EVALUATION OF METHODS USED IN MEDICATION ADHERENCE RESEARCH

AN EVALUATION OF MEASUREMENT OF ADHERENCE AND PATIENT RECRUITMENT METHODS IN PATIENT ADHERENCE TO MEDICATION RESEARCH

By REBECCA JEFFERY, BSc (hons)

A Thesis Submitted to the School of Graduate Studies in Partial Fulfilment of the Requirements for the Degree Master of Science

McMaster University

McMaster University MASTER OF SCIENCE (2013) Hamilton, Ontario

TITLE: An evaluation of measurement of adherence and patient recruitment methods in patient adherence to medication research

AUTHOR: Rebecca Jeffery, B.Sc.
SUPERVISOR: Dr. R. Brian Haynes
NUMBER OF PAGES: x, 112

Abstract

Context: Interventions have been developed to improve adherence to medication, but trials have generally shown only small effects. This may be due in part to poor methodology, especially concerning the measurement of adherence and patient recruitment.

Objective: To provide an overview of the state of trial methodology concerning measurement of patient adherence and patient recruitment, and explore how the quality of these methods impact the adherence results found in trials.

Data sources: Major bibliographic sources, reference lists, and clinicaltrials.gov were searched for relevant trials up to January 2013.

Study selection: Approximately 150 trials were included in the full systematic review, from which 50 trials were selected to represent several methods of measuring adherence.

Results: There were a variety of different measures of adherence with qualities ranging from valid and unobtrusive, to unreliable and subjective. The median overall quality of adherence measures was 5 (IQR 3, maximum score 9, higher is better). The overall correlation of the quality of the measures of adherence and the coefficient of variation (CV) or proportion adherence suggested that adherence measures rated as higher quality were associated with a higher CV but not associated with a lower proportion adherence. The median overall quality of patient recruitment methods was 2 (IQR 1, maximum score 6, higher is better). Only 3 studies recruited only nonadherent patients. The correlation of the power of a trial and the quality of the patient recruitment methods, was slightly positively correlated for both binary and continuous data.

Conclusions: The quality of methods employed in adherence trials varies considerably and affects at least some findings of these trials. Very few studies recruited nonadherent patients or measured adherence at baseline. The importance of these differences in quality merits further study, but it is clear that better standards of adherence measurement are needed to support adherence research.

Acknowledgements

I would like to express my gratitude to everyone who has helped me throughout my thesis. I am very appreciative of all of the guidance and support I have received from my supervisor, Dr. Brian Haynes. Thank you for giving me the opportunity to work in the Health Information Research Unit, a dynamic research environment that provided an enriching educational experience that I am sure will benefit me in my future research endeavours.

This thesis on an evaluation of methodology in intervention trials is a substudy within a larger update on a Cochrane systematic review. This study uses articles found and extracted in that systematic review and therefore borrows the search strategy, screening and extraction interface and some questions.

The following people were involved in the full systematic review: Nancy Wilczynski, Emma Iserman, Dawn Jedras, Thomas Agoritsas, Robby Nieuwlaat, Alfonso Iorio, Niraj Mistry, William Zhang, and Reem Mustafa. Nicholas Hobson provided technical support and developed the online data extraction interface with my many, complicated questions. Special thanks to Tamara Navarro for helping me in the development of the data extraction form and with testing and training data extractors, and for answering my many questions. Brian Haynes (the PI) obtained funding for the full review and provided invaluable guidance throughout this project.

I would also like to thank my committee members, Dr. Lauren Griffith and Dr. Anne Holbrook who provided constructive feedback and help throughout the

development, analysis and writing of this thesis. You, and the rest of the Clinical Epidemiology and Biostatistics community at McMaster, have provided me with an excellent educational experience.

Last but not least, I would like to thank my family and friends for supporting and encouraging me throughout my academic pursuits. Mom, dad and Nick, you have been role models to me all my life and inspired me to pursue this path. And my friends, especially Sameer, Marissa, Melissa, and my HRM friends, you have encouraged me and provided me with many great memories during my Master's degree.

Financial Support

This research was funded by Canadian Institutes of Health Research Knowledge Synthesis Grant: KRS 262115. The author of this thesis was funded by an Ontario Graduate Scholarship, a CIHR Frederick Banting and Charles Best Canadian Graduate Scholarship, and has also received support from McMaster University and the Health Information Research Unit at McMaster University.

Table of Contents

1.0 Introduction

The primary goal of any clinician is to decrease the adverse effects of disease. This goal can be undermined by many factors, including low levels of patient adherence to medication regimens, which is common in chronic diseases such as diabetes and cardiovascular disease[1]. Low patient adherence to medications including placebo, has been shown to be related to patient outcomes, including mortality, in systematic reviews[2]. Adherence to medications in chronic disease has, to-date, been inadequate, averaging around 50%[3]. Adherence to clinician recommendations is even lower for lifestyle modifications such as diet and exercise[1,3].

1.1 Patient adherence defined

Patient adherence may be defined as the extent to which a patients' behaviour corresponds with the instructions they are given for a self-administered treatment [4–6]. Medication adherence is quantified as the amount of medication consumed by the patient divided by the amount of medication the patient should have consumed had they adhered to their prescribed treatment. Other terms historically used in the literature to describe this phenomenon include compliance and concordance. Compliance, or the extent to which a persons' behaviour coincides with medical advice[7], has been deemed too paternalistic, casting the patient in a passive role[8]. Concordance

attempts to improve upon compliance by placing the patient in the role of decision-maker while emphasizing a harmonious doctor-patient relationship[8]. Thus, concordance is what an informed patient decides to do, and there is "disconcordance" only if the patient does not do what they had agreed to do. This is a different concept than adherence, and will not be considered further here. Adherence is currently the preferred terminology for this complex issue; however, adherence and compliance are used interchangeably in the literature.

The problem of poor adherence was first reported by Hippocrates around 400 B.C., when he noted patients pretended to consume their medication, yet the problem persists to this day[8]. Poor adherence is also very common, with typical adherence rates averaging 50%, though adherence can range from 0 to over 100% for different regimens[5,9]. It has been estimated that only about half of patients consume enough medication to experience a therapeutic effect, though it is important to note that the threshold of adherence required for a therapeutic effect varies with different diseases and treatment regimens[6,7,10]. This is of importance to clinicians and patients as multiple studies have indicated the role of good adherence in improving clinical outcomes, such as controlling blood pressure[1,4,6,10]. Further, low adherence has been shown to be an independent predictor of patient outcomes, including mortality[2]. Interestingly, researchers have identified a "healthy adherer effect", wherein participants in trials who adhere to a placebo control have demonstrated better patient outcomes than participants who do not adhere to the active treatment[2,11]. Nonadherence complicates

the treatment of patients, as clinicians often respond to poor patient outcomes such as high blood pressure by prescribing a higher dose, providing a new diagnosis, or discontinuing medication[1,10]. Last, nonadherence is often considered to be a waste of resources in a constrained health care system, and estimates of its costs have been upwards of $100 billion annually[10]. However, the term adherence indicates the patients' ability to decide whether they will adhere to a treatment or not, and although adherence is central in improving health outcomes, respecting a patient's right to refuse treatment is also important[12,13].

Reasons for nonadherence vary among patients, and nonadherence may be either intentional or unintentional, each of which have different causes and should be addressed uniquely. Common reasons for nonadherence include concern over the perceived or real side effects of medication, the complexity of the treatment regimen, the patient's quality of life, the patient's awareness or knowledge about the disease and treatment regimen, health care system issues such as whether the patient has a family doctor, the cost of the medication, and general forgetfulness[4]. Doctors and researchers have attempted to improve adherence by developing interventions that target these reasons.

1.2 Testing interventions to improve adherence to medication

Attempts have been made to improve patient adherence to medications through interventions such as reminders and education, though studies have found singular interventions often fail[6,14]. Adherence is a complex issue requiring complex interventions, and poor adherence is often due to multiple factors[2,15]. Further, the conduct of trials of interventions to enhance adherence to medication can be complicated and must assess both changes in adherence as well as patient outcomes, as interventions must decrease morbidity and mortality to be of use[16]. Though no single intervention is likely to prove a cure-all for nonadherence, methodological issues in intervention trials also may explain some of the failures and inconsistencies in the evaluations of these interventions[6].

Well-designed randomized controlled trials (RCTs) are the cornerstone for testing the efficacy of interventions in healthcare[7,17]. However, flaws in trial methodology can limit the validity and accuracy of the results found[6,18–20]. This was further supported by a meta-analysis by Moher et al. investigating the impact of the quality of the methodology used in RCTs on outcomes[21]. Methodological issues in adherence intervention trials include the measure of adherence, touted to be the single most difficult question in the realm of adherence research, the recruitment of patients into trials, as well as others such as the reporting of outcomes[1,22,23]. Trials from the previous systematic review[6] showed considerable variation in these areas, and it is hypothesized variations in these methods are a source of disparity in the conclusions of the trials. This is supported by a meta-analysis of the correlation between patient outcomes and adherence rates, which

included an analysis of some methodological factors and their impact on the effect size[1]. Dimatteo's review, however, looked at adherence outcome reporting, whether self-report was used, and whether one or multiple adherence measures were used, not the impact of individual measures on this relationship[1]. Further, this study suggested future research into the impact of study methodology on adherence rates and clinical outcomes is necessary, but did not discuss the impact of patient recruitment methods, that is, how subjects were identified, declared eligible and entered into trials, might have on adherence intervention trials. Other literature has discussed the importance of patient recruitment and retention in ensuring reliable results[18].

Despite the acknowledged importance of these methodological aspects, a search for studies that quantify the quality of these methods did not return any results. Therefore, to determine the impact of trial methodology on outcomes, criteria for the quality of adherence measures and patient recruitment were sought based on reviews and studies assessing measurement properties of measures of adherence. A scale was created out of this review of the literature to allow for assessment and comparison of methods between trials of adherence interventions, drawn from an update of a Cochrane review on interventions to improve patient adherence to prescribed medications[6].

This project investigated two key methodological aspects of studies of interventions to improve patient adherence to medication. Such methodological problems as the measurement of adherence and patient recruitment have been understudied, though

adherence measurement has been acknowledged to be the single most difficult question in the realm of adherence research[4,8,14,24]. There have been calls for standardization of measurement of adherence so intervention trials become comparable, yet this is difficult when no gold standard exists [24]. Patient recruitment in particular has a paucity of information available, despite acknowledgments of its impact on study quality and validity[20]. Terms frequently used are defined in the glossary under Appendix 1.

1.3 Research Questions and Hypotheses

Overall Research Question

What is the quality of the methods used in adherence trials, specifically the quality of measurement of adherence methods and patient recruitment methods, and what is the impact of that quality on the adherence results in these trials?

Sub-Research Questions and Hypotheses

1.3.1 Measures of Adherence Primary Research Questions

1. Based on an overview of the literature, what are the advantages, disadvantages and quality aspects, including the validity, reliability, objectivity, unobtrusiveness and longitudinality, of different measures of patient adherence to prescribed medications?

2. What are the types of adherence measures used in a sample of randomized trials of interventions to increase patient adherence to prescribed medications, and what are their measurement qualities? (an analytic survey)

1.3.2 Measures of Adherence Secondary Research Question

1. How does the quality of measures of adherence affect adherence results in randomized trials of interventions to increase patient adherence to prescribed medications?

1.3.3 Hypotheses for Secondary Research Question

a) In trials with continuous adherence rates, higher quality measurement of adherence is correlated with a higher value of a measure of precision of the adherence results, defined by the coefficient of variation (CV).

b) In trials with binary adherence data, the proportion of control group patients who are found adherent decreases with higher quality measures, as bias for lower quality measures tends to overestimate adherence.

1.3.4 Patient Recruitment Primary Research Questions:

1. What is the quality of patient recruitment methodology in a sample of adherence intervention trials, based on important methodological features of patient recruitment, including recruitment of patients with low adherence rates, representativeness of recruited patients, and stratified reporting of adherence levels according to baseline adherence? (an analytic survey)

1.3.5 Patient Recruitment Secondary Research Question:

1. How do patient recruitment methods affect the statistical power of adherence results in adherence intervention trials?

1.3.6 Patient Recruitment Secondary Question Hypothesis:

Studies with higher quality patient recruitment methods are more likely to be powered to detect a clinical effect.

The hypotheses for each research question are based on the assumption that study methods quality, for measuring adherence and selecting patients, and adherence outcomes and patient outcomes would not be associated. This is due to the abundant confounding factors in these trials, such as the heterogeneity and potency of the intervention, the multitude of different diseases being considered, or the variability of patient characteristics. For example, disease states in this review range from asthma and infection to diabetes, and corresponding patient outcomes range from quality of life to glycated hemoglobin (Hb_{A1c}) levels. Thus the above hypotheses will be tested to investigate the relationship between the quality of the methodology, and the precision, magnitude, or power of adherence results.

These variables were chosen based on whether the data were binary or continuous. The precision of the results is an important consideration that may be impacted by the measurement method and is a question that is yet to be answered. For continuous data, precision can be investigated with the standard deviation alone, but

the coefficient of variation (CV) was chosen here to eliminate the units from the analysis, as most outcomes are measured in different ways across trials. For example, adherence results based on a measure of the amount of drug in a patient's blood may be in the units of mol/L, whereas a pill count measure of adherence may be reported as number of pills counted. By dividing the standard deviation by the mean, the units cancel and measures become comparable for their variance in relation to their mean.

This relationship was not explored for binary data, however, because the standard deviation of a proportion is directly related to that proportion. Thus we investigated the correlation of the proportion adherent with the measurement quality instead. This relationship was thought to be a valuable investigation, as it is often suggested that poor quality measures of adherence overestimate adherence. We only explored the proportion adherent in the control group because the proportion adherent in the intervention group would vary with the potency of the intervention, and would therefore not demonstrate the true association between the quality of the measure and the proportion adherent.

Different variables were used for patient recruitment methods as it is self-evident that the value of an estimate of precision of adherence estimates will increase when certain patient recruitment methods are implemented, such as recruiting only nonadherent patients. Thus the relationship of the precision of results to the quality of the methods was not worth exploring. Instead, we chose to investigate the relationship between the power of a study to detect an effect and the quality of patient recruitment

methods employed, as this relationship has not been established. Finding adequately powered results is integral to allowing researchers to draw inferences from those results. Given that results of trials to improve adherence often show no effect, it will be worthwhile to establish whether the methods of recruiting patients into the trial might impact the power of the overall study.

2.0 Methods

This analytic survey is based on data from a subset of randomized controlled trials included in a systematic review update on interventions to improve patient adherence to medications, which was a separate study. Methods for obtaining this dataset are detailed elsewhere[6]. In brief, the systematic review searched MEDLINE, EMBASE, CINAHL, PsychInfo, Sociological abstracts and the Cochrane Library, using key terms such as “compliance”, “medication” and “clinical trial”. Articles retrieved from this search were screened by title and abstract, then by full text in duplicate by independent reviewers and disagreements were resolved through adjudication by a third reviewer. Data was then extracted and disagreements were resolved in the same manner. The 50 articles included in this thesis were extracted first by the review team, prior to extraction of all other trials included in this systematic review update, with me as one of two reviewers on each extraction. The data from these extractions are used here, as my thesis, and was analysed separately by me. Methods to categorize adherence

intervention methods for measuring patient adherence and recruiting patients for such trials are summarized here. The remainder of the methods will describe the approach used to answer the research questions is section 1.3, beginning with the methods used to provide an overview of the measures of adherence currently in the literature, followed by the development of the quality scales, and the extraction and analysis of data from a sample of trials in the full systematic review for the secondary research questions' analyses.

2.1 Measures of adherence: classification

A chart of measurement methods was created based on a review of the literature in order to answer our first research question on what are the advantages and disadvantages of different measures of adherence. This overview involved a key word search in MEDLINE and PubMed Clinical Queries of the terms "weights and measures"[MeSH Terms] OR measure [Text Word] and "patient compliance"[MeSH Terms] OR "compliance"[MeSH Terms] OR compliance [Text Word]. This search strategy was developed in collaboration with a research librarian. The results of the overall systematic review search were also screened for trials that related to this objective, by allowing screeners to flag articles that appeared to discuss adherence measurement for later review. This approach allowed the overview of measures of adherence to be comprehensive. Articles that addressed measurement properties such as reliability and

validity, or systematic reviews that discussed the merits of different measures, were included in this overview. This chart included information on the features listed in Table 1, as well as values for the sensitivity, specificity, or reliability of each measure, to compare the relative merits of each adherence measure to other measures. Information from articles was categorized into each quality aspect and each type of measure by extracting this key information from each article. This extraction was done by myself, not in duplicate, as this aspect of the thesis was not the main objective, and thus it was deemed unnecessary to complete this via a systematic review. Measures were not organized in any sort of hierarchy given that there is no gold standard in adherence measurement and the debatable relative importance of each feature listed in Table 1. For example, though an objective and valid measure may theoretically give an accurate picture of adherence, if that measure is obtrusive to the patient, they may alter their adherence behaviours and pretend to consume their medication. Also, some objective measures such as drug concentration in patient blood depend on patient pharmacodynamics. This overview of measures of adherence and important methodological features of those measures was used during the development of the quality scale to answer the other research questions related to the measurement of adherence outlined above.

2.2 Development of dataset

2.2.1 Desirable features of adherence measurement method

Based on the overview of the literature on measures of adherence, a scale was developed based on important methodological aspects identified through that search. Many aspects of measures of adherence were identified in the literature to be of methodological importance, including reliability and validity, objectivity, unobtrusiveness, and the degree to which the measure provides longitudinal data ("longitudinality"), described in Table 1[1,7,14]. Other features noted in the literature include directness, sensitivity and specificity, and feasibility; however, these aspects were judged to be covered by the five listed previously, thus including these aspects in our scale would double count that aspect of a measure giving it more weight in our quality scale[25]. For example, sensitivity and specificity are ways of expressing validity, and feasibility in part relates to the reliability and obtrusiveness of a measure. Thus we have chosen the five features in Table 1 as a basis for surveying the methods of measurement used in trials of adherence interventions.

Reliability and validity, as aspects of quality of a measure of adherence, refer to whether the scale reproducibly and accurately measures what it purports to[25]. Reliability includes test-retest and inter-rater reliability, while validity includes content validity, criterion validity, and/or construct validity, all of which are important to investigate when developing a measure of adherence. The objectivity of a measure of adherence

relates to the minimization of bias based on the design of the measure. For example, pill counts are designed to be objective, whereas self-report by patients and judgments by providers are inherently subjective. An unobtrusive measure would be one in which the patient is unaware the measurement is being taken, minimizing the risk of bias. Pharmacy refill data provide an unobtrusive measure, as long as patients are unaware it is being used to measure adherence, avoiding the Hawthorne effect[14]. Last, the period of follow-up covered by data is important when measuring adherence. Longitudinal data refers to a measure that provides adherence over a period of time. This period should be longer than 6 months for chronic medication adherence, and timing is important even when measuring adherence to antibiotics over a shorter period of time, as early adherence frequently is higher than adherence towards the end of the prescription period.

