, and higher impact factor were associated with higher CONSORT average score ($p<0.0001$). Recent year of publication was the only factor associated with a decrease in the CONSORT difference score. Two points support the precision of the model. First, the normality of the distribution of CONSORT Average Scores that is evident in the bell-shape distribution shown in Figure (8). Second, the main effect model contains more than 10 observations / variable "Models with fewer than 10 observations/variable, require greater assurance that random errors are normally distributed, i.e. can then not rely on central limit theorem."[97]

The main effects model disproved the second of the two conditions for classical criteria of confounding "a variable is a confounder if it is associated with exposure and causally related to the outcome"[98] by showing that there was no association between each of the potential confounders (Patient Number, Oncology vs. Non-oncology, Cooperative Group, Intervention, Cancer Type, and Trial Site) and the outcome. Disproving this condition was sufficient to rule out the confounding effect in our data.

8.4.1 CONSORT Average Score

Recent year of publication: Results from Table (7) demonstrate a dose response relationship in which later publication year reflects an increase in reporting quality. There was an average increase of 3.1 CONSORT score points comparing an RCT published between 2002-2010 to an RCT published between 1992-1996, and an increase of 1.8 CONSORT score points comparing an RCT published between 2002-2010 to an RCT published between 1997-2001.

Author Number: higher author number was associated with a higher CONSORT Average score ($p<0.0001$). There was an increase of 1 point in the CONSORT Score in published RCTs for an increase of about 7 in the number of authors, Table (7).

Impact Factor: higher impact factor was associated with a higher CONSORT average score ($p<0.0001$). RCTs published in journals of a high Impact Factor have higher CONSORT Average score (1.46 point higher) compared to RCTs published in journals with low impact factor, Table (7)

8.4.2. CONSORT Difference Score

Recent year of publication: This predictor was the only factor associated with a decrease in the CONSORT difference score (This difference reflects disagreement between the two raters, and is considered an estimate of the reliability of the checklist). For example, there was a decrease in CONSORT difference score comparing an RCT published between 2002-2010 to an RCT published between 1992-1996 ($p=0.0085$).

Table (7) presents the results of the main effect model of the association between predictors on one hand and CONSORT average score (reporting quality) and CONSORT score difference (predictors of variability) on the other hand. Table (8) reports category boundaries when continuous variables (Publication Year and Impact Factor) were categorized.

8.5 Other analyses:

Four hundred and eight articles were utilized for calculation of the correlation between CONSORT average score and Impact Factor. A weak Pearson's correlation was seen 0.34. From looking at the entire plot of this correlation Figure (10), we could observe a stronger correlation between these two variables for articles with an impact factor of less than 32.

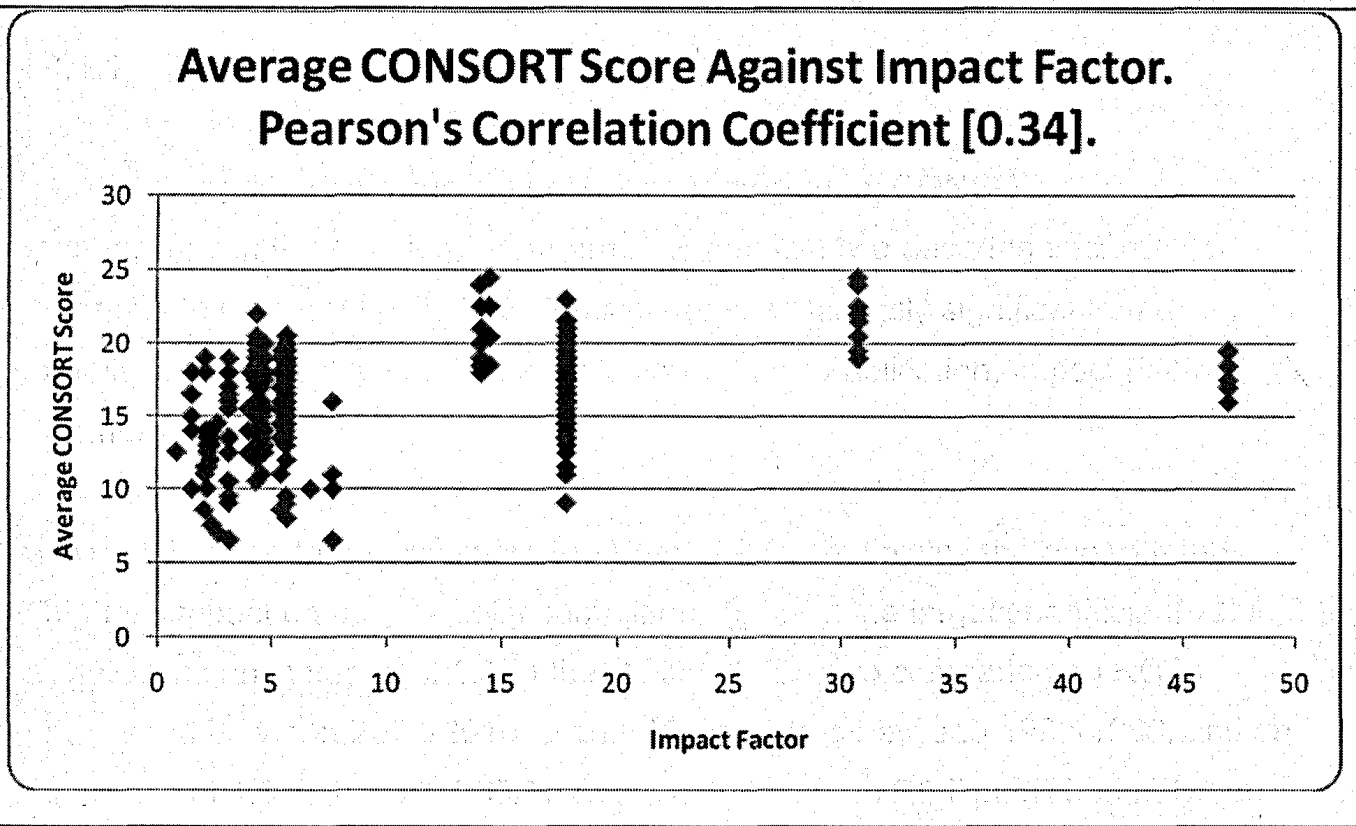
Table 7: Factors Affecting CONSORT Average Score (Reporting Quality), and CONSORT Score Difference (Predictors of Variability) using data from 408 RCTs.

Variable		CONSORT Average		CONSORT Difference	
		Estimate	p-value	Estimate	p-value
Intervention*	Radiation vs. Multiple^	1.54	0.0481	-0.48	0.3694
	Chemo vs. Multiple^	0.63	0.1139	-0.28	0.3020
	Surgical vs. Multiple^	1.21	0.6306	1.91	0.2697
Year of Publication**	(1992-1996) vs. (2002-2010)^	-3.10	**<.0001**	0.70	**0.0085**
	(1997-2001) vs. (2002-2010)^	-1.82	**<.0001**	0.07	0.7491
Trial Site (Single vs. multiple^)		-0.71	0.1510	-0.07	0.8405
Cooperative Group (Non-Cooperative vs. Cooperative^)		-0.31	0.2043	-0.08	0.6186
Journal Type (Non-oncology vs. Oncology^)		0.46	0.3406	0.02	0.9636
No. of Authors		0.15	**<.0001**	-0.03	0.1899
No. of Patients		0.0001	0.6602	-0.0002	0.1670
Impact Factor¥ (High vs. low^)		1.46	**<.0001**	-0.25	0.1661

* The variable Intervention is categorized into four groups: Radiation, Chemo, Surgical, and Multiple Therapy (Any combination of Radiation, Chemo, and Surgical).
** The variable Year of Publication is categorized into three groups: Pre-consort (1992-1996), between the release of CONSORT and its first revision (1997-2001), and between the release of the first CONSORT revision and its second revision (2002-2010).
^ Reference group
¥ The variable Impact Factor is an Ordinal variable that is dichotomized into two groups: Low: ≤ 7.667, and High: ≥ 14.069. Low is the reference group. There was no journal that has an impact factor between these two values

Table 8: Category Boundaries of Publication Year and Impact Factor using data from 408 RCTs.	
Publication Year	
Category 1	(1992-1996)
Category 2	(1997-2001)
Category 3	(2002-2010)
Impact Factor	
Category 1	Low: ≤ 7.667
Category 2	High: ≥ 14.069

Figure 10: The variable Impact Factor is categorized into two groups: Low ≤ 7.667, and High ≥ 14.069. Low is the reference group. There was no journal that has an impact factor between these two values



9.0 Discussion

9.1 Discussion

9.1.1 Predictors of CONSORT Checklist Compliance

Our primary aim was to determine predictors of CONSORT Checklist compliance in the oncology literature over the past two decades and the magnitude of their effects. We identified three statistically significant predictors of quality in the oncology literature including Year of Publication, Impact Factor, and Author Number.

Of these three predictors the Year of Publication was the one with the highest impact on the checklist compliance (i.e with the largest coefficient value). There was an increase of 3.1 in the CONSORT Score comparing an RCT published between 2002-2010 to an RCT published between 1992-1996, and an increase of 1.8 in the CONSORT Score comparing an RCT published between 2002-2010 to an RCT published between 1997-2001.

This result might be explained by several factors that have changed over the two decades. With the increase in the number of researchers trying to publish, publishing became more competitive. Journal editors have raised the publication standards and the editorial instructions have become clearer and stricter. Also the process of peer review became more regulated. The number of peer reviewers increased from one peer reviewer to three or more, and strict rules were put in place to minimize potential biases in the process such as blinding the peer reviewers to the names of authors and institutions.

Another factor that might have increased the quality of reporting is the existence of cooperative group clinical trials. Cooperative groups require internal

peer-review before sending the article for the external peer-review, which may have significant down-stream effects of improved reporting.

Another possible factor is the advancement in technology. This advancement has a clear impact on the way research is conducted. In recent years many research tools became more available such as more robust statistical software, clearer and easier to construct graphs and databases, improvements in publishing and word-processing software, and faster internet access.

Advancement in technology did not only affect the way research is conducted, but also the way it is presented and published. The open access movement enabled the publication of many RCTs that would not be published if online access were not available. Although more access might be viewed as leading to less competitiveness, traditional journals have been striving to distinguish their publications by increasing their quality over time.

The predictor with the second largest coefficient was the journal impact factor. RCTs published in journals of a high impact factor have a higher CONSORT average score (1.46 point higher) compared to RCTs published in journals of low impact factor.

It would have been counterintuitive if the impact factor was not associated with better reporting quality. High impact factor journals are more competitive to successfully publish medical research work. They usually require more strict publication instructions and peer review process. Therefore, RCTs published in these journals have better methodological quality compared to ones published in low impact factor journals. This result is congruent with previous literature,[80,82] in that good methodological quality is associated with good reporting quality in RCT.

The predictor with the smallest coefficient was the total author number. There was an increase of 1 point in the CONSORT Score in published RCTs for an increase of about 7 in the number of authors. Intuitively, larger research teams have the advantages of more feedback and internal reviews. Members from different backgrounds bring to the equation different experiences and perspectives, and thus potentially broaden a team's base of knowledge. For example, having a statistician on the team ensures a more through insight in the statistical part of the reporting.

Although this association is statistically significant, it does not seem to present practical importance. If we considered this finding to be equivalent to a strong linear association, author numbers would go up substantially. For example, if an RCT has 6 authors with a reporting quality of 16 points, and we need to increase the quality of a report to 21 points (Maximum of 25 points), keeping all other variables the same, the number of authors of an article should be increased to 41 authors. This number is not financially feasible, nor sensibly practical.

9.1.2 Reliability of the CONSORT Checklist Items

Kappa agreement for each individual CONSORT checklist item ranged from (0.02-0.92). Percent agreement for each individual CONSORT checklist item ranged from (30.9 -97.8%). Our data showed that a few of the checklist items have poor Kappa values yet high levels of agreement. This counterintuitive relationship was presented and explained by Feinstein in a paper titled "High agreement but low kappa: The problems of two paradoxes"[99] published in the Journal of Clinical Epidemiology. "In a fourfold table showing binary agreement of two observers, the observed proportion of agreement can be paradoxically altered by the chance-corrected ratio that creates kappa as an index of concordance. In one paradox, a high value of proportion of agreement can be drastically lowered by a substantial imbalance in the table's marginal totals either

vertically or horizontally." In other words, for rare finding, very low values of kappa may not necessarily indicate low rates of overall agreement.

CONSORT checklist items [Allocation mechanism (Item # 9), Implementation (Item # 10) Flow for Patients (Item # 13a) Flow for Loss to Follow up (Item # 13b) Generalizability (Item # 21) Interpretation (Item # 22)] were the least clear to interpret by the two reviewers Figure (4). These items were identified based on a combination of the values of the Kappa statistic and the percentage agreement for each item, and then on a discussion between the two reviewers after data collection and analysis.