Though Table 1 outlines examples of measures for each feature, it is important to note that there is no perfect measure and the properties for each measure that appear desirable can be compromised by poor methodology. For example, medication event monitoring systems (MEMS), wherein medication is dispensed in an electronic pill bottle that counts each time the lid is removed and theoretically each time medication is consumed, cannot determine if the medication has actually been consumed by the patient and is therefore an indirect measure of adherence. MEMS can only determine objectively the pattern and number of times the bottle is opened as a proxy for medication consumption. Further, MEMS are obtrusive to patients as it can become

obvious the bulky lid is monitoring bottle use, and this certainly becomes clearer when researchers emphasize the importance of using the bottle properly and returning it at follow up appointments.

An ordinal rating scale was developed in this thesis to assess the quality of measurement employed in each included trial. This scale was based on the principles of measurement, using the features of an ideal measure (valid, reliable, unobtrusive, objective and provides longitudinal data in Table 1) as a reference point for the measurement of adherence. The scale for each aspect was scored as Yes, No or Uncertain[26]. These ratings were given "dummy" values of 0 to 2, so that I could translate the rating of the trials' methodology into a value to correlate with adherence statistically. In this scale, a "No" received a value of 0, a "Yes" received a value of 2, and "Uncertain" received a value of 1, as it is neither certainly absent nor certainly present.

The ratings within each aspect of the scales for the two research questions (the measurement of adherence methods and patient recruitment methods) were derived through a literature search detailed above in the classification of measures. The categorization of each feature is detailed in Table 2 for the measurement of adherence quality scale. The phrasing of each category was revised repeatedly through pretesting with data extractors to try to improve the reliability of this scale. It is important to note the properties of each measure that was found in the review of the literature and reported in Table 4 are general and may not apply in all situations. If a study in our analytic survey mentions that the measure of adherence used is reliable, or any other

property of interest, or if the authors cited a study that demonstrated a property of the measure, that measure was categorized higher than it would be otherwise. For example, self-reports can be elicited in any number of ways (eg, patient diaries, questioning by a doctor or nurse or third party, formal questionnaire), many of which are of unknown reliability and validity. Thus a self-report can be categorized as valid if the authors cited a validity study that documented significant correlation with a criterion or concurrent measure (given there is no gold standard in adherence measures). For example, the ASK-20 self-report questionnaire was validated by correlating the results with those collected from pharmacy refill records[27–29].

2.2.2 Patient recruitment methodology

The methods related to the recruitment of participants in trials of adherence interventions were investigated, as recruitment and the subsequent reporting of patient adherence have important methodological implications[7,18,19]. In general, recruiting patients based on their baseline adherence levels, the representativeness of the sample, and whether the results are reported by baseline adherence, were sought and analyzed[20,30]. This aspect of this thesis is referred to as “patient recruitment methodology”.

Recruiting only nonadherent participants will result in more powerful detection of intervention effects, as adherent patients may blunt the effect of the intervention to

improve adherence through a ceiling effect[30]. An adherence estimator was recently validated to predict patient nonadherence, based on preconceived health beliefs, which could also be used as a recruitment tool[31]. However, lumping patients into groups of compliers or noncompliers on a purely statistical basis (above or below a certain level of adherence) ignores important behavioural and biological considerations, such as the level of adherence required to achieve the desired therapeutic response[24]. Yet recruiting patients with low initial adherence is important to ensuring the full effect of the intervention may be detected.

The representativeness of the study sample was measured by assessing the number of patients asked to participate in the study prior to the investigators reaching their sample size. This aspect is based on the assumption that a sample is less representative (i.e., suffers from selection bias) if a large proportion of eligible patients decline before the study population is recruited. Though this comparison might not take into account other biases such as volunteer bias, this approach was the most feasible and objective method for assessing representativeness.

Last, reporting of results by initial adherence status may clarify whether the intervention had an effect in those who needed it most. If the results were analyzed in two groups when adherent and nonadherent patients were recruited with baseline adherence measurements taken, then the study was considered to have done this. If only nonadherent patients were recruited, the study received the highest rating in this category as results would be reported based on all patients being initially nonadherent.

These aspects of patient recruitment, representativeness and reporting methodology may partially explain inconsistent findings of adherence intervention efficacy across trials. The scale gave each of these three items an assessment of Yes, No or Uncertain[26]. These ratings were then given "dummy" values of 0 to 2, to translate the rating of the trials' methodology into a number to correlate with adherence statistically. In this scale, a "No" received a value of 0, a "Yes" received a value of 2, and "Uncertain" received a value of 1, if it was neither certainly absent nor certainly present. The categorizations of these dimensions are detailed in Table 3. The phrasing of each category was revised repeatedly through pretesting with the data extractors to maximize the reliability of this scale.

2.2.3 Data extraction

Data extraction was completed in combination with the update of the Cochrane review upon which the dataset for this thesis was based. The full extraction form is available in Appendix 2, and the aforementioned scales (Table 2 and 3) were a part of this extraction and were the main focus of this thesis. In brief, data extraction involved duplicate extractors reading included trials and extracting data through an online data management system. Disagreements between extractors were resolved through adjudication by a third extractor. The data extraction form was tested extensively during its development in an attempt to improve the reliability of the form's use.

2.3 Outcome variables

2.3.1 Primary outcome

Adherence outcomes, both binary and continuous, were included during data extraction and used in this analysis. Quality scores were a primary outcome of interest, though only adherence measurement methods that had accompanying adherence results in each trial were used for the quality score for trials. Only one quality score was available for each trial's patient recruitment methodology, and this too was a primary outcome of interest.

When multiple measures of adherence were used in trials, a composite quality score was calculated by taking the maximum score for each domain (2 on a 0-2 scale, for a maximum of 9). This was done when multiple measures of adherence were used and no single measure had the maximum possible quality score for that trial. Using multiple methods to measure adherence may provide more valid adherence data, as one measure may provide information or an aspect that another measure lacks.

2.4 Statistical analysis

2.4.1 Sample selection

The analysis was done on a sample of 50 purposively selected studies, half (25 studies) from the review published in 2008 (for which the literature search was completed in 2007) and half from the current update of the review (2007-2013). A sample of both new and old studies was used to acquire a distribution of quality scores, given the hypothesis that newer studies may have a higher quality score. A purposive sample was selected instead of a random sample given our desire to include all measures of adherence in our analysis, thus studies with various measures of adherence were selected to ensure heterogeneity. Though a stratified random sample may have been the best approach for sampling these articles, this approach was not possible as the article extractions have not yet been completed. In order to choose the 50 studies, the measures of adherence used in each study were assessed for all studies included in the published review, and studies were then selected based on the measures used, aiming for at least 3 studies per measure. However, few studies used some types of measures such as attendance, drug concentration from blood or urine, and direct observation, so studies that included those measures were included and the remaining studies aimed to balance the other measures (self-report, pill count, MEMS, etc.). Some studies were found to be excluded based on inclusion/exclusion criteria for this review after they were included in our sample, and were subsequently replaced in the manner studies were originally selected, that is, through a purposive selection of studies with a certain measure of adherence. For example, some articles were found during extraction

to not meet the inclusion criteria of 80% follow up, which was missed during initial screening due to the complexity of reporting in some articles[32].

2.4.2 Primary analysis

This study was an analytic survey which aimed to describe, classify and rate the measures of adherence and participant recruitment methods in trials of adherence interventions. The primary analysis consisted of descriptive statistics, including a summary of the median quality of the studies' methodologies, the frequency of different types of measures and quality scores, and the range or interquartile range for each median quality score.

2.4.3 Secondary analyses

The association between adherence intervention trial methodology, in terms of measures of adherence and patient recruitment methods, and the precision, magnitude or power of adherence results reported in the trial was investigated. This was quantified by a correlational analysis, using a Spearman correlation, the dependent variable varying for each hypothesis, which differed by the methods aspect and the type of data (binary or continuous), and the independent variable being the summary quality score of methods for measurement of adherence and patient recruitment, quantified by the

scale developed for this thesis. For continuous adherence outcomes for the adherence measurement research question, the dependent variable was the precision, as measured by the coefficient of variation for each trial's adherence outcome. The coefficient of variation (CV) was calculated for each study by dividing the standard deviation of the data for a measure by the mean of that data for that measure (CV =SD / mean) (the "noise to signal" ratio). For example, in an article by Hederos et al.[33], a self-report questionnaire that was quantified as a visual analogue scale (VAS) found a mean of 16.4 for adherence in their population and a standard deviation of 24.2 for this measure[33]. Therefore the CV was 1.48 (=24.2/16.4) in this study for this measure. A higher quality measure of adherence is hypothesized to be correlated with a wider range of adherence in each trial, i.e. the quality will be positively correlated with the CV. This is because a biased method of measuring adherence consistently overestimates adherence, therefore the standard deviation for the distribution of adherence scores would be lower. Binary adherence outcomes for the measurement of adherence research questions used the proportion adherent, as defined by the study, as the dependent variable in a bivariate, Spearman correlation. The proportion adherent was a value ranging from 0-1 reported in the study when adherence was dichotomized through the use of cut-offs. For example, 57.6% of patients took greater than 80% of their pills based on a pill count in a study by Henry and Batey[34].

The patient recruitment method quality investigated how the overall quality score (for both continuous and binary adherence outcomes) correlates with the power

of the study to detect an effect (using the Spearman correlation). The hypothesis was that studies that recruited nonadherent patients ought to have greater power to detect the effect of an intervention to improve adherence. The scores for each quality feature were dichotomized for a within scale item comparison, from 0-1-2 to 0-1, due to the large difference in the number of studies with certain quality scores. The two consecutive scores with relatively smaller values compared to the other score were lumped together. This varied for the three aspects of recruiting patients. For the quality scale items of whether studies recruited patients based on baseline adherence (i.e. only recruited nonadherent patients) and for the item of whether results are reported based on baseline adherence levels, the scale values of 1 and 2 were lumped into a score of 1 and compared to the 0's in this analysis. For the other scale item of recruiting a representative patient sample, the 0 and 1 scores were lumped into a 0 score and the 2 score became a score 1, as there were few 0's in this aspect and we aimed to make the sample size in the two scores most comparable through this dichotomization. A Wilcoxon rank sum test (i.e. Mann-Whitney U test) was used to analyse the effect of each quality aspect, comparing the two quality score's power to detect a difference. Power was evaluated by comparing the study sample size to a calculated theoretical sample size, using a clinically important difference in means or proportions of 25%, 2-tail α= 0.05, and β=0.8. A ratio of actual sample size over the sample size theoretically powered to detect an effect was then calculated and used as the dependent variable (i.e., % recruitment= actual sample size / theoretical sample size). For instance, in a

study by Peveler et al[35], 48 patients were included in the control group, but a sample size of 57 patients in each group was required to detect a 25% difference in the intervention group based on the proportion adherent of 0.51 in the control group. The percentage recruitment in this situation would then be the actual sample size divided by the theoretical sample size, or 48 / 57 = 0.85, or 85% of the recruited sample size was obtained in this study. If the proportion adherent in the control was greater than 75%, thus p_1+MID > 1, a maximum proportion of 100% was imputed for such instances. An example of this is when the reported proportion adherent in the control group in a study is 80%, thus the calculated proportion in the intervention group once the 25% minimal difference is added would be 105%, which would not allow for a sample size calculation. Thus a minimal difference of 20% is imputed instead to create a ceiling of 100% adherence.

Analyses were completed in SPSS version 20 and a 2-tail $p<0.05$ was used to determine statistical significance. Data were originally going to be transformed using an inverse (1/x) or the natural logarithm (ln(x)) transformation. These transformations are used when data is skewed and J-shaped, to make the data more normally distributed. However, even after transforming data to allow for t-test assumptions to be valid (i.e. normality of data), the data was still too skewed for a t-test to be used. Therefore a Wilcoxon rank sum test was used to test differences between the two groups. Inter-rater agreement was quantified for data extractions and scale ratings using an unweighted kappa (κ).

2.4.4 Exploratory Analyses

For the investigation of the quality of measurement of adherence methods, when a composite measure quality score was higher than any single measure's quality, that composite score was substituted for that study. These new quality scores were compared to the overall correlation with the precision or proportion as a sensitivity analysis. For the patient recruitment methods correlation of power with quality, we also tested clinically important differences of 15% and 35% and qualitatively compared these extremes to the main 25% minimum difference. The effect size (=mean $(p_1-p_2)/ SD_{pooled}$, where p=proportion adherent, p_1 is the intervention group and p_2 is the control group) was compared across dichotomized patient recruitment categories to investigate the comparison of low quality patient recruitment methods' mean effect size observed to the effect size found in higher quality patient recruitment methodology studies. Trends in study quality over time was also looked at by comparing the quality in studies published by year of publication using a Spearman correlation and visually examining the scatterplot of this data.

3.0 Results

The flow of studies in this systematic review is shown in Figure 1 and is discussed in greater detail in the full systematic review update[6]. 50 studies were included in this

analytic survey based on the types of measures of adherence they utilized, as detailed in the Methods. Table 4 summarizes the measurement qualities of different measures of adherence in the literature. Tables 5 and 11 provide the methodological quality of each trial for measurement of adherence and patient recruitment methods, respectively. Table 6 reports the frequency of different measures of adherence and the median quality overall and per quality feature for each measure. Table 7 reports the number of studies with different numbers of adherence measures, from one to five measures. Tables 8 and 13 provide all of the data used in this study to compute correlations to answer our secondary research questions for both the measure of adherence and patient recruitment methodological areas. Tables 9 and 10 report the correlations found between the quality of measures of adherence and the precision or proportion adherent. Figure 2 illustrates the correlations found as a graph with a line of best fit for the overall correlation between the quality of the measure of adherence and the proportion adherent. Table 12 reports the patient recruitment methodological quality summary, while Tables 14-19 report the Wilcoxon rank sum test results for each quality feature of patient recruitment quality.

For screening and extracting of data, agreement between extractors was fair, with kappa values of 0.342 (95% CIs 0.114 to 0.678) for 50 extractions. Disagreements between extractors were resolved by a third extractor. Authors were contacted to provide their revisions on our extractions of their article, which resulted in some changes to our original data, based on authors' comments. A substantial proportion of

authors, 29/50 (58%), replied after three emails were sent to them, the maximum number of attempted contacts.

3.1 Classification and description of adherence measures

The overview of the adherence measurement literature to answer the primary research question related to the measurement of adherence methods, found 36 different measures of adherence from the initial review of the literature during the scale development for this thesis. The analytic survey of 50 randomized controlled trials of interventions to improve adherence, for the second primary research question related to the measurement of adherence methods, identified an additional 8 unique measures, giving a total of 44 different measures. These include a generic self-report of adherence, as well as different methods of pill counts. Many of these measures were specific self-report questionnaires, which often contained similar questions and generally had the same advantages and disadvantages of being easy to use and economical but obtrusive to patients and subjective. These have been grouped into the measures listed in Table 4, which summarizes their advantages and disadvantages, as well as what the literature notes for each aspect of quality of measurement.

Twenty of these measures were considered by the literature to be valid in contexts in which validity has been tested, and three were considered not to be valid based on empirical assessments by the study author or by another study of that

measure, including the physician assessment. Other measures have not been tested for validity or may have multiple types of measures within a method of measuring adherence, with some subtypes being valid and others not valid or not validated. For example, this table does not include all possible methods of measuring therapeutic response, though different therapeutic response measures are of variable validity. Twenty-six measures of adherence were considered to be reliable in the literature. Twelve measures were considered objective, though only four of those are direct measures of adherence. Nine measures were considered unobtrusive. Most measures of adherence are capable of measuring adherence over a long period of time (longitudinality), however self-reports are often only capable of measuring the previous week of adherence due to recall bias, and blood drug concentration assessments are only capable of measuring short periods of time when the drug in question has a short half-life. Certain measures are often considered to be a "gold standard" (for example, MEMS[4,28,36–38]), yet Table 4 illustrates the drawbacks of each measurement type, which exemplifies the lack of a true gold standard in this area of measurement, including for MEMS[39].

Other measures that are considered by some to be the best measure of adherence, such as pharmacodynamic measures, were not included in this overview because no trials that discussed measures of adherence or test interventions to improve adherence used these methods. Only measures of adherence that were explicitly stated in an article to be a measure of adherence were included here. However, given the

broad classes of some measures of adherence, such as therapeutic response which can include outcomes such as glycated hemoglobin (HbA1c) so long as the trial states that that measure is quantifying adherence, these other measures might be included.

3.2 Analytic survey of measurement methods used in trials

Table 5 provides the raw data of the quality of each measure per study. Table 6 summarizes the quality per measure overall in the 50 studies included in this analytic survey as well as the median quality in each feature for each measure. The median overall quality of measures of adherence was 5 (IQR=3; maximum, possible score 9). Though studies were selected based on their measure of adherence, few studies in the full systematic review included certain types of measures, such as “appointment keeping”, “clinician judgment”, “direct observation”, and “therapeutic response”. These infrequently used measures of adherence generally had low quality scores, though therapeutic response was an outlier with a median quality score of 6 (out of maximum 9). Other measures of adherence, such as MEMS, pill counts and pharmacy records, were used more frequently in the literature and were of higher quality than other measures. Self-reports were an anomaly in this regard, as they were used quite frequently but were of lower quality (median score of 3). Self-reports are often used in conjunction with other, higher quality measures, as they are user friendly, economical, and able to provide qualitative data on patterns of adherence. Pharmacy refill records

and body drug concentration measures had the highest median quality score of 8 and 7, respectively. These measures fulfill all of our scale criteria, as they are usually unobtrusive to patients. Direct observation is also a good measure of adherence as it is direct and objective; however it is very obtrusive to patients and was thus of lower quality (median score 4.5). Clinician judgment continues to be considered the poorest method of measuring adherence, as it has been previously invalidated[40].

Table 7 reports the number of measures of adherence per study and proportion of studies reporting each number of measures. One study used six and five measures of adherence for the adherence assessment, two studies used four measures for the adherence assessment, five studies used three measures, and 18 studies used two measures (Table 7). It is thought that using more measures will lead to a more representative assessment of adherence, as different aspects of adherence can be evaluated by different measures. This was found to be the case in eight of 27 (29%) trials that included multiple measures of adherence, which had a composite quality score that was higher than any single measure of adherence quality. This was because different measures often have different drawbacks, such as a pill count being obtrusive or a questionnaire not being valid. In the other 20 studies the quality of a single measure was higher in all quality aspects (validity, reliability, etc.) than any other measure included in that study.

3.3 Secondary analysis of measurement of adherence data

Table 8 provides the data from each study. Out of 50 studies, 21 (42%) reported adherence data in a continuous manner, though four of those studies did not report exact p-values or standard deviations to allow the data to be used in the analysis, but did include adherence data reported in a binary manner and therefore were still included in this study. The proportion of studies reporting adherence data as a binary variable was 38/50 (76%). Ten of 50 (20%) studies reported data both continuously and as a binary adherence outcome. As some articles that included multiple types of measures of adherence did not report data per measure, not all measures were included in secondary statistical analysis; however, these measures were still included in the overall descriptive analyses.