Previous works in the field have shown similar results. For example Moberg-Mogren *et al* found that the kappa statistics computed on individual items ranged from high levels of agreement for some items to ones that fell below the moderate level of agreement.[86, 93]

Recent year of publication was the only factor associated with an increase in reliability of the CONSORT checklist items (decrease in the CONSORT difference score which is considered an estimate of variability $p<0.0001$). There was a decrease in CONSORT Score difference comparing an RCT published between 2002-2010 to an RCT published between 1992-1996 ($p<0.0001$). This result could be explained by the same possible reasons mentioned above to explain the increase of the reporting quality in recent studies.

The variability in the level of agreement for some items (a proxy for item reliability) is likely multi-factorial. One possible explanation of the variability in the levels of agreement for some items is that the fact that the CONSORT statement was adopted by many journals does not mean that these journals literally use the CONSORT checklist as their guidelines; rather they integrate the statement's recommendations in their own guidelines. For example, although the Journal of General Internal Medicine (JGIM) adopted the CONSORT Statement, authors or

peer reviewers interacting with the journal should use the JGIM Review Guideline as a guideline. Variability in the integration of the statement's recommendations may result in lowering the reliability of some of the checklist items. A second possible explanation is that some of the checklist items might be more applicable for certain specialties or procedures. For example, allocation concealment might be easier to explain and report in drug trials than in surgical trials. A third possible explanation is that the wording of some of the checklist items and the instructions on how to use them are indeed unclear. Also, a baseline level of variation is expected with any interpretative activity including this audit of the oncology literature.

In all cases, collaborative efforts to improve the wording of the checklist items and the working definitions related to these items are sought. Such collaboration would be more effective if all stakeholders were involved in the process, whether be journal auditors, authors, peer reviewers, librarians, and research consumers etc. Further incorporation of the CONSORT statement in graduate and post-graduate studies may result in a more universal understanding of its definitions. Finally, further research in this area may help in improving the reliability of the CONSORT items too.

9.1.3 RCT Reporting Quality over the Past 20 Years

The results that have been noted by previous research describing RCT reporting [86, 89] has been confirmed herein. Although improvements in RCT reporting have been observed over time in the cancer literature, the overall quality of reporting remains suboptimal (Mean average CONSORT score was 16.6 [SD 3, max 25]). By looking at these numbers from a percentage point of the view we can safely say that 50% of the published literature has a reporting quality of 66.4% or less (16.6/25 x 100), and 85% (one standard deviation above

the mean) has a reporting quality of 78.4% or less (19.6/25 x 100) based on the CONSORT statements consensus definition of reporting standards.

By looking at these percentages in the context of the CONSORT statement, we can see that the reporting quality of published RCTs in oncology RCT-based research is less than optimal. This may be the result of several obstacles. Identifying and overcoming these obstacles may help health-care practitioners and research consumers to make informed inferences about the validity of the studies upon which they base their decisions. Further research in this area is warranted.

In the 515 reports included in the full article review, 107 had unclear abstracts. These abstracts included information about the study design that presented the report as an RCT while the full article review revealed that the report was not. Many research consumers depend only on the abstract to obtain the research results and level of evidence. It is therefore alarming that a fifth (107/515) of the literature might provide inaccurate information in this regard. Other studies found similar or less optimal results. For example, the initial electronic search and abstract review in a similar study identified 274 cancer RCTs, while their full article review revealed that only 104 of these articles were indeed RCTs.[89] In another example, the initial electronic search and abstract review identified 482 clinical trials, while their full article review revealed that only 193 were indeed RCTs.[94]

Confounding: Unlike other studies, our study did not find any significant confounding effects of the study potential confounders (as described in the methods section) on the association between the study predictors and the reporting quality.

9.2. Generalizability

This study does have several factors affecting its generalizability. The study sample was derived from a single database, PubMed, this increases the reliability (internal validity) of the study results, yet reduces the generalizability (external validity) of these results when applying the same question to RCTs published in other databases. From searching the literature we found two studies by Plint *et al*, and Farrokhyar *et al* with similar results to our study. Both studies obtained their data from electronic searches of (MEDLINE, EMBASE, and Cochrane CENTRAL) and of (MEDLINE, the Cochrane Library, CINAHL, HealthSTAR and EMBASE) respectively, and both used the CONSORT checklist to score the quality.[87, 93]

The study investigated the four most common cancer types. These types are responsible for around 60% of all cancer incidences and cancer caused death. Since research studying the other cancers is conducted in almost similar pattern (same countries, institutions, and journals), it might be possible, with caution, to generalize the study results to the research dealing with the other types of cancer. The study sample included RCTs published only in English language journals. We cannot infer whether it is safe to generalize the study results to RCTs published in other languages.

The article review was done by qualified reviewers who have different backgrounds (Oncology, Epidemiology), which increased the generalizability of the study results. The study sample is large and includes RCTs conducted in many different institutions, groups, and countries. This variety also ensures good generalizability of the study results. This study investigated RCTs that were published in journals that had published more than four articles in the past 20 years. Therefore, the study results may not be applicable for RCTs published in journals that publish cancer research infrequently.

9.3 Future Research

There are several directions for future study. Given the observed kappa agreement heterogeneity, further work in the assessment of the reliability of individual CONSORT items is warranted. One potential method to conduct this assessment is by using the checklist to score a sample of RCTs by multiple reviewers. The fact that our data was extracted from published RCTs may be a limitation to the strength of evidence obtained in our study. Performing a study with prospective database might provide a higher level of evidence with regards to addressing questions related to reporting quality.

Studies have shown that several scales were used to evaluate the reporting quality of RCTs, but most of these scales have not been adequately developed, nor have been adequately validated. Our study provides complementary results to those from Moher *et al*[53], and highlights the need for a more standardized method to assess the reporting of RCTs.

Furthermore, It might be a good idea that the reporting guidelines recommend that publishing journals require authors to use the standardized scale to generate a score for their RCT as a mandatory step in the publication process. This score could be included in the article index. Such score would give the reader clear information on the article's reporting quality before reading it. This in turn might be a factor that draws the authors' attention to the importance of reporting quality and encourages them to strive for excellence in this regard.

Future scale development is likely to be most favourable if items common to all trials are assessed, if the scale is straightforward to use, and if it is developed with sufficient reliability. General consideration for the development of a scale may include: definition of the quality assessed by the scale, definition of the sample to be scored (published and/or unpublished RCTs in a specific field or

in all medical fields), definition of research consumers that are going to use the scale (same or different backgrounds), definition of application method (The use of scoring sheet or training the reviewers), and open or blind-trial scoring.

9.4 Limitations

This study has a number of limitations and sources of bias. For example, multiple testing can result in inclusion of "noise variables", and different "best" models may be obtained using other algorithms (e.g. forward stepwise). To minimize the bias resulting from these two points, our analysis did not include unnecessary tests; we only performed the necessary analyses required to achieve credible results and to rule out confounding. We also used SAS software that properly accounts for the dummy variables in our model. More importantly, a combination of epidemiological insight, the size of estimated regression coefficients, the width of the confidence intervals, and the size of P-values rather than relying exclusively on hypothesis tests drove the main effect modeling.

Another source of bias in this study was the fact that the two reviewers could not be blinded to the journals' names or authors. There might be a theoretical incline to give high impact journals a higher score because of their reputation. If this bias takes place in this study, it will similarly affect older and newer publications resulting in no effect on the study's conclusion regarding the increase in reporting quality in recent publications.

Although we have included in our investigation all published RCTs, and although there was no missing data, there is no clear way to evaluate how much of these reports were "improved" by the journal editors and peer-reviewers after submission to the journal. If this "improvement" in fact exists, the variation from journal to journal depending on their reviewers and editors remains unknown.

In the recent years, the Open Access Movement (the ability to publish in online journals) has provided a new medium to publish, which increases the chances of publishing RCTs.[100] Therefore, the more recent an RCT is, the greater its chances to be published. We have not included those RCTs that were rejected for publication. The exclusion of these RCTs might have biased the p-values downward (especially when testing the association of Year of Publication with CONSORT Average Scores and Difference of Scores). Therefore, results reported as statistically significant may in fact not be significant. The fact that unpublished RCTs do not usually influence clinical practice justifies their exclusion of the study sample.

Another limitation of our study was that we assessed publications only in English and only involving Lung, Breast, Prostate and Colorectal cancer; the study excluded other malignant diseases. Also, we did not have access to the original study protocols nor did we interview the investigators who had undertaken the studies for additional information.

Another limitation of our study was the fact that several trials included in our sample did not specify age range for their inclusion/exclusion criteria. Thus, our inclusion/exclusion criteria did not specify a range too.

We did not examine our data to see whether some of the subjects in our sample were published by the same author (research team). Reporting quality of RCTs published by the same author are likely to be more similar than reporting quality of RCTs published by different authors. Hence the assumption of statistical independence, an assumption that is central for the validity of the hypothesis testing, could be violated. The violation of this assumption produces p-values that are too small. We did not factor in this effect in our analysis because of its small magnitude, since the number of RCTs published by the same author in the sample is likely very small relative to the large sample size.

9.5 Conclusion

The findings of this study are summarized in the following points:

This study demonstrates a dose-response relationship in which more recent publication year reflects an increase in reporting quality. There was an increase of 3.1 CONSORT score points comparing an RCT published between 2002-2010 to an RCT published between 1992-1996, and an increase of 3.1 CONSORT score points comparing an RCT published between 2002-2010 to an RCT published between 1997-2001. The higher the impact factor, the higher the reporting quality was (higher CONSORT average score $p<0.0001$). RCTs published in journals of a high impact factor have higher COSORT Average score (1.46 point higher) compared to RCTs published in Journals of low Impact Factor. The higher the author number, the higher the reporting quality was (higher CONSORT Average score $p<0.0001$). There was an increase of 1 point in the CONSORT Score in published RCTs for an increase of about 7 in the number of authors.

An overall two-way intraclass correlation coefficient of 0.71 (95%CI 0.61-0.78) for comparison of overall CONSORT score between the two reviewers. Kappa agreement for each individual CONSORT checklist item ranged from (0.02-0.92). Percent agreement for each individual CONSORT checklist item ranged from (30.9 -97.8%). CONSORT Checklist items [Allocation mechanism (Item # 9), Implementation (Item # 10) Flow for Patients (Item # 13a) Flow for Loss to Follow up (Item # 13b) Generalizability (Item # 21) Interpretation (Item # 22)] were the least clear to interpret by the two reviewers. Table (6) presents simple Kappa statistic and percent agreement for each of the CONSORT Checklist items.

Recent year of publication was the only factor associated with an increase in reliability of the CONSORT checklist items (decrease in the CONSORT difference score which is considered predictor of variability $p<0.0001$). There was a decrease in CONSORT Score difference comparing an RCT published between 2002-2010 to an RCT published between 1992-1996 ($p<0.0001$). Although improvements in RCT reporting have been observed over time in the cancer literature, the quality of reporting remains suboptimal (Mean average CONSORT score was 16.6 [SD 3, max 25]).

Appendix I

SAS Programming – Kappas

```
options ls=120 ps=65;

%let root=S:\LRCP\CRU STATS\consort audit;

libname save "&root\save";
libname library "&root\save";

ods rtf file="&root\kappas.rtf";

data consort;
  set save.consort;

proc freq;
  table rand1*rand2
        desi1*desi2
        .  back1*back2
           obje1*obje2
           desib1*desib2
           part1*part2
           sett1*sett2
           inte1*inte2
           prim1*prim2
           samp1*samp2
           sequ1*sequ2
           gene1*gene2
           allo1*allo2
           impl1*impl2
           blin1*blin2
           stat1*stat2
           flow1*flow2
           flowb1*flowb2
           recr1*recr2
           base1*base2
           numb1*numb2
           primb1*primb2
           harm1*harm2
           limi1*limi2
           geneb1*geneb2
           inteb1*inteb2
           regi1*regi2
           prot1*prot2
           fund1*fund2/agree;
*  exact agree;
RUN;
```