3.3.1 Continuous data

The overall Spearman's correlation of the quality of the measures of adherence and the coefficient of variation of adherence results for a given measure of adherence was 0. 663 (95% CIs 0.386 to 0.830, $p<0.001$), meaning that higher quality scores were associated with a larger variation in the estimate of adherence, as quantified by the coefficient of variation. This indicates that the quality of the measure may impact the adherence results that are found in a trial. Table 9 contains the Spearman correlations

for each measure. Several measure types did not have data reported continuously or not enough data to compute a correlation. While the values could easily have arisen by chance as the correlation for all three measures was nonsignificant, the direction of correlation corresponded to our hypothesis of a positive correlation between quality and precision for three measures, MEMS, pharmacy refill record, and pill count.

3.3.2 Binary data

The overall correlation and correlations for individual correlations were not statistically significant. The overall Spearman's correlation of the quality of the measures of adherence and the proportion adherent was -0.210 (95% CIs -0.436 to 0.041, p=0.101), meaning a higher quality measure of adherence may be associated with a lower proportion adherent in a trial. This further supports the idea that the quality of the measures of adherence used might impact the adherence results found in a trial. Table 10 contains the Spearman correlation of each measure. Although none of the results for individual measures was statistically significant, some individual measures' correlations corresponded to the hypothesis of a negative correlation between quality and proportion of patients who were reported as adherent, while others had a positive correlation contrary to the hypothesis.

3.4 Patient recruitment methods

The overall median quality of patient recruitment methods was 2 (IQR=1, maximum possible score 6). Table 11 details the patient recruitment quality scores in each study while Table 12 summarizes that information. Few studies (3/50) recruited only nonadherent patients. This is a key finding of the study. Most studies in our sample recruited "representative groups" of patients. A study's sample was considered to be representative in our quality scale if they recruited more than half of the patients that were eligible for entry into the trial, as reported in the trial. For instance, a study by Rickles et al. reported that 98 patients who met the eligibility criteria were approached for entry into the study of which 63 were included, which exceeds our required ratio of less than 2:1 eligible to included patients[41]. Thus in this study, the sample was considered to be representative of the eligible population. Though this approach to estimating representativeness does not take other biases such as volunteer bias into account, this method was considered the most feasible and objective way to estimate this quality aspect. Fewer studies recruited patients based on their baseline or background adherence rate than studies that recruited a representative number of patients, and few studies reported their results based on baseline adherence. The data from each study and calculated percentage recruitment are provided in Table 13.

3.5 Patient recruitment secondary analysis

3.5.1 Continuous data (Tables 14-16)

The overall Spearman correlation between the quality score and the percentage of sample size recruited in a study was not statistically significant (0.038 (95% CIs -0.347 to 0.412) p=0.851). Each of the three quality aspects that were included in the patient recruitment methods quality scale were investigated individually. The quality score of each scale aspect was dichotomized to 0 or 1, rather than the original 0, 1 or 2, due to the uneven spread of the scores, as discussed in the Methods section. These dichotomized quality scores were then compared to the power of a trial to detect an effect using a Wilcoxon rank sum test, to compare the two groups. This test was used in order to determine if there was a difference between the power of a trial to detect an effect based on the quality of the method used for each quality aspect. The results of the Wilcoxon tests for each quality aspect, comparing the percentage recruited across dichotomized quality scores, are reported in Tables 14-16. Overall, with the exploratory nature of these analyses and lack of significant differences in mind, our findings on the relationship between the quality of patient selection methods and the power to detect an effect for continuous data were inconclusive. Thus we cannot determine with certainty whether any of the quality aspects impacts the power of a trial to detect an effect.

Though p-values for ANOVA and Wilcoxon rank sum tests were similar, only the results of Wilcoxon tests were reported because analysis by three quality score categories had largely disparate numbers in each category making a dichotomization of the quality score more appropriate for analysis. The statistical appendix provides distributions of the three quality scale scores for the three aspects of patient recruitment quality (Figures 1-3). The two consecutive scores with smaller values were lumped and compared to the other score, which varied for the three aspects. For the quality scale items of whether studies recruited patients based on baseline adherence (i.e., only recruited nonadherent patients) and for the item of whether results are reported based on baseline adherence levels, the scale values of 1 and 2 were lumped into a score of 1 and compared to the 0's in this analysis. For the other scale item of recruiting a representative patient sample, the 0 and 1 scores were lumped into a 0 score and the 2 score became a score 1, as there were few 0's in this aspect and I aimed to make the sample size in the two scores more comparable through this dichotomization.

3.5.2 Binary data (Tables 17-19)

The overall correlation of the overall quality and the percent recruitment was also not statistically significant (0.140 (95% CIs -0.127 to 0.388), p=0.303). The results of the Wilcoxon rank sum tests comparing transformed percentage recruited across

dichotomized quality scores are reported in Tables 17-19. Overall, the relationship between the quality of patient recruitment methods and the power of the trial to detect an effect for binary data was indeterminate. No comparisons were significant, thus it remains to be seen whether the quality of these methods might impact studies' power.

Though p-values for ANOVA and Wilcoxon rank sum tests were similar, only results of Wilcoxon tests were reported because analysis by three quality score categories had largely disparate numbers in each category making a dichotomization of the quality score more appropriate for analysis. The statistical appendix provides distributions of the three quality scale scores for the three aspects of patient recruitment quality (Figures 1-3). The two consecutive scores with smaller values were lumped and compared to the other score, which varied for the three aspects. For the quality scale items of whether studies recruited patients based on baseline adherence (i.e. only recruited nonadherent patients) and for the item of whether results are reported based on baseline adherence levels, the scale values of 1 and 2 were lumped into a score of 1 and compared to the 0's in this analysis. For the other scale item of recruiting a representative patient sample, the 0 and 1 scores were lumped into a 0 score and the 2 score became a score 1, as there were few 0's in this aspect and we aimed to make the sample size in the two scores more comparable through this dichotomization.

3.6 Exploratory Analyses

The impact of a composite adherence measure score was tested by investigating the overall correlation of the quality of the measures to the adherence results for both binary and continuous data. The addition of a composite measure, wherein a total highest score was assigned to each study if the highest score in each quality scale category was not already in a single measure, altered the correlation between the quality and the precision slightly, but the correlations remained not statistically significant. The new Spearman's rho became 0.546 (95% CIs 0.232 to 0.757, p=0.002), a slight decrease in the positive correlation found without the inclusion of composite quality scores. The new Spearman's rho for binary data, between the quality and the proportion, also changed slightly to -0.162 (95% CIs -0.383 to 0.077, p=0.183). These slight changes in the correlation for continuous and binary data indicate that the addition of a composite measure of adherence does not substantively alter the relationship between the quality of a measure and the adherence outcomes in a trial.

The effect size in the two quality score groups for the patient recruitment methods questions was investigated in another exploratory analysis. One of the six comparisons between the effect sizes for dichotomized high versus low quality scores (1 vs. 0) for continuous and binary data were statistically significant ($p<0.05$) (Table 20-21). It was hypothesized that the effect size in groups that recruited nonadherent patients (high quality, score of 1) would be higher than that of studies that did not recruit

patients based on their baseline adherence rates. This was because the difference between the mean in the intervention group and the mean in the control group would be greater if baseline adherence was lower. This was not supported through this investigation of the effect size between these two groups, with studies recruiting nonadherent patients having a median effect size of 0.278 (IQR 0.353) and studies that did not include this eligibility criterion had a median effect size of 0.370 (IQR 0.626) for continuous data. This hypothesis was also not supported for binary data, with studies recruiting nonadherent patients having a median effect size of 1.508 (IQR 3.177), and studies not recruiting based on baseline adherence having a mean effect size of 2.434 (IQR 3.614). However, this difference was statistically significant for binary data; yet this analysis was exploratory and thus no conclusions may be drawn from this finding. No hypotheses were made for the other two quality scale aspects.

Different minimally important differences were tested to determine the effect of choosing a difference other than 25%, on the power of the trial to detect an effect for the patient recruitment questions. Using a 15% minimally important difference in adherence resulted in not meeting the required sample size for a study to be powered much more frequently and sample sizes needed to be moderate to large more frequently. A 35% minimally important difference in adherence, conversely, resulted in much smaller required sample sizes and therefore studies were adequately powered more frequently. The distribution of results for the percent recruitment for each minimally important difference is illustrated in Figures 3-5. I believe a 25% minimally

important difference is the most appropriate as the other two values test, 15% and 35%, resulted in extremes of sample size, compared to the moderate 25% difference that has been noted in the literature previously.

The quality of the methods in these trials was correlated with the year of the trial's publication to determine if there was an association between quality and time. Figure 6 illustrates this relationship of quality of measurement of adherence over time, based on publication year. The Spearman correlation of measurement quality and time was -0.141 (95% CIs -0.338 to 0.068). Figure 7 illustrates the relationship of quality of patient recruitment methods over time. The Spearman correlation of patient recruitment quality and time was -0.022 (95% CIs -0.244 to 0.202). Neither correlation was statistically significant, thus this relationship remains inconclusive. Interestingly, the correlation between the quality of the methods and time (year of publication) was slightly negative, contrary to the hypothesis that quality might increase over time. Though the reason for this result is unclear, it is important to keep in mind this analysis was exploratory and inconclusive.

4.0 Discussion

4.1 Summary of findings

This thesis investigated the types and quality of methodology used in trials of interventions to improve patient adherence to medication. A scale to assess the quality of the methods used in these trials was developed here to allow for a descriptive analysis of the quality of the methods. The types of measures of adherence discussed in the literature and used in trials of interventions to improve adherence ranged from poor quality, such as attendance, to high quality, such as pharmacy refill records. Lower quality measures of adherence were used less frequently in the literature than those with overall higher median quality scores, such as pharmacy refill records and MEMS. Over half of included studies used more than one measure of adherence, which might explain the high frequency of self-report use in these trials, as self-reports are of lower quality but were the second most frequently used measure. In terms of the patient recruitment methodology, the overall quality was low as few studies completed steps that we hypothesized to be important to detect an effect from the intervention. For the secondary research questions involving analyses of the relationship between the quality of the methods and the results of the trials, all results were not statistically significant. The quality of measures correlated positively and moderately for most measures with the precision of adherence results. The quality of measures correlated slightly negatively for most measures with the proportion adherent. The hypothesis for patient recruitment methodology was also supported through our analysis and this result was also not statistically significant. A larger sample size may have allowed us to detect an effect in the patient recruitment methodology questions. Few trials reported adherence results

continuously, despite a recent meta-analysis finding that reporting continuous adherence data compared to reporting binary data was significantly related to the adherence outcome effect[1].

4.2 Interpretation of results

The methodological aspects of each measure of adherence were summarized here and it is thought that Table 4 can be used by future trialists to determine which measure(s) of adherence best suit their trial, while considering key methodological features such as validity and objectivity. Though it is difficult to create a hierarchy of these measures as different trials prioritize different aspects of a measure of adherence, researchers may consult this table as well as the scale created here to maximize the quality of the measures they use. Based on the overview in this thesis, some methods of measuring adherence have been invalidated, are subjective, and are obtrusive to patients, and therefore should not be used or only be used in combination with at least one other, higher quality measurement of adherence to increase the overall quality of the measurement. Many methods have been validated, though such validations are limited without a true gold standard, thus the validation column in Table 4 must be interpreted with caution. Though few measures were objective, researchers can improve the quality of a subjective measure by blinding outcome assessors to the patients' treatment allocation to minimize bias, as the Hawthorne effect may impact the

accuracy of trial results. Many measures may be done unobtrusively, though some of the measures that were considered to be the most accurate such as direct observation and MEMS are obtrusive to patients, which can bias results. Last, most measures were longitudinal in their measurement, and those that only measure the past week of adherence may become longitudinal by administering serial measurements. This, however, may be infeasible. Overall, some measures were better in certain categories of quality than others, therefore composite measures of adherence with complementary strengths might provide the best estimate of adherence.

Similarly to the overview of the literature on measures of adherence, multiple types of measures were used in the sample of 50 randomized controlled trials of interventions to improve adherence. These range from poor quality such as clinician judgments to high quality such as pharmacy refill records. Given the low quality of adherence measurement in some studies, it would appear not all researchers are currently aware of the importance of this aspect of methodology in adherence trials. On the other hand, some trials used three to six measures of adherence, as they noted the importance of composite outcomes in establishing a fuller, more detailed picture of adherence. It is also interesting to note that some of the best, yet most difficult to implement measures, such as MEMS and pharmacy refill records, were used frequently in trials. Though these measures may be of high quality based on this quality scale, the translatability of these measures into clinical practice may be an important feature of a measure that was not assessed in this thesis.

Based on the results of this analytic survey, the quality of measures of adherence appears to relate to the adherence results in these trials. This finding, though exploratory, further supports the importance of this methodological aspect in trials. Only one correlation was statistically significant, but most measures for which this relationship was calculated upheld the pre-specified hypothesis.

It may be important to establish clearer guidelines for selecting measures of adherence in these trials, given the potential relationship between the quality of the measure of adherence and the outcomes. As there is currently no gold standard for measuring adherence, most trials should use composite measures of adherence so that one measure is able to fill the gaps in quality that another may leave. This might be accomplished by first selecting measures that are appropriate for the trial design, then assessing the quality score of each measure and combining those scores until a composite score of 9/9 is reached. For example, a trial on adherence to statins in patients with high cholesterol might select a pill count as their primary measure of adherence as it is able to provide longitudinal, objective data that can be obtained unobtrusively if the pill count is done once at the patients' homes. In order to fill in the gaps in quality in the reliability and validity of this measure, a researcher could add a validated self-report interview as a secondary measure of adherence to reach a composite quality score of 9/9. Though this may be impractical in a clinical setting, this approach may be feasible and appropriate for a research setting. Using this approach, future trials may achieve a more representative adherence result and therefore be

better able to assess the true potency of the intervention being tested. Future trials might also determine if there is a quality score threshold, wherein 7 or 8 out of 9 may be adequate to ensure accurate adherence results.

In the sample of 50 studies, the patient recruitment methodological quality also varied considerably. Few studies used baseline adherence rates as an inclusion criterion for participants, despite the potential importance of this step in establishing the true effect of an intervention. Logically, recruiting either just adherent or both nonadherent and adherent patients makes it more difficult to detect an increase in adherence resultant of an intervention. The exploratory analysis was underpowered and did not establish whether this methodological step did correlate with the power of the study or not. An exploratory analysis of the effect size for high and low quality studies did find that studies that recruited nonadherent patients had a smaller effect size than studies that did not recruit patients based on their adherence levels. This analysis was exploratory and thus even the statistically significant difference found for this comparison may have arisen by chance. This was true for the other two methodological aspects, of whether a representative sample was recruited and whether results were reported based on baseline adherence, in that no analysis detected a relationship between the quality and percent recruitment. Though questions remain as to whether patient recruitment methods might impact adherence results or the power of the study to detect an effect, it is clear researchers do not often consider these aspects of methodology in their trials, as demonstrated by the low overall median quality score.

4.3 Comparison to past studies

Over 18 studies in the past have discussed the quality of methodology in interventions to improve adherence trials, though no articles that investigated the quality of patient related methods or the impact of methods on the results found in these studies[1,2,4,7,14–24,31,42]. The information from these studies has been incorporated into the scale development and Table 4, the summary of adherence measures.

4.4 Limitations

This study is limited by several factors, including the sample size, the fact the scale developed was not validated nor tested for reliability before implementation, and the selection and nature of these intervention trials.

The sample size for this analytic survey was chosen for feasibility reasons, thus the sample size chosen did not allow for a highly powered analysis. However, the primary goal of this thesis was not to determine the correlation between quality and adherence outcomes, but was to provide an overview of the state of the trial methodology, which was accomplished. The analyses were secondary and exploratory, thus the fact the sample size was small and underpowered is not considered to be a major limitation to this study. Due to sample size limitations correlations for several measures of adherence were not calculated. Some measures of adherence such as

pharmacodynamic measures (i.e. International Normalized Ratio, glucose), were also not included in the review of measures because they were not included in the sample of trials included in this study. This limitation of the analysis may be overcome by increasing the sample size.

The scale that was developed in this study to assess the quality of the methodology was not assessed for validity or reliability prior to applying it to these studies. The online data extraction form was extensively tested, including the scale questions, thus the fair inter-rater reliability, calculated via a kappa, was surprising. Further, the validity of this scale was not assessed, which may be an important step in determining the accuracy of these results. Not only was the validity of the scale not assessed, but the scale was based on a semi-systematic review of related literature. It was determined that conducting a systematic review to ensure all papers on this topic are included in this table for scale development was unnecessary. This was because the purpose of this thesis was not to perform a systematic review on the aspects of the quality of the methods, but to provide some basis for the quality scale developed for this analytic survey. Further, the application of this scale to each trial took the information provided by the authors at face value. That is, a measure was considered reliable in this scale if the author of the trial noted that measure was reliable. This was done to allow the scale to be applied to each trial consistently, rather than searching or insisting on hard evidence of each quality aspect in a measure. The validity of this scale was supported qualitatively by comparing it to narrative assessments of measures of

adherence from the literature. However, it is possible this scale did not include some aspects of quality that may be important in distinguishing the quality of different measures. For example, though pharmacy refill occasionally received a perfect quality score, it is indirect and this may be a limitation of the pharmacy record that should be accounted for in its quality. Also, feasibility of use in research or clinic settings could be taken into account directly in this scale. The scale could have been developed using item generation involving a Delphi method with experts, which may have led to better confidence in the validity of this scale. The quality scale may also be insensitive in detecting differences in qualities for a given type of measure (for example, home versus clinic pill counts). This may be overcome by adding more dimensions to our quality scale and therefore increasing the discrimination between good and bad quality methods of implementation of the same measure of adherence.

Poor reporting in trials and incomplete reporting may also limit or bias the results of this thesis. Some trials did not report adherence data per measure of adherence used, thus some measures of adherence included in the overall quality assessment could not be included in the secondary analysis of the data. Completed extraction forms were sent to all authors in the hopes they would provide further data to reduce this limitation, but no authors have responded with more outcomes data. Further, poor reporting could have affected the analyses concerning patient recruitment, with an average of 16.5/50 (33%) having unclear ratings across the three

scale items. Therefore the quality of the methods may have been rated lower overall than if studies had appropriately reported their methods.

The method used to select the trials that were included in this thesis could have biased the results, as trials were not randomly selected. Trials were included in this study based on the type of measure of adherence used in the trial, selecting them purposively. This approach may have been biased, as the full set of articles included in the separate systematic review had not been identified and extracted at that point. Though a stratified random sample, stratifying studies by the type of measure used in them, may have been the best approach with the least selection bias, this approach was not possible given the stage of the full review, and an inability to stratify all studies. Further, only trials that tested interventions to improve adherence were included in this study, which also might limit the comprehensiveness of the types of measures of adherence included in this overview. For example, pharmacodynamic measures of adherence are often not explicitly stated to be a measure of adherence, and thus would not have been included in this study. Other measures of adherence might have been included had the inclusion criteria for this review involved other types of studies that tested ways to improve patient outcomes as well. However, including these trials would have become unfeasible, given the large quantity of such trials, thus limiting the sample to trials focused on adherence was appropriate.