Appendix II

SAS Programming – Descriptive Analysis and Main Effects Model

```
proc freq;
      tables journal cancer year yeargroup intervention trial_site
multipleco primaryco coop journalonc consort1 consort2 consortav
consortdiff IFgroupA;
      run;

proc univariate;
      var authorno patientno IF2009 consort1 consort2 consortav
consortdiff;
      run;

proc glm;
class Cancer Intervention Trial_Site Coop Journalonc Yeargroup;
model consortav = Cancer Intervention Trial_Site Coop Journalonc
Yeargroup authorno patientno IFgroupA / solution;
run;

proc glm;
class Cancer Intervention Trial_Site Coop Journalonc Yeargroup;
model consortdiff = Cancer Intervention Trial_Site Coop Journalonc
Yeargroup authorno patientno IFgroupA / solution;
run;


Cancer_lung=0; Cancer_breast=0;Cancer_prostate=0;
if Cancer=1 then Cancer_lung=1;
if Cancer=2 then Cancer_breast=1;
if Cancer=3 then Cancer_prostate=1;
Intervention_radiation=0;Intervention_chemo=0;Intervention_surgical=
0;
if Intervention=1 then Intervention_radiation=1;
if Intervention=2 then Intervention_chemo=1;
if Intervention=3 then Intervention_surgical=1;
Trial_Site=0;
if Trial_Site=1 then Trial_site=1;
Coop=0;
if coop=1 then coop=1;
Journalonc=0;
if Journalonc=1 then Journalonc=1;
Yeargroupq=(Yeargroup le 1996) + 2*(Yeargroup ge 1997 and lwt le
2001) + 3*(Yeargroup ge 2002);
Yeargroup1=0;Yeargroup2=0;
if (Yeargroupq=1) then Yeargroup1=1;
if (Yeargroupq=2) then Yeargroup2=1;



proc reg;
model consortav=

{Cancer_lung Cancer_breast Cancer_prostate}
{Intervention_radiation Intervention_chemo Intervention_surgical}
{Yeargroup1 Yeargroup2}
```

```
Trial_Site Coop Journalonc
authorno patientno IFgroupA /
selection=backward

/* Creating Revised Predictors and Dummy Variables */

ftvb=0;
if ftv ge 1 then ftvb=1;
race_white=0;race_other=0;
if race=1 then race_white=1;
if race=3 then race_other=1;
lwtq=(lwt le 109) + 2*(lwt ge 110 and lwt le 120)
+ 3*(lwt ge 121 and lwt le 138)
+ 4*(lwt ge 139);
lwt2=0;lwt3=0;lwt4=0;
if (lwtq=2) then lwt2=1;
if (lwtq=3) then lwt3=1;
if (lwtq=4) then lwt4=1;
ptlb=0;
if ptl ge 1 then ptlb=1;


/* Main Effects Model - FTV binary */
proc reg;
model bwt=smoke {age age2} ftvb
{race_white race_other}
{lwt2 lwt3 lwt4}
ht ptlb ui /
selection=backward
groupnames="Mother's smoking status"
"Mother's age"
"Num. doc visits"
"Mother's race"
"Mother's pre-pregnancy weight"
"Mother's history of hypertension"
"Number of premature labors"
"Uterine irritability"
include=3;
```

Appendix III

SAS Programming – Backward Elimination

```
proc reg;
model consortav = Cancer Intervention Trial_Site Coop Journalonc
Yeargroup authorno patientno IFgroupA /
selection=backward;
run;


proc reg;
model consortdiff = Cancer Intervention Trial_Site Coop Journalonc
Yeargroup authorno patientno IFgroupA /
selection=backward;
run;
```

Appendix XII

Permission Letters

Citation: Material appearing in this publication may be reproduced or copied without permission; however, the following citation must be used to indicate the source:

Canadian Cancer Society's Steering Committee on Cancer Statistics. Canadian Cancer

Statistics 2011. Toronto, ON: Canadian Cancer Society; 2011.

May 2011

ISSN 0835-2976

From: Canadian Cancer Society <hamiltoncis@cis.cancer.ca>
To: iarra2@uwo.ca
Sent: Mon, June 27, 2011 3:08:25 PM
Subject: Permission to use figures as reference (Date 06/27/2011 User 239)

Dear Ian,

Thank you for contacting the Canadian Cancer Society's Cancer Information Service. Your email has been forwarded to us by our National office.

In the Canadian Cancer Statistics 2010 booklet, if you go to the first page titled Steering Committee Members, towards the bottom of the page you will see a section titled Citation. There you will read how the exact wording should be when you cite information from the statistics book.

If you have any questions or comments, don't hesitate to contact us again, and good luck with your thesis.

Sincerely,

Lynn, Cancer Information Specialist

Citing the CONSORT Statement from the website (for authors)

When referring to the CONSORT Statement, we recommend using journal article citations rather than referring to the CONSORT Statement website. If you are not already using a journal article citation, please cite one of the following original publications of CONSORT 2010:.

CONSORT 2010 Statement

- Schulz KF, Altman DG, Moher D, for the CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. Ann Int Med 2010;152. Epub 24 March.
- Schulz KF, Altman DG, Moher D, for the CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. BMC Medicine 2010, 8:18. (24 March 2010)
- Schulz KF, Altman DG, Moher D, for the CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. BMJ 2010;340:c332.
- Schulz KF, Altman DG, Moher D, for the CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. J Clin Epi 2010;
- Schulz KF, Altman DG, Moher D, for the CONSORT Group. CONSORT 2010
  statement: updated guidelines for reporting parallel group randomized trials. Obstet Gynecol 2010;115(5):[pages TBD].
- Schulz KF, Altman DG, Moher D; for the CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomized trials. Open Med 2010;4(1):60-68.
- Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 Statement: Updated Guidelines for Reporting Parallel Group Randomised Trials. PLoS Med 2010;7(3): e1000251. doi:10.1371/journal.pmed.1000251
- Schulz KF, Altman DG, Moher D, for the CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. Trials 2010, 11:32. (24 March 2010)

CONSORT 2010 Explanation and Elaboration

Since the CONSORT Statement should be read in conjunction with the CONSORT Explanation and Elaboration document, you should also be using and citing:

- Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, Elbourne D, Egger M, Altman DG, for the CONSORT Group. CONSORT 2010 Explanation and Elaboration: updated guidelines for reporting parallel group randomised trial. BMJ 2010;340:c869.
- Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, Elbourne D, Egger M, Altman DG, for the CONSORT Group. CONSORT 2010 Explanation and Elaboration: updated guidelines for reporting parallel group randomised trial. J Clin Epi 2010;

Using the CONSORT Statement

However, because the guidelines represent a consensus agreed through successive drafts by the CONSORT Group, they should not be edited or modified in any way, although it is acceptable to publish portions (e.g., the summary).