Last, the nature of these intervention trials to improve adherence makes a quantitative analysis of their results complicated due to the diversity of outcomes

reported and multiple confounding factors that would need to be controlled for. A correlation between adherence and patient outcomes was originally planned, as it was thought that this correlation would be stronger in trials with better quality methodology. Unfortunately, due to the variability in reporting and diversity of disease conditions, this analysis was not possible. Even with the analysis as it is, it is possible the quality of the methods is confounded by an overall poor quality of a trial.

4.5 Future Directions

Many trials in the past have lamented the lack of a gold standard for measuring adherence as well as the ineffectiveness of tested interventions. Though this study has taken steps to address those issues, the limitations listed previously call into question the validity of these results. In order to improve the reliability of these results, there are steps that may be taken in the future. One such future research direction that would improve the validity of this study would be to first improve the reliability of the scale we created here, perhaps by rewording each scale item for better clarity. Once an acceptable level of reliability is found, the validity of this scale can then be tested by comparing it to similar scales or having experts in this field assess the validity.

Another useful next step in this area would be to come to a consensus among adherence experts on a gold standard for measuring adherence. This gold standard could either be a new measurement method, such as the emerging “adherence

necklace"[43], an existing method, or a combination of existing methods. If experts were able to deem a measure of adherence a gold standard, this could ensure future trials report the most accurate and precise results. Failing that, establishing a quality threshold would be useful to ensure a single or composite adherence measure will report accurate results.

The methodological steps outlined in this thesis should be applied to all future adherence trials to ensure the most accurate results are found. This may reduce the confounding across trials and reduce the variability in efficacy of a given intervention. This may then reduce the heterogeneity across trials to allow for future meta-analyses of these studies. Other steps that could be taken by future trials to allow for meta-analysis include standardizing other aspects of trials in addition to these two methodological areas. Standardizing the reporting of results, both clinical and adherence outcomes, may increase the likelihood that studies can be pooled. However, even with these steps, future meta-analyses may be impossible given the large variability in diseases and intervention types.

5.0 Conclusions

The state of the methodology in trials to improve patient adherence to medication has been described here, in terms of the measurement of adherence and the patient recruitment methods. The qualities of these methods have been quantified

through a scale developed here for this purpose, and how this quality relates to the adherence outcomes from each study was investigated. The lack of a gold standard for measuring adherence is a major issue that needs to be addressed in future research. The lack of significant results in these provisional analyses may be mainly related to our limited sample size, so this topic deserves further exploration with a larger sample of adherence intervention trials.

6.0 Tables and Figures

Table 1: Desirable features of measures of adherence and currently available measures that fulfill that feature

Desirable features	Definition of feature	Current measures that correspond with this definition
Reliability	Reflects the amount of variation in repeated measures of the same adherence state[25]	-MEMS* -biologic measure, such as concentration of drug in blood
Validity	The extent to which a measure assesses adherence compared with a "gold standard" or criterion of adherence [25]	-MEMS -validated questionnaires
Objectivity	The extent to which a measure requires little judgment on the part of the person measuring (ie, less prone to experimenter bias)	-biologic measure (blood/ urine) -pharmacy fill record
Unobtrusiveness	The extent to which the measure is unapparent to the patient, minimizing risk of bias (obtrusive measures can cause the Hawthorne effect)[14]	-pharmacy fill record -physician assessment -therapeutic response -attendance at appointments
Longitudinality	The extent to which a measure provides adherence data over a period of time	-MEMS -pharmacy fill record

* MEMS= medication event monitoring system

Table 2: Categorization of features of adherence measures (reliability, validity, objectivity, unobtrusiveness, longitudinality) into scale scores based on whether the feature is absent (No), uncertainly absent or present (Uncertain), or present (Yes)

Feature	Score 0 (No)	Score 1 (Uncertain)	Score 2 (Yes)
Reliability[38]	Documented not to be reliable	Reliability not assessed	Measure documented be reliable
Validity[38]	Documented not to be valid	Validity not assessed	Measure documented to be valid in comparison with a

			criterion standard
Objectivity[7,44]	Subjective measure without appropriate blinding to patients' treatment group (appropriate blinding = method of blinding stated and blinding of patient and assessor, or blinding of assessor when impossible to blind patients)	Subjective measure with uncertain blinding (method or blinded group not explicitly stated)	Objective measure (MEMS, pharmacy refill data, biologic measure of drug[45]) or subjective measure with method of blinding explained and includes appropriate blinding
Unobtrusiveness	Obtrusive to patient leading to potential Hawthorne effect[46] (eg, electronic monitoring[14,45])	Unclear whether the patient is aware adherence is being measured or the extent to which the measure would interfere with their usual medication consumption	Patient is unaware the measure is being taken and the measure does not interrupt the normal pattern of medication consumption (pharmacy refill record[14])
Longitudinality[47]	Data provided by measure covers the past 1-7 days of adherence for a chronic (long term) regimen	Data by measure covers a longer period of time (>7 days) for a chronic medication regimen	

Table 3: Categorization of features of patient recruitment methods into scale scores based on whether that feature is absent (no), uncertainly absent or present (Uncertain), and present (Yes)

Feature	Score 0 (No)	Score 1 (Uncertain)	Score 2 (Yes)
Nonadherent patients selected[48]	0- No (no mention of past adherence in any capacity, assume both adherent and nonadherent patients recruited)	1- Indeterminate (eg, only included patients with high blood pressure or patients who had that physiologic state possibly due to a lack of adherence in the past but not explicitly stated that adherence was measured prior to selection)	2- Yes (adherence was measured prior to study inclusion and only those with low adherence were included in the study, or those with low adherence to begin with are separately reported)
Representativeness of sample (= # recruited/ # screened who met eligibility criteria)	0- Number of patients asked to join is much higher than sample size (>2:1 non-enrolled to enrolled[18])	1- Does not report number of patients asked before reaching sample size	2- The number of patients asked to join is similar to sample size ($\leq$2:1 non-enrolled to enrolled)

Results reported based on baseline adherence	0- Baseline adherence was not measured	1- Baseline adherence measured but results were not reported according to initial adherence level.	2- Yes, results were reported based on baseline adherence level or if only non-adherent patients are recruited, if intention to treat analysis is followed

Table 4: Overview of desirable features of adherence measurement methods based on a review of the literature and their advantages and disadvantages

Measure/ test	Advantages	Disadvantages	Reliability	Validity	Objectivity	Unobtrusiveness	Longitudinality
Direct observation	- most direct	- unfeasible in most settings[49] - patients can still not swallow the pills - need to check "cheek pouches"	- reliable	- valid	- direct (know patient takes medicine)[50] - subjective (the observer must make a judgment and could alter the judgment according to the patient and other influences, eg, intervention group status)	- most obtrusive measure (and often intended to enhance adherence)	- could be long term - depends on period of observation
Pill count techniques							
Therapeutic drug monitoring (MEMS-medication event monitoring system)	- seen as the gold standard [37,51] - only method able to provide adherence data over time without inconvenience to outpatients (time series)[52] - using bottle may be incentive to increase adherence - detects white coat compliance[53] - more feasible than unannounced pill counts[53]	- Hawthorne effect - high cost, proprietary software[55] - not accepted by some patients, difficult to use, faulty at times[49] - indirect method - ingestion cannot be confirmed based on data, does not tell how much med taken at one time[49] - requires that patient return the pill container for downloading	- "most reliable"[56,57]	- valid[53,58] - gold standard to validate other methods[37] - eye dropper Travatan Dosing Aid (TDA) validated[59] - Smartinhaler (metered-dose inhaler (MDI))[60]	- objective[63,64] - indirect	- noninvasive though patients may feel "observed"[56] - bottle / system often bulky and its purpose is inferred by patients	- long term

Measure/ test	Advantages	Disadvantages	Reliability	Validity	Objectivity	Unobtrusiveness	Longitudinality
	- generally acceptable [53] - considered tamper-proof to some[54] (though patients can still pocket dose or open without consuming) - difficult to "consistently bias the recordings"[52]	adherence data (if not automatically transmitted)		- Informedix Med-eMonitor System[61] -DMAS device found to be feasible and correlated with MEMS[62]			
IDAS (Intelligent Drug Admin-istration System) II [56]	- some patients preferred this to MEMS in a trial [56]	- Hawthorne effect	- uncertain[56]	- uncertain	- objective - indirect	- patients may feel "observed"[56]	- long term
Pharmacy refill record/ medication possession ratio (MPR)	- collecting from multiple pharmacies is feasible[65] - repeat refilling over time is a good proxy for good adherence[66] - important measure, should be used in clinical practice[67]	- overestimates nonadherent if do not collect pill count in addition to pharmacy records [50,65] - may differ from pill count [68] - potential for documentation errors[49]	- more reliable than self-report [69,70]	- generally valid - New Prescription Medication Gaps (NPMG) has been validated[40,71]	- indirect[49] - objective	- noninvasive[49]	- long term data[49]
Pill count at clinic visit	- more accurate than self-report[72] - gives quantitative indication of adherence	- tedious and difficult to administer, laborious, indirect (do not know if medication removed from pill bottle is taken)[68] - pill counts at patient visits require patient cooperation: attend the visit, bring all the pills, don't 'fix' the count	- easy for patient to alter medication, patients who are nonadherent often "forget" to bring medication to visit or don't attend	- valid compared to MEMS, correlated[29]	- objective measure[73,74] - indirect	- patient cannot know the purpose of pill count (though often purpose is inferred) for this to be unobtrusive	- long term

Measure/ test	Advantages	Disadvantages	Reliability	Validity	Objectivity	Unobtrusiveness	Longitudinality
		- easy to bias data[52]					
Pill count at home	- more accurate than self-report - well correlated to drug concentration in body if single pill count at patient's home with purpose not revealed ahead of time [54]	- pill counts at home are costly and logistically difficult[54,74]	- unannounced pill counts are reliable when done in the home or on the telephone[73,74]	-unannounced pill counts are valid when done in the home or on the telephone[73,74]	- objective measure[73,74] Indirect	- must be unannounced or patient cannot know the purpose of pill count (though often purpose is inferred)	- long term
Body drug concentration							
Pharmacodynamic monitoring (INR)[51]	- can provide an accurate picture of adherence if patient variables are taken into consideration	- influenced by many patient variables that may or may not be taken into account in the measurement	- depends on measure and laboratory[49]	- depends on measure and laboratory[49]	- direct[50] - objective	- unobtrusive as usual care generally involves such samples	- long term (>1 week)
Body fluid drug/ metabolite concentration for drugs with a long half life (>12hr) (eg. digoxin, epilepsy medicines (eg, Dilantin, phenytoin, phenobarbital), lithium, etc.[75]	- long half-life drugs can reach a steady state in the body and can demonstrate if the patient has been adherent over the course of several days[52,75] - can add small sub-therapeutic amount of these to short acting drugs to determine compliance[49,52,76]	- affected by lab errors, time of sampling and pharmacodynamics/ metabolism[49] - easily manipulated with single doses	- depends on measure and laboratory[49]	- depends on measure and laboratory[49]	- direct[50] - objective	- unobtrusive as usual care generally involves such samples	- long term (>1 week)

Drugs with a short half life	- serial extractions provide longitudinal data - drugs with short half-life give "snap shot" for the prior few hours of adherence[49,52] - average for prior period of detectable drug up to 6 times the half-life[75] (for example an assay of a drug with a half-life of 3 hours would indicate adherence for the prior 18 hours)	- affected by lab errors and pharmacodynamics/ metabolism[49] - white coat compliance can reduce reliability[58] - only demonstrates adherence for the previous few hours to day	- depends on measure and laboratory[2]		- direct[50] - objective	- may be unobtrusive if usual care involves taking such samples	- short half-life drugs may provide long-term assessment of adherence if measurements are done serially
Self-report							
Self-report, summary[77]	- easy, simple, economical, revealing of patient nonadherence reasons [78] - most feasible[70] - often can involve simple question of "How many doses of the drug did you miss since the last time we met?"[79] - could be done using a VAS[33] - nonjudgmental question- "People often have difficulty taking their pills for one reason or another. How many times do you think you may have missed taking your pills in the last week?"[80]	- biased by social desirability[52,81] - often overestimates adherence, compared to other higher quality measures such as pharmacy refill records	- some self-report methods are reliable - reliability lower due to recall errors and social desirability bias[73]	- some self-report methods are valid - others need to be validated[82] - VAS scale may be more reliable	- subjective	- minimally invasive, yet obvious to the patient the measure is being taken[73] - computer interview/ questionnaire are less obtrusive[83]	- longer recall periods result in larger non-adherence rates[81] - inherent error in human memory[84] - some are 7 day recalll no difference between 3, 4, and 7 day recall [78,85]

Measure/ test	Advantages	Disadvantages	Reliability	Validity	Objectivity	Unobtrusiveness	Longitudinality
Simplified Medication Adherence Questionnaire (SMAQ) [86]	- validated, very simple, economical, and easy to apply - can be used semi-quantitatively by assigning a percentage of adherence - sensitivity 0.91, specificity 0.72, positive likelihood ratio 7.94 (95% CI 6.44, 9.45)[87] with MEMS as the gold standard	- tends to overestimate adherence	- reliable	- valid	- subjective	- obtrusive	- usually only covers past 3 days
Haynes-Sackett [86,88–90]	- valid method is one of the most simple to use in clinical practice[86] - brief, economical[86] - sensitivity 0.96, specificity 0.5, positive likelihood ratio = 1.92 (95% CIs , 1.45, 2.49)[89]	- overestimates adherence	- reliable	- valid[86,89]	- subjective	- obtrusive	- usually only covers past 3 days
Morisky-Green [86,91,92] **(MMAM/ MGT)**	- brief and very easy to apply - validated and applied in numerous pathologies - can provide information about reasons for non-adherence - Economical - 0.81 sensitivity (0.61 in one trial[93]), 0.44 specificity, LR+= 1.43 (95% CIs 1.12, 1.84)[91]	- overestimates adherence - may have poor psychometric properties (acceptability)[94] - invalid for patients with IBS (irritable bowel syndrome)[95]	- reliable in Taiwan setting[96]	- validated in Taiwan setting[96]	- subjective	- obtrusive	- usually only covers past 3 days

Measure/ test	Advantages	Disadvantages	Reliability	Validity	Objectivity	Unobtrusiveness	Longitudinality
Brief Medication Questionnaire (BMQ) [86,97]	- self-applied test - not too long and allows for analysis of adherence and barriers to adherence - high sensitivity and can be used to validate other tests specificity 0.8, sensitivity 1, positive likelihood ratio 5.0 (95% CIs 0.87, 28.9) → small sample size used for validation[98]	- presents a complete validity procedure	- reliable	- good predictive validity[97]-	- subjective	- obtrusive	- usually only covers past 3 days
Medication Adherence Self-report Inventory (MASRI)[99]	- best when combined with MD scale (physician assessment)[99] - sensitivity 0.68, specificity 0.67, +LR 2.1 (95% CIs 1.1, 4.3)[99]	- specific to lupus [99]	- reliable [99]	- valid [99]	- subjective	- obtrusive	- usually only covers past 3 days
Rating of Medication Influences (ROMI) scale [86,100]	- validates patient attitudes about medication - Cronbach's alpha 0.41-0.57 (no values for sensitivity/ specificity reported)[101]	- specifically for schizophrenia - excessively long scale	-reliable[100]	- valid[100]	- subjective	- obtrusive	- usually only covers past 3 days
Medication adherence self-efficacy scale (MASES) [102–104]	- simple / fast - captures useful data on adherence self efficacy[104] - disease specific - hypertension and glaucoma have been validated[103,104] - factor loadings (validity) 0.26-0.72, no sensitivity /	- only 2 diseases have been validated	- reliable[102,103]	- use in hypertension or glaucoma validated[102,103]	- subjective	- obtrusive	- usually only covers past 3 days

Measure/ test	Advantages	Disadvantages	Reliability	Validity	Objectivity	Unobtrusiveness	Longitudinality
	specificity reported						
Maastricht Utrecht Adherence in Hypertension (MUAH) questionnaire [86,105]	- self-applied test validated for hypertension - presents excellent psychometric properties and can he used to identify factors that provide barriers to or facilitate adherence - Cronbach's alpha 0.63-0.8, no sensitivity/ specificity reported[105]	- specific to hypertension - excessively long, requires awareness of adherence phenomenon that may not be universal (known by all participants)	- reliable[105]	- valid[105]	- subjective	- obtrusive	- usually only covers past 3 days
European Heart Failure Self-care Behaviour Scale (EHFScBS)[106]	- reliability score 0.8 [107]	- item total correlation ranged from 0.14-0.65 for construct validity [107] - appears invalid based on values[107]	- good psychometric properties[106]	- not validated	- subjective	- obtrusive	- usually only covers past 3 days
Hill Bone medication adherence (HBMA)[108]	- simple, easy to apply, economical	- only validated in Korean high blood pressure population (Cronbach's alpha >0.4 - no sensitivity or specificity)[108] - may have poor acceptability and insufficiency, though only one trial[94]	- reliable in Korean setting[108]	- valid in Korean setting[108]	- subjective	- obtrusive	- usually only covers past 3 days
Medication adherence report scale (MARS) [86,109,110]	- self-applied test - used in chronic disorders (COPD) - validated, significantly correlated with other self-report measures[109]	- excessively long, thus some patients do not retake the test or do not complete the test - no sensitivity/ specificity reported	- reliable[109]	- valid [109]	- subjective but well correlated to objective measures[97]	- obtrusive	- usually only covers past 3 days
Herz 5-point	- simple to use, 5-point scale	- likely overestimates	- "high degree of	- not tested	- subjective	- obtrusive	- only covers

Measure/ test	Advantages	Disadvantages	Reliability	Validity	Objectivity	Unobtrusiveness	Longitudinality
scale[111,112]	(1 = always and 5 = never)	compliance as noncompliance was conservatively defined as "never taking medication"	reliability"[111]				past 3 days
Measurement of Treatment Adherence (MTA) for Brazilian patients[113]	- simple - dichotomous scale once scores added, 6 items[113]	- biased by social desirability[52,81] - often overestimates adherence	- weak evidence of reliability in this population[113]	- weak evidence of validity in this population[113]	- subjective	- obtrusive	- usually only covers past 3 days
Adult AIDS Clinical Trials Group studies (AACTG) Adherence Instrument[62,114,115]	- AIDS specific - simple, feasible[115]	- may not be generalizable to non USA populations[115]	- not tested	- not tested	- subjective	- obtrusive yet feasible[115]	- covers past 2-4 days[115]
Bogus-Pipeline[86]	- very simple, economical brief and easy to use	- tends to overestimate adherence - not validated	- not tested	- not tested	- subjective	- obtrusive	- usually only covers past 3 days
Prochasca-Diclemente[86]	- brief, comprehensive, economical, and very easy to apply - applicable to chronic disorders	- tends to overestimate adherence - not validated	- not tested	- not tested	- subjective	- obtrusive	- usually only covers past 3 days
Hermes[86]	- brief, comprehensive, economical, and very easy to apply - applicable to any disorder	- tends to overestimate adherence - not validated	- not tested	- not tested	- subjective	- obtrusive	- usually only covers past 3 days
Herrera Caranza[86]	- brief, comprehensive, economical, and very easy to apply	- can overestimate adherence -not validated	- not tested	- not tested	- subjective	- obtrusive	- usually only covers past 3 days

Measure/ test	Advantages	Disadvantages	Reliability	Validity	Objectivity	Unobtrusiveness	Longitudinality
	- applicable to any disorder						
Drug Attitude Inventory (DAI)-30 [86]	- self-applied scale with true/ false questions	- specifically for schizophrenia - does not cover patient motivation for taking or not taking medication - extremely long	- not tested	- not tested	- subjective	- obtrusive	- usually only covers past 3 days
Drug Attitude Inventory (DAI)-10 [86]	- self-applied scale with true/ false questions	-specifically for schizophrenia - does not cover patient motivation for taking or not taking medication - although it is shorter that the DAI-30, it is still longer than other self reports	- not tested	- not tested	- subjective	- obtrusive	- usually only covers past 3 days
Neuroleptic Dysphoria (ND) [86]	- brief, economical and very simple to apply - possible predictor of immediate and long-term adherence	- specifically for schizophrenia - does not cover patient motivation for taking or not taking medication	- not tested	- not tested	- subjective	- obtrusive	- usually only covers past 3 days
Multidimensional Adherence Classification System (MACS)[116]	- comprehensive method of understanding patient adherence [116] - no values on validity/ reliability provided	- specific to transplant patients[116]	- not tested	- not tested	- subjective	- obtrusive	- usually only covers past 3 days
A-14 adherence questionnaire [63]	- shows good psychometric properties[63]	- requires validation study - study on psychometric properties flawed, not compared to gold standard, 50% response[63]	- not tested	- not tested	- subjective	- obtrusive	- usually only covers past 3 days
Adherence Self-report	- easy, simple, economical, revealing of patient	- biased by social desirability[52,81]	- not reliable	- invalid based on	- subjective	- obtrusive	- usually only covers past 3

Measure/ test	Advantages	Disadvantages	Reliability	Validity	Objectivity	Unobtrusiveness	Longitudinality
Questionnaire (ASRQ)[55]	nonadherence reasons [78] - most feasible[70]	- often overestimates adherence		comparison to MEMS[55]			days
Patient diary Ex. Computer based[117]	- more long term information than self-report	- may forget to enter information - may enter wrong information - desirability	- more reliable than self-report - theoretically highly reliable, not tested	- uncertain	- subjective	- obtrusive	- long term
Other methods of measuring adherence							
Capsule photograph[118]	- less expensive than MEMS as cell phones are ubiquitous[118]	- more work for the patient, they must remember to take photo before each pill consumed for this to be accurate[118] - underestimates adherence[118]	- seems reliable compared to MEMS, but only one reliability study[118]	- uncertain	- indirect measure and objective	- obtrusive	- long term- depends on instructions
Adherence necklace[43]	- essentially foolproof method of measuring adherence, wherein the patient wears a "necklace" (monitoring device worn around the throat) that tracks each time a pill is swallowed, by manipulating the prescribed medication to contain small data chips that the "necklace" counts - direct measure	- very obtrusive to patients - likely not feasible in most settings	- uncertain	- uncertain	- objective and direct measure	- obtrusive	- long term, similar to MEMS
Therapeutic outcomes	- adherence rates correlated to improved therapeutic outcomes[50] - therapeutic outcomes can include anything from blood	- variable reliability and accuracy due to lab/clinical measurement errors	- depends on measure- blood pressure less reliable than viral load	- depends on specific measure	- may be subjective or objective depending on outcome	- may be unobtrusive- depends on measure	- long term

Measure/ test	Advantages	Disadvantages	Reliability	Validity	Objectivity	Unobtrusiveness	Longitudinality
	pressure to viral load				measure used		
Attendance	- easy, readily available data	- studies have shown appointment keeping not well correlated with medication adherence (for certain conditions)[1]	- uncertain	- attendance not a good overall adherence indicator[12]	- objective	- unobtrusive if patient not told measure taken	- long term generally
Physician assessment	- convenient; includes all medication types	- overly rely on patient outcomes as indicators of adherence [49] - 10% specificity[54] - biased[1,52]	- uncertain	- invalid - overestimates adherence[40]	- subjective	- noninvasive	- long term

Table 4 includes jargon defined here: *Sensitivity*= proportion "compliant" correctly identified, 80% cut point for adherence/nonadherence dichotomy[89]; *Specificity* = proportion "noncompliant" correctly identified[89]; *Likelihood ratio*= (sensitivity)/ (1-specificity); S*ocial desirability bias*= a systematic deviation from the truth caused by the desire to respond in what is thought to be a favourable way

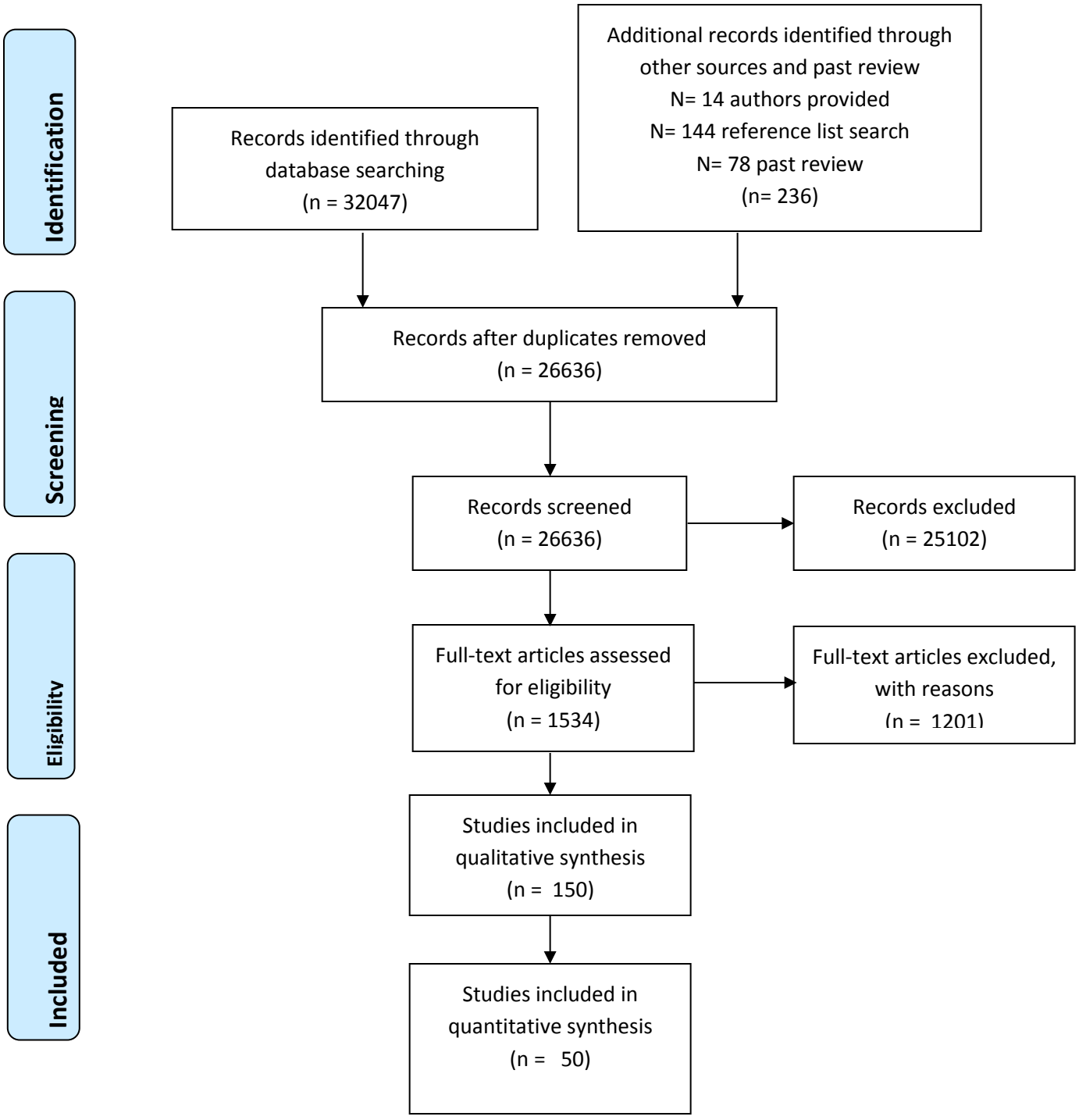


Figure 1: Flow of studies through inclusion process

Table 5: Table of scores of measurement quality of measures of adherence from analytic survey, by study, including score in each feature of measurement quality and overall quality score

Study ID	Year of publication	Measurement method	Validity	Reliability	Objective	Obtrusive	Longitudinal	Total measurement method quality score (0-9)
Abrahams	2010	pill count	1	1	2	0	1	5
Abrahams	2010	self report – diary	1	1	0	0	1	3
Abrahams	2010	self report - interview	1	0	0	0	0	1
Abrahams	2010	attendance at scheduled follow up appointments	1	1	0	1	1	4
Al-Eidan	2002	pill count	1	1	2	1	1	6
Al-Eidan	2002	self report – questionnaire	1	1	0	0	1	3
Andrade	2005	MEMS	2	2	0	0	1	5
Andrade	2005	MEMS	2	2	2	0	1	7
Andrade	2005	self report – questionnaire	2	2	0	0	0	4
Ansah	2001	self report - interview	1	1	0	0	0	2
Baird	1984	drug concentration in body fluid (quantitative)	1	1	2	1	1	6
Baird	1984	pill count	1	1	0	0	1	3
Becker	1986	pill count	1	1	2	2	1	7
Becker	1986	self report - interview	2	2	0	0	1	5
Becker	1986	therapeutic response	1	1	2	1	1	6
Becker	1986	therapeutic response	1	1	2	1	1	6
Becker		composite measure	2	2	2	2	1	9
Berrien	2004	pharmacy refill record	1	1	2	2	1	7
Berrien	2004	self report – questionnaire	1	1	0	0	0	2
Brus	1998	pill count	1	1	2	0	1	5
Chang	2011	pill count	1	1	2	0	1	5
Chang	2011	pill count	1	1	2	0	1	5
Chang	2011	self report – questionnaire	1	1	0	0	0	2
Choudhry	2011	pharmacy refill record	1	1	2	2	1	7
Colcher	1972	drug concentration in body fluid (quantitative)	1	2	2	1	1	7

Cote	1997	pill count	1	1	2	2	1	7
Dilorio	2008	MEMS	2	2	2	0	1	7
Farooq	2011	pill count	1	1	2	0	1	5
Farooq	2011	self report – questionnaire	1	2	1	0	0	4
Farooq	2011	composite measure	1	2	2	0	1	6
Gallefoss	1999	pharmacy refill record	1	1	2	2	1	7
Gallefoss	1999	pharmacy refill record	1	1	2	2	1	7
Ginde	2003	pharmacy refill record	1	1	2	2	1	7
Ginde	2003	self report - interview	0	0	0	0	0	0
Girvin	1999	MEMS	2	2	2	0	1	7
Girvin	1999	MEMS	2	2	2	0	1	7
Girvin	1999	MEMS	2	2	2	0	1	7
Girvin	1999	MEMS	2	2	2	0	1	7
Girvin	1999	pill count	1	1	2	1	1	6
Girvin	1999	composite measure	2	2	2	1	1	8
Haynes	1976	pill count	1	1	2	2	1	7
Hederos	2005	MEMS	1	1	2	1	1	6
Hederos	2005	self report – diary	1	1	0	0	1	3
Hederos	2005	clinician judgment	0	0	0	0	1	1
Hederos	2005	self report – questionnaire	1	1	0	0	1	3
Henry	1999	pill count	1	1	2	0	1	5
Henry	1999	self report - interview	1	1	1	1	1	5
Henry	1999	composite measure	1	1	2	1	1	6
Kalichman	2011	pill count	2	2	2	2	1	9
Katon	2001	pharmacy refill record	2	2	2	2	1	9
Katon	2001	self report - interview	2	2	1	0	0	5
Kimmel	2012	MEMS	1	1	2	0	1	5
Klein	2007	drug concentration in body fluid (quantitative)	1	1	2	2	1	7
Klein	2007	MEMS	1	2	2	0	1	6
Klein	2007	pill count	1	1	2	0	1	5
Klein	2007	self report – questionnaire	1	1	0	0	0	2
Klein	2007	self report – questionnaire	2	2	0	0	0	4
Klein	2007	therapeutic response	1	1	2	2	1	7
Klein	2007	composite measure	2	2	2	2	1	9

Lai	2011	self report – questionnaire	1	1	0	0	0	2
Lai	2011	self report – diary	1	1	0	0	1	3
Lai	2011	pill count	1	1	2	0	1	5
Lai	2011	pharmacy refill record	1	1	2	2	1	7
Lai	2011	therapeutic response	1	1	2	1	1	6
Laporte	2003	MEMS	1	1	2	0	1	5
Laporte	2003	pill count	1	1	0	0	1	3
Matsumura	2012	pill count	1	1	2	0	1	5
Moshkovska	2011	drug concentration in body fluid (quantitative)	1	1	2	1	1	6
Mullan	2009	pharmacy refill record	2	2	2	2	1	9
Mullan	2009	self report - interview	2	2	0	0	0	4
Nieuwkerk	2012	self report – questionnaire	2	2	0	0	0	4
Otsuki	2009	pharmacy refill record	2	2	2	2	1	9
Otsuki	2009	self report – questionnaire	1	1	1	0	1	4
Perrin	2010	MEMS	2	2	2	2	1	9
Perrin	2010	MEMS	2	2	2	2	1	9
Peterson	1984	attendance at scheduled follow up appointments	0	0	0	1	1	2
Peterson	1984	drug concentration in body fluid (quantitative)	1	1	2	2	1	7
Peterson	1984	drug concentration in body fluid (quantitative)	1	1	2	2	1	7
Peterson	1984	drug concentration in body fluid (quantitative)	1	1	2	2	1	7
Peterson	1984	pharmacy refill record	1	1	2	2	1	7
Peveler	1999	MEMS	2	2	2	2	1	9
Peveler	1999	self report – questionnaire	2	2	0	0	1	5
Purcell	2007	MEMS	1	1	2	0	1	5
Purcell	2007	self report – questionnaire	2	2	0	0	0	4
Purcell	2007	composite measure	2	2	2	0	1	7
Rawlings	2003	MEMS	1	1	2	0	1	5
Remien	2005	MEMS	2	2	2	0	1	7
Rickles	2005	pharmacy refill record	2	2	2	2	1	9
Rickles	2005	self report - interview	2	2	1	0	0	5
Sackett	1975	drug concentration in body fluid (quantitative)	1	1	2	2	1	7
Sackett	1975	drug concentration in body fluid (quantitative)	1	1	2	2	1	7

Sackett	1975	pill count	1	1	2	2	1	7
Sarna	2008	direct observation	1	1	2	0	0	4
Sarna	2008	pill count	1	1	2	0	1	5
Sarna	2008	self report – questionnaire	2	2	0	0	0	4
Sarna	2008	composite measure	2	2	2	0	1	7
Sherrard	2009	self report – questionnaire	1	1	0	0	0	2
Sirey	2010	self report – questionnaire	1	2	1	0	0	4
Solomon	2012	pharmacy refill record	2	2	2	2	1	9
Sorensen	2007	MEMS	2	2	2	0	1	7
Sorensen	2007	pill count	1	1	2	0	1	5
Sorensen	2007	self report – questionnaire	1	1	0	0	0	2
Velligan	2008	pharmacy refill record	1	1	2	2	1	7
Velligan	2008	pill count	1	1	2	1	1	6
Villeneuve	2010	pharmacy refill record	2	2	2	2	1	9
Walley	2001	direct observation	1	1	2	0	1	5
Weber	2004	MEMS	1	1	2	0	1	5
Weber	2004	self report – questionnaire	1	1	0	0	0	2
Wilson	2010	pharmacy refill record	2	2	2	2	1	9
Wu	2012	MEMS	2	2	2	1	1	8

Table 6: Measurement qualities of all generic types of adherence measures from a sample of randomized trials of interventions to increase patient adherence

Measure[a]	Frequency of use in sample of RCTs (n)	Median validity score (max 2) and range[b]	Median reliability score (max 2) and range[b]	Median objectivity score (max 2) and range[b]	Median unobtru-siveness score (max 2) and range[b]	Median longitu-dinality score (max 1) and range[b]	Median total score (max 9) and IQR
Attendance	2	0.5 (0-1)	0.5 (0-1)	0	1	1	3 (1)
Clinician judgment	1	0	0	0	0)	1	1 (0)
Direct observation	2	1	1	2	0	0.5 (0-1)	4.5 (0.5)
Drug concentration in body	6	1	1 (1-2)	2	1.5 (1-2)	1	7 (0.75)
MEMS	16	2 (1-2)	2 (1-2)	2 (0-2)	0 (0-2)	1	6.5 (2.25)
Pharmacy refill record	14	1.5 (1-2)	1.5 (1-2)	2	2	1	8 (2)
Pill count	20	1 (1-2)	1 (1-2)	2 (0-2)	0 (0-2)	1	5 (1.25)
Self-report questionnaire	17	1 (1-2)	1 (1—2)	0 (0-1)	0	0 (0-1)	3 (2)
Self-report	3	1	1	0	0	1	3 (0)

diary							
Self-report interview	8	1.5 (0 – 2)	1.5 (0 – 2)	0 (0 – 1)	0 (0 – 1)	0 (0 – 1)	4.5 (3.25)
Therapeutic response	3	1	1	2	1 (1-2)	1	6 (0.5)

a- Measures are not meant to be compared against each other as the nature of each quality category (i.e. validity, etc.) is not comparable between measure types as criterion standards can differ. Some measures are from the same study as some studies used multiple measures of adherence.

b- Range is only included in brackets if a range is possible to include based on the data

Table 7: Number and proportion of studies with one to five measures of adherence in a given study

	One measure of adherence	**Two measures of adherence**	**Three measures of adherence**	**Four measures of adherence**	**Five measures of adherence**	**Six measures of adherence**
Number of studies	23	18	5	2	1	1
Percent of studies (/50)	46%	36%	10%	4%	2%	2%

Table 8: Outcomes for each study, divided into binary and continuous outcomes, for measures of adherence secondary analysis

	CONTINUOUS			**BINARY**
Study ID	Adherence outcome mean control group	Adherence outcome SD control group	Adherence outcome CV = SD/ mean control group	Adherence outcome proportion control group
61894 Villeneuve				0.81
58320 Solomon				0.303
58692 Lai	96.17	10.95	0.113861	
58693 Lai	96.46	10.17	0.105432	
58918 Farooq				0.45
60196 Kalichman	0.66	0.22	0.333333	
60980 Abrahams				0.319
60980 Abrahams				0.516
60980 Abrahams				0.362