Page last edited: 16 May 2011

References:

1. http://www.cancer.ca/Canada-wide/About%20cancer/Cancer%20statistics/Stats%20at%20a%20glance/General%20cancer%20stats.aspx?sc_lang=en

2. Johnston SC, Rootenberg JD, Katrak S, Smith WS, Elkins JS. Effect of a US national institutes of health programme of clinical trials on public health and costs. Lancet 2006 Apr 22;367(9519):1319-27.

3. Canadian Cancer Society's Steering Committee on Cancer Statistics. Canadian Cancer. Statistics 2011. Toronto, ON: Canadian Cancer Society; 2011. May 2011, ISSN 0835-2976

4. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA 1995 Feb 1;273(5):408-12.

5. Moher D, Pham B, Jones A, Cook DJ, Jadad AR, Moher M, Tugwell P, Klassen TP. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? Lancet 1998 Aug 22;352(9128):609-13.

6. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: Updated guidelines for reporting parallel group randomised trials. BMJ 2010 Mar 23;340:c332.

7. Thornley B, Adams C. Content and quality of 2000 controlled trials in schizophrenia over 50 years. BMJ 1998 Oct 31;317(7167):1181-4.

8. Dickinson K, Bunn F, Wentz R, Edwards P, Roberts I. Size and quality of randomised controlled trials in head injury: Review of published studies. BMJ 2000 May 13;320(7245):1308-11.

9. Stolberg HO, Norman G, Trop I. Randomized controlled trials. AJR Am J Roentgenol 2004 Dec;183(6):1539-44.

10. Louis PCA. Research into the effects of bloodletting in some inflammatory diseases and on the influence of tartarized antimony and vesication in pneumonitis. Am J Med Sci1836; 18:102 –111

11. STREPTOMYCIN treatment of pulmonary tuberculosis. Br Med J 1948 Oct 30;2(4582):769-82.

12. Louise Forsetlund[a]; Iain Chalmers[b]; Arild Bjørndal[a]. When Was Random Allocation First Used To Generate Comparison Groups In Experiments To Assess The Effects Of Social Interventions? Economics of Innovation and New Technology, Volume 16, Issue 5 July 2007 , pages 371 - 384

13. Meldrum ML. A brief history of the randomized controlled trial. from oranges and lemons to the gold standard. Hematol Oncol Clin North Am 2000 Aug;14(4):745,60, vii.

14. Freedman B. Equipoise and the ethics of clinical research. N Engl J Med 1987 Jul 16;317(3):141-5.

15. Hopewell S, Dutton S, Yu LM, Chan AW, Altman DG. The quality of reports of randomised trials in 2000 and 2006: Comparative study of articles indexed in PubMed. BMJ 2010 Mar 23;340:c723.

16. Edwards SJ, Braunholtz DA, Lilford RJ, Stevens AJ. Ethical issues in the design and conduct of cluster randomised controlled trials. BMJ 1999 May 22;318(7195):1407-9.

17. Newcombe RG, Duff GR. Eyes or patients? traps for the unwary in the statistical analysis of ophthalmological studies. Br J Ophthalmol 1987 Sep;71(9):645-6.

18. Edwards SJ, Lilford RJ, Hewison J. The ethics of randomised controlled trials from the perspectives of patients, the public, and healthcare professionals. BMJ 1998 Oct 31;317(7167):1209-12.

19. Sackett DL. Superiority trials, noninferiority trials, and prisoners of the 2-sided null hypothesis. ACP J Club 2004 Mar-Apr;140(2):A11.

20. Piaggio G, Elbourne DR, Altman DG, Pocock SJ, Evans SJ, CONSORT Group. Reporting of noninferiority and equivalence randomized trials: An extension of the CONSORT statement. JAMA 2006 Mar 8;295(10):1152-60.

21. Treweek S, Zwarenstein M. Making trials matter: Pragmatic and explanatory trials and the problem of applicability. Trials 2009 Jun 3;10:37.

22. Donner A. Design and Analysis of Clinical Trials. 2010 page 24.

23. Zelen M. A new design for randomized clinical trials. N Engl J Med 1979 May 31;300(22):1242-5.

24. Schulz KF, Grimes DA. Generation of allocation sequences in randomised trials: Chance, not choice. Lancet 2002 Feb 9;359(9305):515-9.

25. Jones, Byron; Kenward, Michael G. (2003). Design and Analysis of Cross-Over Trials (Second ed.). London: Chapman and Hall.

26. Torgerson DJ, Roland M. What is zelen's design? BMJ 1998 Feb 21;316(7131):606.

27. Moher D, Hopewell S, Schulz KF, Montori V, Gotzsche PC, Devereaux PJ, Elbourne D, Egger M, Altman DG. CONSORT 2010 explanation and elaboration: Updated guidelines for reporting parallel group randomised trials. BMJ 2010 Mar 23;340:c869.

28. Schulz KF, Grimes DA. Allocation concealment in randomised trials: Defending against deciphering. Lancet 2002 Feb 16;359(9306):614-8.

29. Zwarenstein M, Treweek S, Gagnier JJ, Altman DG, Tunis S, Haynes B, Oxman AD, Moher D, CONSORT group, Pragmatic Trials in Healthcare (Practihc) group. Improving the reporting of pragmatic trials: An extension of the CONSORT statement. BMJ 2008 Nov 11;337:a2390.

30. Pildal J, Chan AW, Hrobjartsson A, Forfang E, Altman DG, Gotzsche PC. Comparison of descriptions of allocation concealment in trial protocols and the published reports: Cohort study. BMJ 2005 May 7;330(7499):1049.

31. Wood L, Egger M, Gluud LL, Schulz KF, Juni P, Altman DG, Gluud C, Martin RM, Wood AJ, Sterne JA. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: Meta-epidemiological study. BMJ 2008 Mar 15;336(7644):601-5.

32. Jadad AR. Randomised controlled trials: a user's guide. London, England: BMJ Books, 1998.

33. Marson AG, Al-Kharusi AM, Alwaidh M, Appleton R, Baker GA, Chadwick DW, Cramp C, Cockerell OC, Cooper PN, Doughty J, Eaton B, Gamble C, Goulding PJ, Howell SJ, Hughes A, Jackson M, Jacoby A, Kellett M, Lawson GR, Leach JP, Nicolaides P, Roberts R, Shackley P, Shen J, Smith DF,

Smith PE, Smith CT, Vanoli A, Williamson PR, SANAD Study group. The SANAD study of effectiveness of valproate, lamotrigine, or topiramate for generalised and unclassifiable epilepsy: An unblinded randomised controlled trial. Lancet 2007 Mar 24;369(9566):1016-26.