61577 Sirey	2.9	1.34	0.462069	0.43
62574 Otsuki	0.53	0.8	1.509434	
62574 Otsuki				0.9496
62874 Mullan				1
62874 Mullan				0.81
64892 Sarna				0.588
64892 Sarna				0.953
64892 Sarna				0.588
65330 Velligan				0.53
65442 DiIorio	0.55	0.364	0.661818	
65867 Purcell				0.859
65867 Purcell				0.859
66971 Sorensen	0.531	0.2922	0.550282	
66971 Sorensen	0.784	0.2835	0.361607	
66971 Sorensen	0.716	0.2554	0.356704	
75329 Perrin	0.767	0.305	0.397653	0.553
75329 Perrin	0.737	0.36	0.488467	0.5
77245 Klein				0.51
77245 Klein	0.808	0.124	0.153465	0.79
77245 Klein	0.972	0.136	0.139918	
77245 Klein				0.75
77245 Klein				0.63
77245 Klein	0.808	0.124	0.153465	0.51
83030 Chang				0.742
83030 Chang				0.979
83030 Chang				0.815
85044 Choudhry				0.089
86690 Nieuwkerk	8.86	1.48	0.167043	
86695 Matsumura				0.98
86842 Hederos	72			0.7
86842 Hederos				0.85
86844 Andrade	0.64	0.082	0.128125	
86847 Rickles	48.6	39.2	0.806584	

86850 Remien	0.6	0.34		0.39
86856 Berrien	31.9	3.873	0.121411	
86859 Weber	0.889	0.04	0.044994	0.5
86864 Rawlings				0.74
86866 Ginde				0.742
86866 Ginde				0.968
86868 Laporte	0.807	0.194	0.240397	
86868 Laporte				0.997
86873 Al-Eidan				0.237
86877 Ansah				0.42
86882 Katon	0.497	0.4325	0.870221	
86883 Walley				0.03
86889 Girvin	0.901	0.1139	0.126415	0.726
86889 Girvin				0.296
86889 Girvin				0.478
86889 Girvin				0.611
86889 Girvin	0.949	0.0509	0.053635	
86889 Girvin	0.901	0.1139	0.126415	0.726
86890 Gallefoss				0.32
86890 Gallefoss				0.58
86891 Peveler				0.51
86894 Henry	0.969			0.576
86894 Henry				0.39
86894 Henry				0.576
86897 Brus	0.84	0.21	0.25	
86899 Cote				0.91
86906 Becker				0.753
86906 Becker				0.541
86906 Becker				0.835
86906 Becker				0.553
86906 Becker				0.753
86907 Peterson				0.65
86907 Peterson	1.9	1.5	0.789474	0.652
86907 Peterson	7.1	4.6	0.647887	
86907 Peterson	20.2	7.9	0.391089	
86907 Peterson				0.5

86908 Baird				0.9
86910 Haynes	43.2	42.85	0.991898	0.39
86911 Sackett				0.51
86911 Sackett				0.56
86912 Colcher				0.58
59167 Moshkovska				0.32
87229 Kimmel				0.77
88919 Wu				0.36
63023 Sherrard				0.497

Table 9: The correlation of measurement quality and the coefficient of variation of continuous adherence results in a sample of randomized trials of interventions to increase patient adherence

Measure	Number of studies with this measure	Median quality score (9 maximum)	Mean coefficient of variation (CV = SD/mean) for control	Spearman correlation of quality with coefficient of variation (precision) of adherence results (r and 95% CIs)
Attendance[a]	0	3	----------	----------
Clinician judgment	0	2.5	---------	----------
Direct observation[a]	1	4.5	----------	----------
Drug concentration in body	3	7	0.609	Insufficient data[b]
MEMS	10	6	0.336	0.567 (-0.111 to 0.903)
Pharmacy refill record	4	7	----------	----------
Pill count	7	5	0.319	0.296 (-0.588 to 0.857)
Self-report diary	1	3	0.1139	Insufficient data[b]
Self-report questionnaire	6	3	0.450	0.447 (-0.901 to 0.984)
Therapeutic response[a]	0	6	---------	--------

a- These measures did not report any adherence results continuously and therefore do not have corresponding coefficients of variation

b- Measures with "insufficient data" either had no variability in their quality score therefore no correlation was calculated in SPSS, or only one study reported continuous adherence results for that measure and therefore no correlation was calculated in SPSS

Table 10: The correlation of measurement quality and proportion adherent in a sample of randomized trials of interventions to increase patient adherence reporting data as a binary variable

Measure	Number of studies with this measure	Mean quality score (9 maximum)	Average proportion adherent based on adherence results across studies in control	Pearson correlation of quality with proportion adherent in control group (r and 95% CIs)
Attendance	1	4	0.450	Insufficient data[a]
Clinician judgment	2	1.5	0.705	Insufficient data[a]
Direct observation	1	4	0.516	Insufficient data[a]
Drug concentration in body	1	7	0.740	0.725 (-0.436, 0.98)
MEMS	11	7	0.767	-0.482 (-0.806 to 0.065)
Pharmacy refill record	10	8	0.543	0.394 (-0.43, 0.859)
Pill count	11	5	0.580	-0.411 (-0.784 to 0.180)
Self-report interview	4	3.5	0.587	-0.203 (-0.871 to 0.728)
Self-report questionnaire	17	4	0.629	0.180 (-0.559 to 0. 69)
Self-report diary	3	3	0.522	Insufficient data[a]
Therapeutic response	2	6.5	0.536	Insufficient data[a]

a- Measures with "insufficient data" either had no variability in their quality score therefore no correlation was calculated in SPSS, or only one study reported binary adherence results for that measure and therefore no correlation was calculated in SPSS

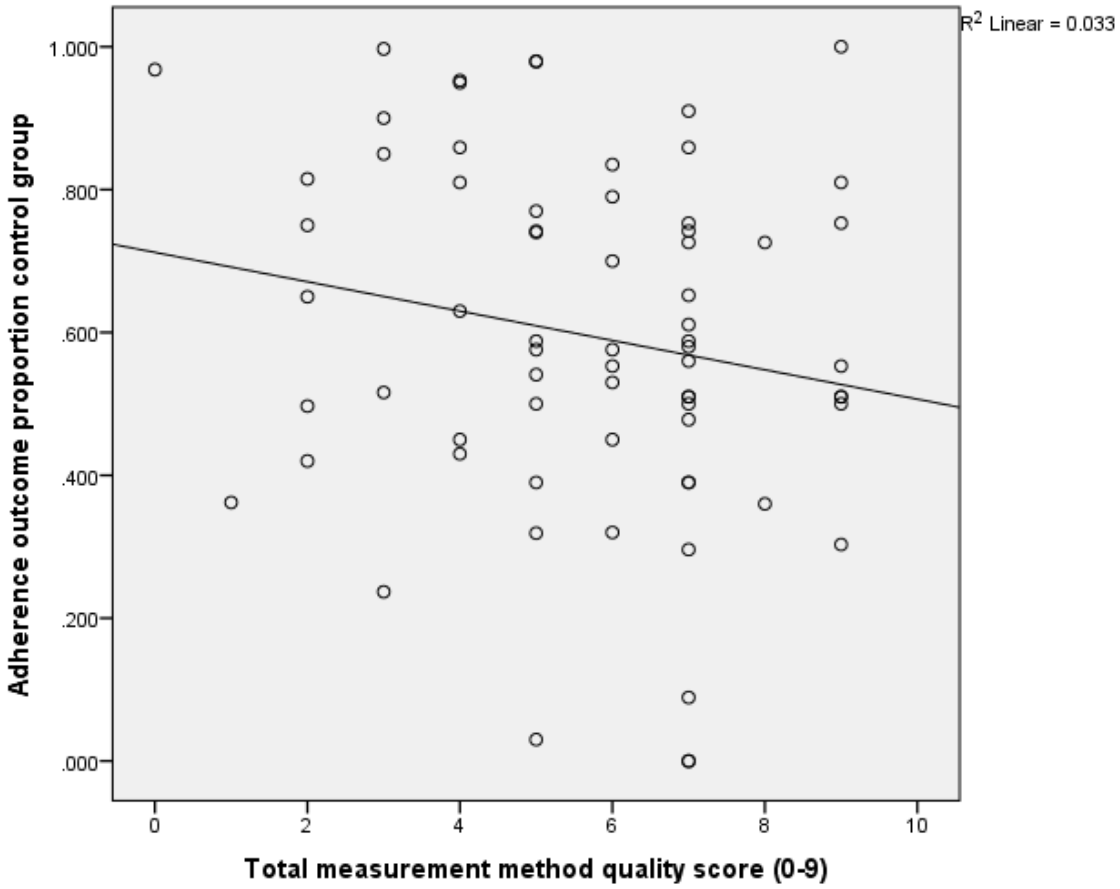


Figure 2: Graph of overall correlation of proportion adherent and measurement quality with linear line of best fit

Table 11: Patient recruitment methods quality scores in included studies from analytic survey, by study, per feature of quality and overall quality score

Study ID	Nonadherent patients selected score	Representativeness	Baseline adherence	Overall score (maximum 6)
Haynes, 1976	2	1	2	5
Cote, 1997	0	1	0	1
Brus, 1998	0	1	1	2
Girvin, 1999	0	1	0	1
Al-Eidan, 2002	0	1	0	1

Laporte, 2003	1	1	1	3
Weber, 2004	0	1	1	2
Remien, 2005	2	0	2	4
Ansah, 2001	0	1	0	1
Hederos, 2005	1	1	0	2
Peveler, 1999	0	2	0	2
Andrade, 2005	0	1	0	1
Ginde, 2003	0	1	0	1
Berrien, 2004	1	2	0	3
Katon, 2001	0	2	0	2
Walley, 2001	0	2	0	2
Gallefoss, 1999	1	1	1	3
Henry, 1999	0	2	0	2
Becker, 1986	1	1	1	3
Peterson, 1984	0	1	0	1
Baird, 1984	0	1	0	1
Rickles, 2005	0	2	1	3
Rawlings, 2003	1	1	0	2
Sackett, 1975	0	2	0	2
Colcher, 1972	0	1	1	2
Abrahams, 2010	0	2	0	2
Lai, 2011	0	1	0	1
Farooq, 2011	0	2	0	2
Kalichman, 2011	0	2	1	3
DiIorio , 2008	0	1	1	2
Chang, 2011	0	1	0	1
Villeneuve , 2010	1	1	0	2
Otsuki, 2009	1	0	1	2
Mullan, 2009	1	2	0	3
Sarna, 2008	0	2	1	3
Purcell, 2007	0	2	1	3
Velligan, 2008	0	2	0	2
Sorensen, 2007	2	2	2	6
Perrin, 2010	0	0	0	0
Klein, 2007	0	1	0	1
Choudhry, 2011	0	2	0	2
Matsumura, 2012	0	2	1	3
Sirey, 2010	0	2	0	2

Nieuwkerk, 2012	0	2	0	2
Solomon, 2012	1	0	0	1
Moshkovska, 2011	0	0	1	1
Wilson, 2010	1	2	1	4
Kimmel, 2012	0	0	0	0
Wu, 2012	0	0	1	1
Sherrard, 2009	0	0	0	0

Table 12: Quality of patient recruitment methods in a sample of randomized trials of interventions to increase patient adherence

Dimension of patient recruitment	Frequency of studies that did not fulfill quality dimension n(%)	Frequency of studies that are unclear if quality dimension fulfilled n(%)	Frequency of studies that fulfill that quality dimension n(%)	Median score in this dimension and range
Nonadherent patients selected	36 (72%)	11 (22%)	3 (6%)	0 (0-2)
Representativeness of sample	8 (16%)	22 (44%)	20 (40%)	1 (0-2)
Reporting of results	30 (60%)	17 (34%)	3 (6%)	0 (0-2)

Table 13: Data for patient recruitment secondary analysis by study, for binary and continuous data

	BINARY				CONTINUOUS			
Study ID	**Control N at follow-up (1)**	**Control group proportion adherent at follow-up from primary adherence measure**	**Calculated N needed for each group to detect 25% absolute increase in proportion of patients who are adherent**	**% recruitment of calculated for 25% increase**	**Adherence outcome mean control group**	**Adherence outcome SD control group**	**Calculated N needed for each group to detect 25% absolute increase in proportion of patients who are adherent**	**% recruitment of calculated for 25% increase**
86910 Haynes	18	0.39	61	0.293	0.432	0.4285	23.058	0.781

86899 Cote	54	0.91	42	1.277				
86897 Brus	30				0.84	0.21	10.683	2.808
86889 Girvin	25	0.726	25	0.996	0.949	0.0509	7.818	3.198
86889 Girvin	25	0.296	60	0.418	0.901	0.1164	10.850	2.304
86889 Girvin	25	0.478	59	0.427				
86889 Girvin	25	0.611	46	0.541				
86873 Al-Eidan	38	0.237	56	0.675				
86868 Laporte	42	0.997	1307	0.032	0.807	0.194	7.930	5.296
86859 Weber	24	0.5	57	0.419	0.889	0.196	24.472	0.981
86850 Remien	94	0.39	61	1.531	0.66	0.3	11.302	8.317
86877 Ansah	144	0.42	61	2.364				
86842 Hederos	24	0.7	31	0.767				
86891 Peveler	48	0.51	57	0.849				
86844 Andrade	29				0.64	0.442	24.534	1.182
86866 Ginde	31	0.742	20	1.552				
86856 Berrien	15				0.862	0.105	4.544	3.301
86882 Katon	145				0.497	0.4325	23.491	6.173
86883 Walley	162	0.03	27	6.072				
86890 Gallefoss	66	0.32	61	1.088				
86890 Gallefoss	66	0.58	50	1.319				
86894 Henry	59	0.39	61	0.961				
86894 Henry	59	0.576	50	1.169				
86906 Becker	85	0.541	54	1.574				
86906 Becker	85	0.753	15	5.860				
86907 Peterson	24	0.65	40	0.593				
86907 Peterson	24	0.5	57	0.419				
86908 Baird	196	0.9	38	5.170				
86847 Rickles	32				0.514	0.392	19.297	1.658
86864 Rawlings	99	0.74	21	4.771				
86877 Ansah	144	0.42	61	2.364				
86911 Sackett	112	0.51	57	1.980				
86911 Sackett	113	0.56	52	2.165				
86912 Colcher	100	0.58	50	1.999				
60980	128	0.516	56	2.281				

Abrahams								
60980 Abrahams	128	0.362	61	2.084				
60980 Abrahams	128	0.319	61	2.111				
58691 Lai	89				0.9646	0.1017	64.780	1.374
58691 Lai	89				0.9617	0.1095	64.156	1.387
58918 Farooq	46	0.45	60	0.767				
60196 Kalichman	210				0.66	0.22	6.078	34.550
65442 DiIorio	106				0.55	0.364	16.639	6.371
83030 Chang	461	0.815	20	23.222				
83030 Chang	461	0.742	20	23.077				
83030 Chang	461	0.979	186	2.484				
61894 Villeneuve	110	0.68	35	3.118				
62574 Otsuki	77	0.9496	77	1.006	0.53	0.8	80.373	0.958
62574 Otsuki	77	0.9496			0.6	0.88	97.251	0.792
62874 Mullan	33	0.81	19	1.711				
64892 Sarna	94	0.953	82	1.144				
64892 Sarna	94	0.588	49	1.914				
65867 Purcell	121	0.859	26	4.569				
65330 Velligan	29	0.53	55	0.527				
66971 Sorensen	28				0.716	0.2554	8.192	3.418
66971 Sorensen	28				0.784	0.2835	13.521	2.071
66971 Sorensen	28				0.531	0.2922	10.722	2.611
75329 Perrin	49	0.5	57	0.856	0.767	0.305	13.449	3.643
75329 Perrin	49	0.553	53	0.927	0.737	0.36	16.275	3.011
77245 Klein	21	0.63	44	0.482	0.808	0.124	3.274	6.415
77245 Klein	21	0.75	14	1.467	0.972	0.136	185.169	0.113
77245 Klein	21	0.51	57	0.371				
77245 Klein	21	0.79	17	1.213				
85044 Choudhry	3010	0.089	39	77.823				
86695 Matsumura	99	0.98	195	0.508				
61577 Sirey	30	0.43	61	0.495	0.483	0.223	6.245	4.804

86690 Nieuwkerk	97				0.984	0.167	855.068	0.113
58320 Solomon	1025	0.303	60	17.062				
59167 Moshkovska	34	0.32	61	0.561				
61989 Wilson	189				0.43	0.3485	3723.688	0.051
87229 Kimmel	43	0.77						
88919 Wu	28	0.36	61	0.456				
63023 Sherrard	143	0.497	57	2.489				

Table 14: Correlation of quality of patient recruitment and power of trials in a sample of randomized trials with continuous adherence data of interventions to increase patient adherence

Patient selection	Score	Median (IQR) proportion recruitment of calculated for 25% increase for continuous adherence outcomes[a]	Wilcoxon rank sum test results
Nonadherent and uncertainly adherent patients selected	1	2.34 (2.56)	p=0.660
Unspecified adherence level patients selected	0	2.81 (3.43)	

a- % recruitment is a measure of the power of the trial and refers to the proportion of the actual versus theoretical required sample size to achieve an alpha of 0.05 and a beta or power of 80%.

Table 15: Correlation of quality of patient recruitment methods based on the representativeness of the sample and the power of trials in a sample of randomized trials with continuous adherence data of interventions to increase patient adherence

Representativeness	Score	Median (IQR) proportion recruitment of calculated for 25% increase for continuous adherence outcomes[a]	Wilcoxon rank sum test results
Sample representative	1	2.96 (2.70)	p=0.675
Sample not representative or uncertainly representative	0	2.30 (2.66)	

a- % recruitment is a measure of the power of the trial and refers to the proportion of the actual versus theoretical required sample size to achieve an alpha of 0.05 and a beta or power of 80%.

Table 16: Correlation of quality of patient recruitment methods based on whether results are reported based on adherence and the power of trials in a sample of randomized trials with continuous adherence data of interventions to increase patient adherence

Reporting of results	Score	Median (IQR) proportion recruitment of calculated for 25% increase for continuous adherence outcomes[a]	Wilcoxon rank sum test results
Reported based on baseline adherence levels	1	2.34 (3.86)	p=0.905
Not reported based on baseline adherence levels	0	3.01 (2.27)	

a- % recruitment is a measure of the power of the trial and refers to the proportion of the actual versus theoretical required sample size to achieve an alpha of 0.05 and a beta or power of 80%

Table 17: Correlation of dichotomized quality of patient recruitment methods based on whether nonadherent patients were recruited and the transformed power of trials in a sample of randomized trials with binary adherence data of interventions to increase patient adherence

Patient selection	Score	Median (IQR) proportion recruitment of calculated for 25% increase for binary adherence outcomes[a]	Wilcoxon rank sum test results
Nonadherent and uncertainly adherent patients selected	1	1.53 (2.11)	p= 0.560
Unspecified adherence level patients selected	0	1.19 (1.75)	

a- % recruitment is a measure of the power of the trial and refers to the proportion of the actual versus theoretical required sample size to achieve an alpha of 0.05 and a beta or power of 80%

Table 18: Correlation of quality of patient recruitment methods based on the representativeness of the sample and the power of trials in a sample of randomized trials with binary adherence data of interventions to increase patient adherence

Representativeness of sample	Score	Median (IQR) proportion recruitment of calculated for 25% increase for binary adherence outcomes[a]	Wilcoxon rank sum test results
Sample representative	1	1.81 (1.28)	p= 0.313
Sample not representative	0	1.21 (1.87)	

a- % recruitment is a measure of the power of the trial and refers to the proportion of the actual versus theoretical required sample size to achieve an alpha of 0.05 and a beta or power of 80%.