34. Chan R, Hemeryck L, O'Regan M, Clancy L, Feely J. Oral versus intravenous antibiotics for community acquired lower respiratory tract infection in a general hospital: Open, randomised controlled trial. BMJ 1995 May 27;310(6991):1360-2.

35. Fukase K, Kato M, Kikuchi S, Inoue K, Uemura N, Okamoto S, Terao S, Amagai K, Hayashi S, Asaka M, Japan Gast Study Group. Effect of eradication of helicobacter pylori on incidence of metachronous gastric carcinoma after endoscopic resection of early gastric cancer: An open-label, randomised controlled trial. Lancet 2008 Aug 2;372(9636):392-7.

36. Devereaux PJ, Manns BJ, Ghali WA, Quan H, Lacchetti C, Montori VM, Bhandari M, Guyatt GH. Physician interpretations and textbook definitions of blinding terminology in randomized controlled trials. JAMA 2001 Apr 18;285(15):2000-3.

37. Haahr MT, Hrobjartsson A. Who is blinded in randomized clinical trials? A study of 200 trials and a survey of authors. Clin Trials 2006;3(4):360-5.

38. Koepsell T., Weiss N. Epidemiology Methods Studying the Occurrence of Illness. Page 323-224.

39. Vittinghoff *et al.*, Regression methods in biostatistics: linear, logistic, survival, and repeated measures models, 2005 page 51.

40. Anderson JL, Pratt CM, Waldo AL, Karagounis LA. Impact of the food and drug administration approval of flecainide and encainide on coronary artery disease mortality: Putting "deadly medicine" to the test. Am J Cardiol 1997 Jan 1;79(1):43-7.

41. Sanson-Fisher RW, Bonevski B, Green LW, D'Este C. Limitations of the randomized controlled trial in evaluating population-based health interventions. Am J Prev Med 2007 Aug;33(2):155-61.

42. Kjaergard LL, Als-Nielsen B. Association between competing interests and authors' conclusions: Epidemiological study of randomised clinical trials published in the BMJ. BMJ 2002 Aug 3;325(7358):249.

43. Lexchin J, Bero LA, Djulbegovic B, Clark O. Pharmaceutical industry sponsorship and research outcome and quality: Systematic review. BMJ 2003 May 31;326(7400):1167-70.

44. Als-Nielsen B, Chen W, Gluud C, Kjaergard LL. Association of funding and conclusions in randomized drug trials: A reflection of treatment effect or adverse events? JAMA 2003 Aug 20;290(7):921-8.

45. Chan AW, Altman DG. Epidemiology and reporting of randomised trials published in PubMed journals. Lancet 2005 Mar 26-Apr 1;365(9465):1159-62.

46. Ross JS, Hill KP, Egilman DS, Krumholz HM. Guest authorship and ghostwriting in publications related to rofecoxib: A case study of industry documents from rofecoxib litigation. JAMA 2008 Apr 16;299(15):1800-12.

47. Ioannidis JP, Haidich AB, Pappa M, Pantazis N, Kokori SI, Tektonidou MG, Contopoulos-Ioannidis DG, Lau J (2001). "Comparison of evidence of treatment effects in randomized and nonrandomized studies". *J Am Med Assoc* 286 (7): 821–30. doi:10.1001/jama.286.7.821.

48. Bentzen SM. Towards evidence based radiation oncology: Improving the design, analysis, and reporting of clinical outcome studies in radiotherapy. Radiother Oncol 1998 Jan;46(1):5-18.

49. Bekelman JE, Yahalom J. Quality of radiotherapy reporting in randomized controlled trials of hodgkin's lymphoma and non-hodgkin's lymphoma: A systematic review. Int J Radiat Oncol Biol Phys 2009 Feb 1;73(2):492-8.

50. Hellman S, Hellman DS. Of mice but not men. problems of the randomized clinical trial. N Engl J Med 1991 May 30;324(22):1585-9.

51. Benson K, Hartz AJ. A comparison of observational studies and randomized, controlled trials. N Engl J Med 2000 Jun 22;342(25):1878-86.

52. Einhorn LH. Curing metastatic testicular cancer. Proc Natl Acad Sci U S A 2002 Apr 2;99(7):4592-5.

53. Moher D, Jadad AR, Nichol G, Penman M, Tugwell P, Walsh S. Assessing the quality of randomized controlled trials: An annotated bibliography of scales and checklists. Control Clin Trials 1995 Feb;16(1):62-73.

54. Huwiler-Muntener K, Juni P, Junker C, Egger M. Quality of reporting of randomized trials as a measure of methodologic quality. JAMA 2002 Jun 5;287(21):2801-4.

55. Devereaux PJ, Manns BJ, Ghali WA, Quan H, Guyatt GH. The reporting of methodological factors in randomized controlled trials and the association with a journal policy to promote adherence to the consolidated standards of reporting trials (CONSORT) checklist. Control Clin Trials 2002 Aug;23(4):380-8.

56. Schulz KF, Grimes DA, Altman DG, Hayes RJ. Blinding and exclusions after allocation in randomised controlled trials: Survey of published parallel group trials in obstetrics and gynaecology. BMJ 1996 Mar 23;312(7033):742-4.

57. Altman DG. The scandal of poor medical research. BMJ 1994 Jan 29;308(6924):283-4.

58. Pocock SJ, Hughes MD, Lee RJ. Statistical problems in the reporting of clinical trials. A survey of three medical journals. N Engl J Med 1987 Aug 13;317(7):426-32.

59. Three universities refuse to extend agreement for Elsevier journals. In: Duke news and communications; 2004.

60. Hotten R. Reed Elsevier wards off threats as scientists keep publishing. In: Times online; 2004.

61. Lerner E. Fraud shows peer-review flaw. The Industrial Physicist 2003;8 (6):12–7.

62. Altman DG. Better reporting of randomised controlled trials: The CONSORT statement. BMJ 1996 Sep 7;313(7057):570-1.

63. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: Updated guidelines for reporting parallel group randomised trials. J Clin Epidemiol 2010 Aug;63(8):834-40.

64. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: Updated guidelines for reporting parallel group randomised trials. PLoS Med 2010 Mar 24;7(3):e1000251.

65. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: Updated guidelines for reporting parallel group randomised trials.

BMC Med 2010 Mar 24;8:18.

66. Schulz KF, Altman DG, Moher D, Consort Group. CONSORT 2010 statement: Updated guidelines for reporting parallel group randomised trials (chinese version). Zhong Xi Yi Jie He Xue Bao 2010 Jul;8(7):604-12.

67. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: Updated guidelines for reporting parallel group randomised trials. J Pharmacol Pharmacother 2010 Jul;1(2):100-7.

68. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: Updated guidelines for reporting parallel group randomised trials. Trials 2010 Mar 24;11:32.

69. Moher D, Hopewell S, Schulz KF, Montori V, Gotzsche PC, Devereaux PJ, Elbourne D, Egger M, Altman DG, Consolidated Standards of Reporting Trials Group. CONSORT 2010 explanation and elaboration: Updated guidelines for reporting parallel group randomised trials. J Clin Epidemiol 2010 Aug;63(8):e1-37.

70. Moher D, Schulz KF, Altman D, CONSORT Group (Consolidated Standards of Reporting Trials). The CONSORT statement: Revised recommendations for improving the quality of reports of parallel-group randomized trials. JAMA 2001 Apr 18;285(15):1987-91.

71. Moher D, Schulz KF, Altman DG. The CONSORT statement: Revised recommendations for improving the quality of reports of parallel-group randomised trials. Lancet 2001 Apr 14;357(9263):1191-4.

72. Laxdal OE. Abridged index medicus. Can Med Assoc J 1970 Jun 20;102(13):1413.

73. Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, McQuay HJ. Assessing the quality of reports of randomized clinical trials: Is blinding necessary? Control Clin Trials 1996 Feb;17(1):1-12.

74. http://www.consort-statement.org/database/impact-of-and-adherence-to-consort/

75. Andrew E. Method for assessment of the reporting standard of clinical trials with roentgen contrast media. Acta Radiol Diagn (Stockh) 1984;25(1):55-8.

76. Goodman SN, Berlin J, Fletcher SW, Fletcher RH. Manuscript quality before and after peer review and editing at annals of internal medicine. Ann Intern Med 1994 Jul 1;121(1):11-21.

77. DerSimonian R, Charette LJ, McPeek B, Mosteller F. Reporting on methods in clinical trials. N Engl J Med 1982 Jun 3;306(22):1332-7.

78. Grant A. Reporting controlled trials. Br J Obstet Gynaecol 1989 Apr;96(4):397-400.

79. MAHON WA, DANIEL EE. A method for the assessment of reports of drug trials. Can Med Assoc J 1964 Feb 29;90:565-9.

80. Lai R, Chu R, Fraumeni M, Thabane L. Quality of randomized controlled trials reporting in the primary treatment of brain tumors. J Clin Oncol 2006 Mar 1;24(7):1136-44.

81. Al-Namankany AA, Ashley P, Moles DR, Parekh S. Assessment of the quality of reporting of randomized clinical trials in paediatric dentistry journals. Int J Paediatr Dent 2009 Sep;19(5):318-24.

82. Ethgen M, Boutron L, Steg PG, Roy C, Ravaud P. Quality of reporting internal and external validity data from randomized controlled trials evaluating stents for percutaneous coronary intervention. BMC Med Res Methodol 2009 Apr 9;9:24.

83. Taghinia AH, Liao EC, May JW,Jr. Randomized controlled trials in plastic surgery: A 20-year review of reporting standards, methodologic quality, and impact. Plast Reconstr Surg 2008 Oct;122(4):1253-63.

84. Adetugbo K, Williams H. How well are randomized controlled trials reported in the dermatology literature? Arch Dermatol 2000 Mar;136(3):381-5.

85. Scales CD,Jr, Norris RD, Keitz SA, Peterson BL, Preminger GM, Vieweg J, Dahm P. A critical assessment of the quality of reporting of randomized, controlled trials in the urology literature. J Urol 2007 Mar;177(3):1090,4; discussion 1094-5.

86. Moberg-Mogren E, Nelson DL. Evaluating the quality of reporting occupational therapy randomized controlled trials by expanding the CONSORT criteria. Am J Occup Ther 2006 Mar-Apr;60(2):226-35.

87. Plint AC, Moher D, Morrison A, Schulz K, Altman DG, Hill C, Gaboury I. Does the CONSORT checklist improve the quality of reports of randomised controlled trials? A systematic review. Med J Aust 2006 Sep 4;185(5):263-7.

88. Pat K, Dooms C, Vansteenkiste J. Systematic review of symptom control and quality of life in studies on chemotherapy for advanced non-small cell lung cancer: How CONSORTed are the data? Lung Cancer 2008 Oct;62(1):126-38.

89. Mathoulin-Pelissier S, Gourgou-Bourgade S, Bonnetain F, Kramar A. Survival end point reporting in randomized cancer clinical trials: A review of major journals. J Clin Oncol 2008 Aug 1;26(22):3721-6.

90. Ziogas DC, Zintzaras E. Analysis of the quality of reporting of randomized controlled trials in acute and chronic myeloid leukemia, and myelodysplastic syndromes as governed by the CONSORT statement. Ann Epidemiol 2009 Jul;19(7):494-500.

91. Kober T, Trelle S, Engert A. Reporting of randomized controlled trials in hodgkin lymphoma in biomedical journals. J Natl Cancer Inst 2006 May 3;98(9):620-5.

92. Barbui C, Cipriani A, Malvini L, Tansella M. Validity of the impact factor of journals as a measure of randomized controlled trial quality. J Clin Psychiatry 2006 Jan;67(1):37-40.

93. Farrokhyar F, Chu R, Whitlock R, Thabane L. A systematic review of the quality of publications reporting coronary artery bypass grafting trials. Can J Surg 2007 Aug;50(4):266-77.

94. Mills E, Loke YK, Wu P, Montori VM, Perri D, Moher D, Guyatt G. Determining the reporting quality of RCTs in clinical pharmacology. Br J Clin Pharmacol 2004 Jul;58(1):61-5.

95. Marashi SA. On the identity of "citers": Are papers promptly recognized by other investigators? Med Hypotheses 2005;65(4):822.

96. http://forums.studentdoctor.net/archive/index.php/t-384581.html

97. Vittinghoff *et al.*, Regression methods in biostatistics: linear, logistic, survival, and repeated measures models, 2005 page 149.

98. Vittinghoff *et al.*, Regression methods in biostatistics: linear, logistic, survival, and repeated measures models, 2005 page 138/

99. Feinstein AR, Cicchetti DV High agreement but low kappa: I. The problems of two paradoxes. J Clin Epidemiol. 1990;43(6):543-9.

100. http://creativecommons.org/licenses/by/3.0/us/