Table 19: Correlation of quality of patient recruitment methods based on whether results are reported based on adherence and the power of trials in a sample of randomized trials with binary adherence data of interventions to increase patient adherence

Reporting of results	Score	Median (IQR) proportion recruitment of calculated for 25% increase for binary adherence outcomes[a]	Wilcoxon rank sum test results
Reported based on baseline adherence levels	1	1.32 (1.71)	p= 0.972
Not reported based on baseline adherence levels	0	1.24 (1.65)	

a- % recruitment is a measure of the power of the trial and refers to the proportion of the actual versus theoretical required sample size to achieve an alpha of 0.05 and a beta or power of 80%

Table 20: The mean effect size for each dichotomized quality score (0= low, 1= high quality) for each aspect of the patient recruitment quality scale for continuous data and significance level for difference between high and low quality scores

Quality scale aspect	Quality score (dichotomized) (0= low quality, 1= high quality)	Number of studies in this group	Number of patients in this group from all included studies	Median of effect sizes for group (ES= ($mean_{intervention}$- $mean_{control}$)/ SD_{pooled}[e])	Interquartile range (IQR) of effect size for each group	Between group differences (Mann Whitney U test) p-value
Nonadherent patients recruited	No- 0[a]	13	1783	0.370	0.626	p= 0.150
	Yes or Uncertain- 1[b]	7	926	0.278	0.353	
Representative sample	No or Uncertain- 0[c]	12	1218	0.266	0.243	p=0.336
	Yes- 1[d]	8	1491	0.433	0.448	
Report results based on baseline adherence	No- 0[a]	9	1029	0.402	0.676	p=0.053
	Yes or Uncertain- 1[b]	11	1680	0.266	0.330	

a- 0 quality score for aspect of recruiting nonadherent patients and reporting results based on baseline adherence refers to a score of 0 on the original scale before dichotomization, wherein the item was NOT done

b- 1 quality score for aspects of recruiting nonadherent patients and reporting results based on baseline adherence refers to a score of 1 or 2 on the original scale before dichotomization, wherein the item was done (YES) or it is uncertain whether the item was done (UNCERTAIN)

c- 0 quality score for aspect of recruiting a representative sample refers to a score of 0 or 1 from the original scale, where the item was NOT done or was UNCERTAIN whether it was done

d- 1 quality score for aspect of recruiting a representative sample refers to a score of 2 from the original scale, where the item was (YES) done, i.e. the sample was representative

e- $SD_{pooled} = \sqrt{[(SD_1^2 + SD_2^2)/2]}$

Table 21: The mean effect size for each dichotomized quality score (0= low, 1= high quality) for each aspect of the patient recruitment quality scale for binary data and significance level for difference between high and low quality scores

Quality scale aspect	Quality score (dichotomized)	Number of studies in this group	Number of patients in this group over all included studies[g]	Median effect size for group (ES= ($mean_{intervention}$ - $mean_{control}$)/ SD_{pooled}[e,f])	Interquartile Range (IQR) of effect size for each group	Between group differences (Mann Whitney U test) p-value
Nonadherent patients recruited	No- 0[a]	28	10513	2.434	3.614	p=0.020
	Yes or Uncertain- 1[b]	11	3429	1.508	3.177	
Representative sample	No or Uncertain- 0[c]	26	5930	2.454	3.210	p= 0.173
	Yes- 1[d]	13	8012	1.083	3.174	
Report results based on baseline adherence	No- 0[a]	25	11153	2.245	2.862	p= 0.141
	Yes or Uncertain- 1[b]	14	2789	1.864	3.013	

a- 0 quality score for aspect of recruiting nonadherent patients and reporting results based on baseline adherence refers to a score of 0 on the original scale before dichotomization, wherein the item was NOT done

b- 1 quality score for aspects of recruiting nonadherent patients and reporting results based on baseline adherence refers to a score of 1 or 2 on the original scale before dichotomization, wherein the item was done (YES) or it is uncertain whether the item was done (UNCERTAIN)

c- 0 quality score for aspect of recruiting a representative sample refers to a score of 0 or 1 from the original scale, where the item was NOT done or was UNCERTAIN whether it was done

d- 1 quality score for aspect of recruiting a representative sample refers to a score of 2 from the original scale, where the item was (YES) done, i.e. the sample was representative

e- SD for intervention and control groups = $\sqrt{}$ ($p*q / n_{intervention\ or\ control}$)

f- $SD_{pooled} = \sqrt{} [(SD_1^2 + SD_2^2)/2]$

g- Some studies did not report the number of patients in one group at follow up so this value is not the complete sample size

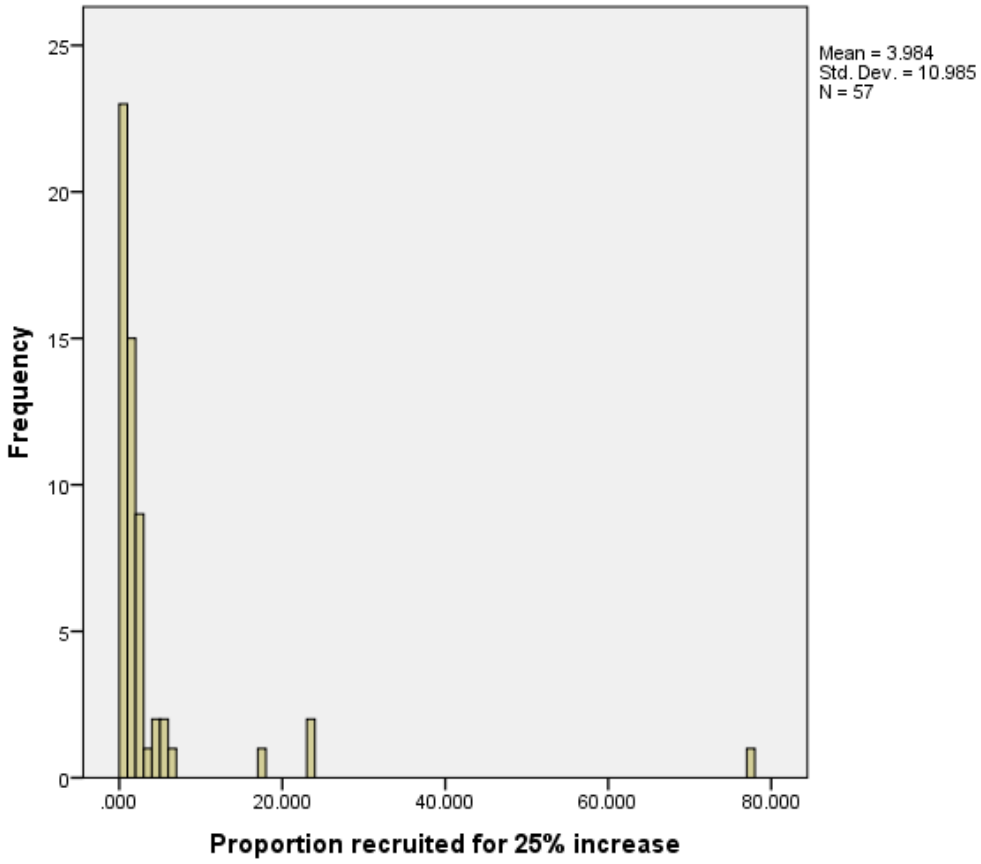


Figure 3: Distribution of results for patient recruitment for a 25% minimally important difference for the sample size calculation, where frequency refers to the number of studies with that proportion of the theoretical sample size recruited, with possible values for the proportion recruitment lumped into values with intervals of 1 and the majority of studies had 0-3 times the required theoretical sample size recruited

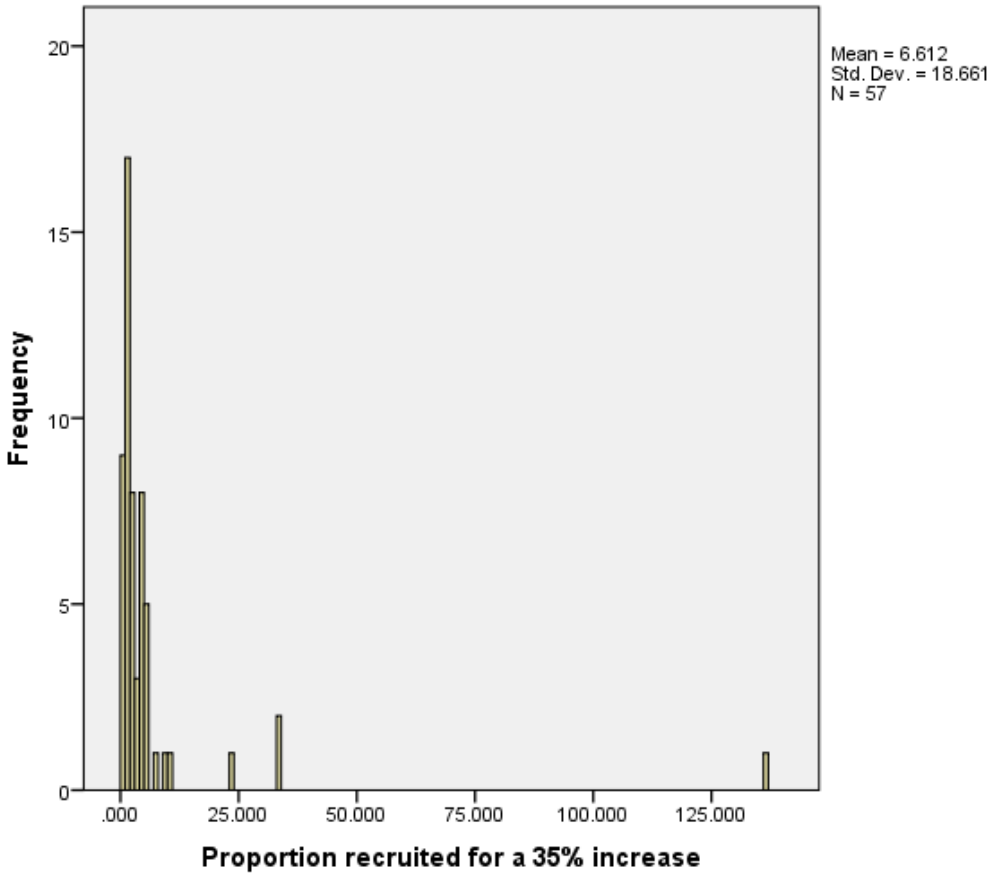


Figure 4: Distribution of results for patient recruitment for a 35% minimally important difference for the sample size calculation, where frequency refers to the number of studies with that proportion of the theoretical sample size recruited, with possible values for the proportion recruitment lumped into values with intervals of 1.5 and the majority of studies had 0-6 times the required theoretical sample size recruited

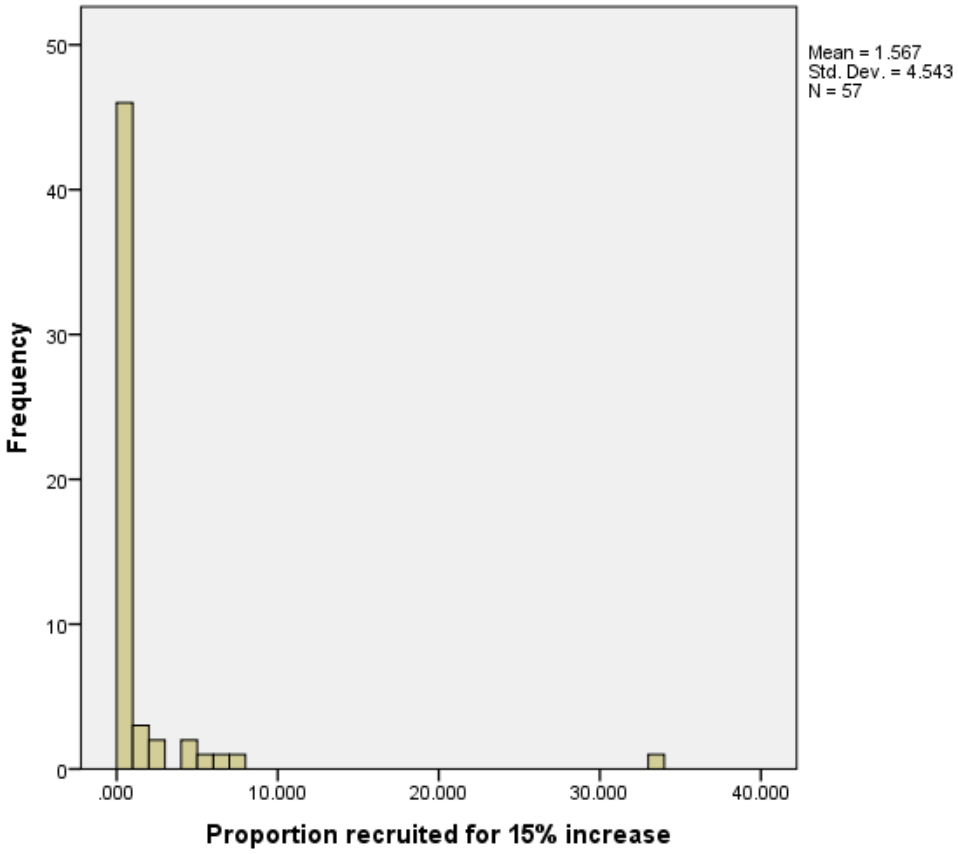
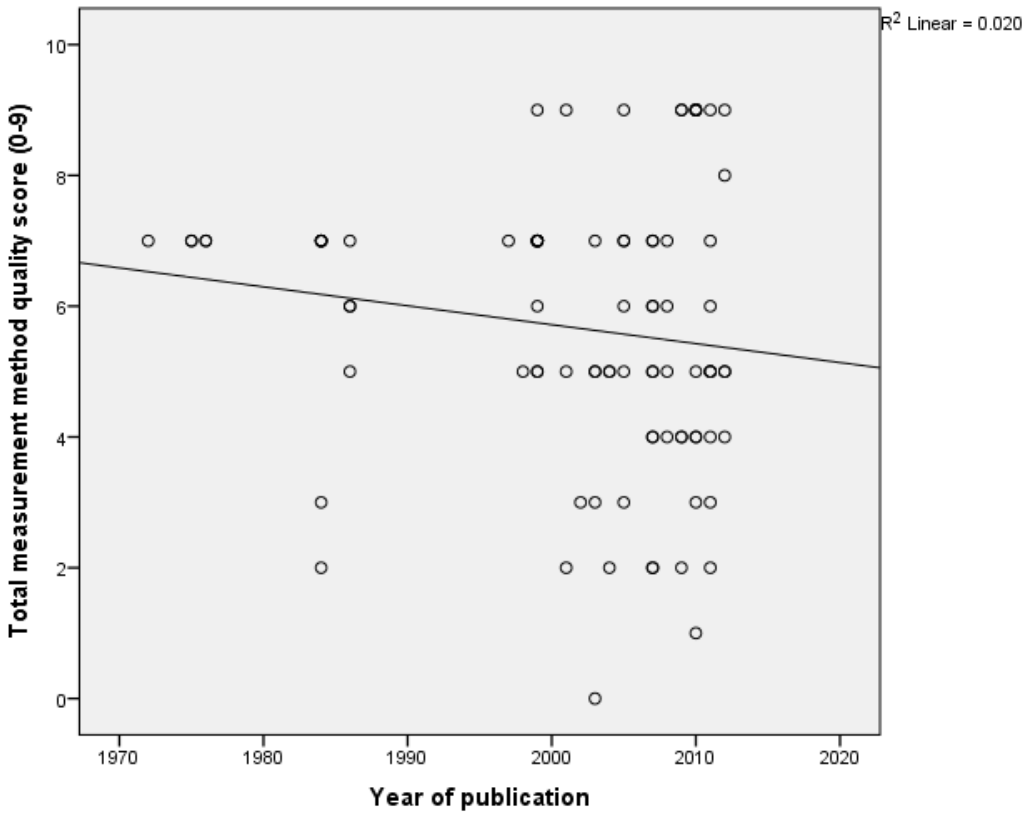
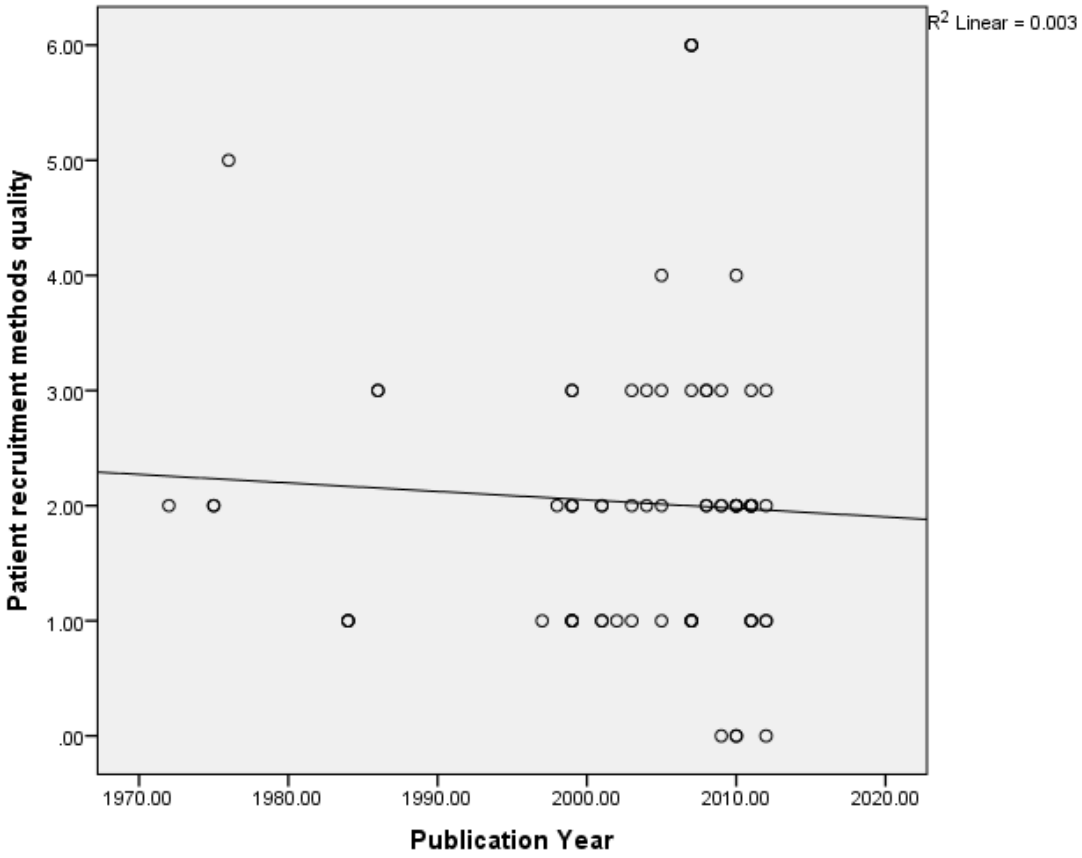


Figure 5: Distribution of results for patient recruitment for a 15% minimally important difference for the sample size calculation, where frequency refers to the number of studies with that proportion of the theoretical sample size recruited, with possible values for the proportion recruitment lumped into values with intervals of 0.5 and the majority of studies had 0-1 times the required theoretical sample size recruited



* R^2 refers to the coefficient of determination, which is correlation coefficient (r) squared, and indicates the strength of the correlation

Figure 6: Study quality versus year of publication of article for measurement of adherence methodology



* R^2 refers to the coefficient of determination, which is correlation coefficient (r) squared, and indicates the strength of the correlation

Figure 7: Patient recruitment quality scores versus year of publication of article

Statistical appendix

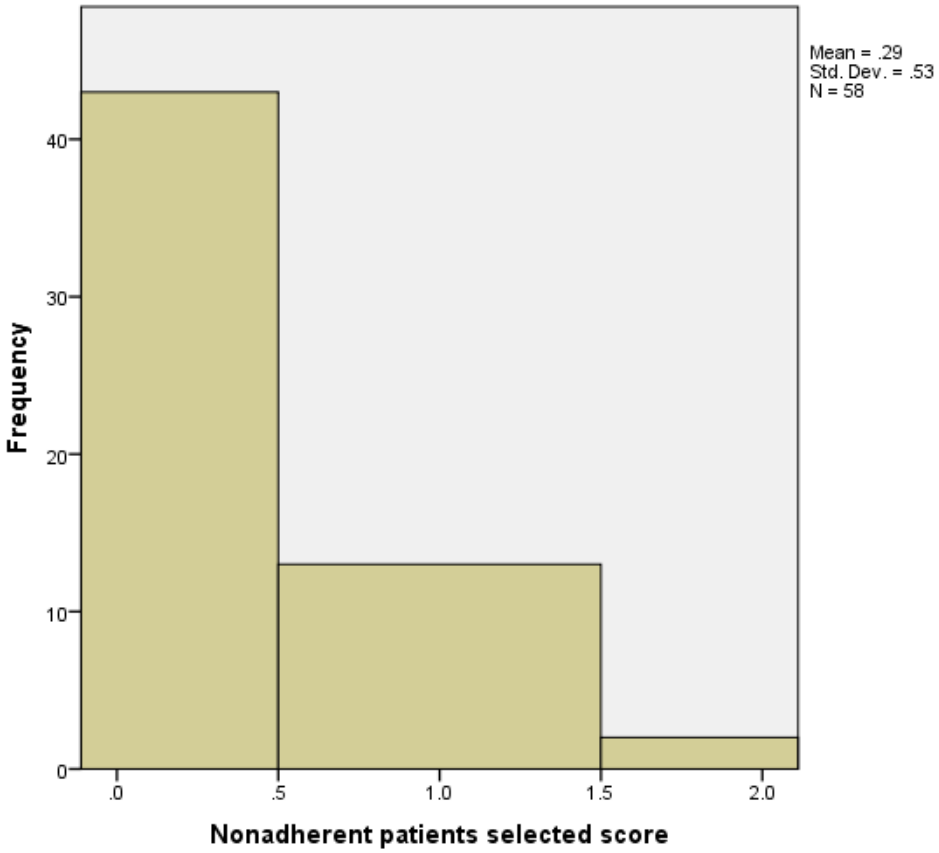


Figure 1: Distribution of results for binary patient recruitment methodology data for scores of 0, 1 and 2 for one aspect of the patient recruitment scale, whether nonadherent patients were selected

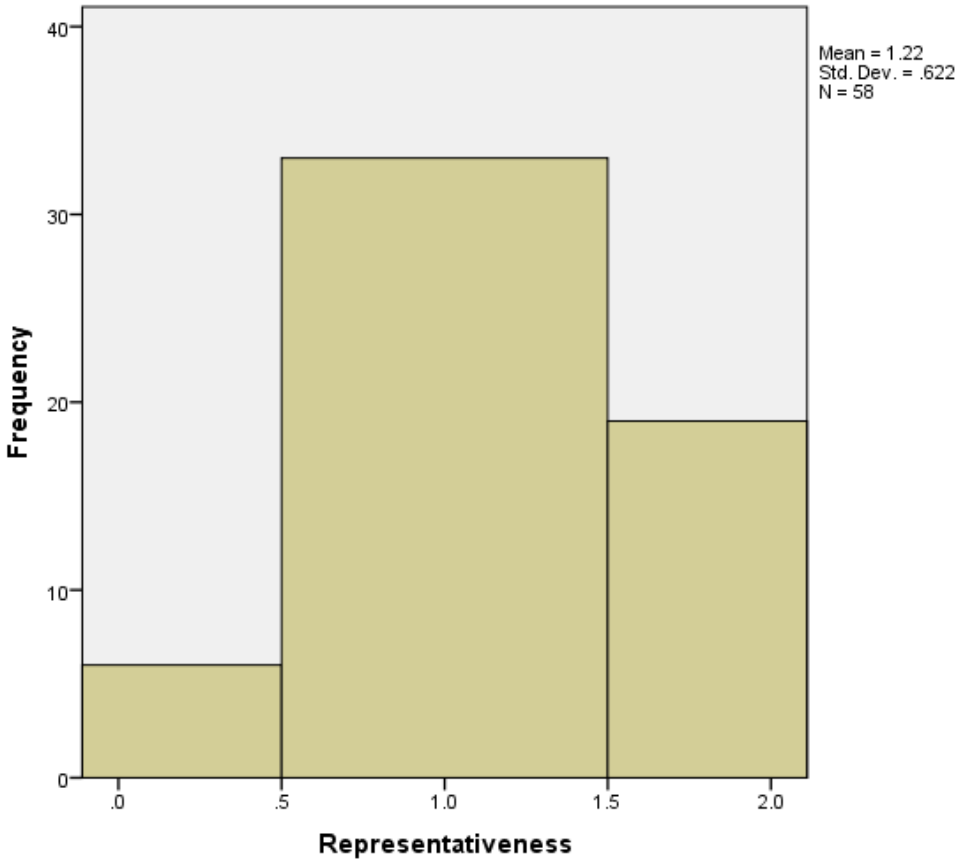


Figure 2: Distribution of results for binary patient recruitment methodology data for scores of 0, 1 and 2 for one aspect of the patient recruitment scale, whether the study sample was representative

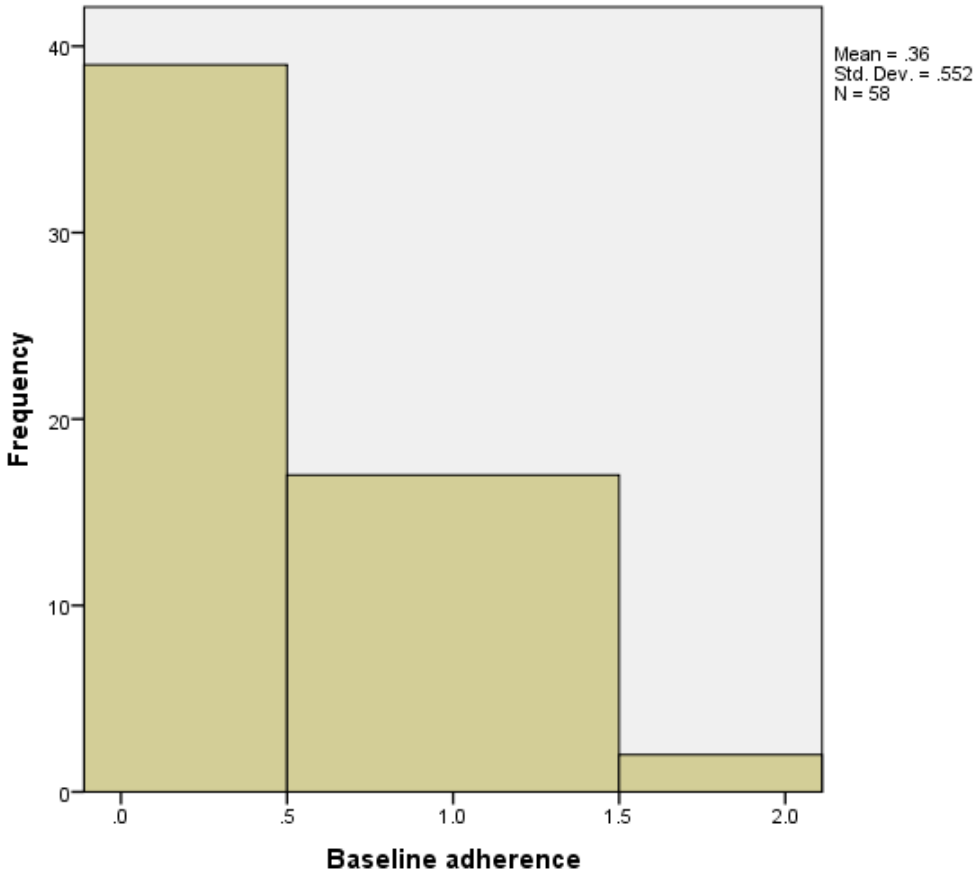


Figure 3: Distribution of results for binary patient recruitment methodology data for scores of 0, 1 and 2 for one aspect of the patient recruitment scale, whether the study reported the results from the trial based on baseline adherence levels

7.0 Appendix

Appendix 1: Glossary of terms

Adherence results: The outcomes in randomized controlled trials of the amount of medication the patient consumed relative to the amount of prescribed medication the patient should have consumed had they followed their prescription, quantified by the measure of adherence.
Composite measure: A measure of adherence to medication that is made up of two or more measures in a single trial. For example, a single trial might include both a self-report questionnaire and a pharmacy refill record to quantify adherence.
Precision: Related to the standard deviation of an outcome or result from a trial, this refers to how close the measured outcome is to the true outcome
Quality: Refers to the superiority of a method over another. For example, a high quality measure of adherence may be more objective than a lower quality measure of adherence.

Appendix 2: Full data extraction form

***Note**: Online extraction form is broken into questions for Objective 2 and 3 and questions for Objectives 1 and 4 but has been merged here. Info buttons that appear online have been deleted here for ease of reading.*

1. **Title of article**: [automatically inserted]

2. **Citation**: [automatically inserted]

3. **PubMed ID/Ovid accession number/Database ID**: [automatically inserted]

4. **Author Abstract**: [automatically inserted]

5. **Behavioural Theory Questions**

a). Was the intervention explicitly theory based?
Yes/No/Unclear
If yes (indicate theory or theories)_____

a) Were stakeholders involved in developing the intervention?
Yes/No/Not reported
If yes (check all that apply)

i. Patients
ii. Policy makers

iii. Healthcare providers
iv. Others _____

b) Were the reasons for not adhering to the medication(s) reported assessed in the studied population? (Check all that apply)

i. No
ii. Yes, assessed in the studied population in advance of the RCT (copy text): _____
iii. Yes, assessed during the RCT (copy text):_____

c) Was the level of participation with the adherence intervention in the RCT reported?

i. No
ii. Yes (Please copy text; specify numbers, e.g., % attendance to education visits):_____

If Yes, were reasons for not participating with the adherence intervention in the current study assessed?
i. No
ii. Yes (please copy text):_____

d) How many medications were targeted with the adherence intervention?

i. 1
ii. >1
iii. Uncertain

e) From the patients' perspective, when was the adherence intervention started? (Check all that apply)

i. At the time of newly starting medication
ii. Following a major event (e.g., Hospitalization),
iii. After being on treatment for some time, no triggering event
iv. Uncertain

f) What was the frequency and duration of the adherence intervention application? [Copy and paste text]:_____

6. **Study Setting**: [cut and paste text from the article and/or type your answer in the format to complete the sentence "The study location was:" and enter the site/ State Province/ Country]

7. **Clinical Problem** for which the participants are being recruited: [For insertion in the Cochrane Adherence and Outcomes Table; typed answer:_____]

8. **Participant inclusion/exclusion criteria**:

   a) Cut and paste text from the article :_____
   b) Type the answer___ in format of
   text box 1: The inclusion criteria were: __
   text box 2: The exclusion criteria were (if given): __]
   c) Number randomized to Intervention group: [text box to enter number or "not specified"]
   d) Number randomized to Control group: [enter number or "not specified"]
   e) Nature of medicine regimen:
      i. ☐ study is about a new or old medication prescribed for an acute condition (i.e., to be prescribed for < 6 months)

      ii. ☐ study is about patients who will be prescribed a <u>new</u> medication for a chronic condition at the start of the period of intervention (i.e., to be prescribed for ≥ 6 months)

      iii. ☐ study is about testing an intervention to increase adherence with a medication(s) that has (have) been prescribed for a chronic condition before the study began; <u>if so</u>:

         a) Did they measure adherence at baseline?
            Yes/ No/ Uncertain

         b) Did they use the adherence status of the patients for recruitment/ as part of inclusion criteria?
            Yes/ No/ Uncertain

9. **Recruitment and drop outs**:

   a) Number of patients eligible who were asked to join before reaching sample size, enter number or N/A if not specified []

   b) Number of patients in sample size at baseline (enter number or N/A) []

   c) Total number of patients randomized (enter number or N/A if not specified) [ ]

   d) Number assessed for adherence in Intervention group at least 24 weeks after entry for chronic regimens or at final assessment for acute regimens [enter number or "not specified" and any details]

e) Time period for drop out report (e.g., 1 week; 24 weeks; 1 year, etc) [enter time periods given as # weeks/months/years]

f) Number assessed for adherence in Control group at least 24 weeks after entry for chronic regimens or at final assessment for acute regimens : [enter number or "not specified" and any details]

g) Time period for drop out report (e.g., 1 week; 24 weeks; 1 year): [enter time periods given as # weeks/months/years]

10. **Risk of bias** *refer to info buttons for judgment:

a) Random sequence generation: Extractor's judgment: High risk of bias / Unclear / Low risk of bias
Support for judgment – [ ]

b) Allocation concealment: Extractor's judgment: High risk of bias / Unclear / Low risk
Support for judgment – [ ]

c) How was adherence data collected? [List]
Blinding of staff and personnel to study group among those collecting adherence data: Extractor's judgment: High risk / Unclear / Low risk
Support for judgment – []:

d) How many outcomes were collected? [list them]
Blinding of staff and personnel to study group among those collecting patient outcomes: Extractor's judgment: High risk / Unclear / Low risk
Support for judgment – **[]**

e) Blinding of patients to study group. Extractor's judgment: High risk / Unclear / Low risk
Support for judgment – [cut and paste text from the article or typed answer]:

f) How many study groups were there? [list them]Blinding of key study personnel to study group. Extractor's judgment: High risk / Unclear / Low risk
Support for judgment – [cut and paste text from the article or typed answer]:

g) How much incomplete outcome data was identified? [list it]
Incomplete outcome data (refer to Cochrane rules): Extractor's judgment: High risk / Unclear / Low risk
Support for judgment – [cut and paste text from the article or typed answer]:

h) Was there selective reporting bias (i.e. did they fail to report what they said they would report)?
Extractor's judgment: High risk / Unclear / Low risk
Support for judgment – [cut and paste text from the article or typed answer]:

i) Was there adjustment for multiple comparisons? Yes / Unsure / No /

Support for judgment: [text box]

j) Was the primary outcome clearly stated? Yes /No /Unsure; If Yes, what was the primary outcome:

k) Was the analysis plan clearly stated? Yes /No /Unsure; If Yes, copy & paste the analysis plan:

l) Was the study sufficiently powered? Yes /No /Unsure

m) Was there any other type of bias noted? High risk / Unclear / Low risk
Support for judgment – [cut and paste text from the article or typed answer]:

n) Was there statistical adjustment for clustering? Yes / Unsure / No / Not applicable (ie, not a cluster RCT)

11. Intervention:

a) Provide a complete description of the intervention(s) received by one group (if more than 1 intervention group, report each intervention group separately here) [cut and paste text from the article]:_____

b) Use your own words and fill out the following headings, when possible:
Heading - Nature of intervention: [textbox]
Heading- Who did what to whom: [textbox]
Heading - Number of sessions: [textbox]
Heading - Duration of follow up: [textbox]
Heading - Intensity of intervention (if mentioned): [textbox]

c) Based on b), who received the adherence intervention?

i. Patient
ii. Caregivers
iii. Patients' family or friends
iv. Research staff
v. Other [typed answer _____]

d) Based on the description in a) what type(s) of intervention(s) was (were) tested in the intervention group (<u>check all that apply</u>)?

i. Increased supervision <u>from physician prescriber</u> (e.g., more frequent visits, phone calls)
ii. Increased supervision from nurse
iii. Increased supervision by pharmacist
iv. Increase supervision by other provider
[Typed answer: _____]
v. Additional medication instructions for patients in the intervention group (e.g., verbal, written, or visual material)

vi. Additional instructions/ education for patients in the intervention group on their disease or adherence (e.g., verbal, written, or visual material)
vii. Patient counseling (e.g., about the target disease, importance of therapy, compliance with therapy, possible side-effects, patient empowerment, couple-focused therapy to increase social support)
viii. Automated patient monitoring and counseling (telephone, cell phone, or computer assisted)
ix. Manual telephone follow-up
x. Family/social intervention (e.g., involvement of family members/significant others in the intervention)
xi. Increased convenience of care (e.g., provision at the workplace or at home)
xii. Simplified dosing (i.e. changing frequency of medication use)
xiii. Increased self-care (involving patients more in their care, e.g., through self-monitoring their blood pressure)
xiv. Reminders (e.g., programmed medication devices)
xv. Special pill packaging (e.g., calendar packs)
xvi. Dose-dispensing units of medication and medication charts
xvii. Appointment and prescription fill reminders
xviii. Reinforcements or rewards for improved adherence
xix. Reinforcement or rewards for improved treatment response (e.g., reduced frequency of visits and partial payment for blood pressure monitoring equipment)
xx. Different medication formulations (e.g., tablet vs. syrup)
xxi. Crisis intervention conducted when necessary (e.g., for attempted suicide, aggressive and destructive behaviour)
xxii. Direct observation of treatments (DOTS) by health workers or family members
xxiii. Lay health mentoring
xxiv. Augmented pharmacy services
xxv. Psychological therapy (e.g., cognitive behaviour therapy, multisystemic therapy)
xxvi. Mailed communications
xxvii. Group meetings
xxviii. Social media (eg, Twitter, Facebook)
xxix. Other(s) [_____]

e) Provide a description of the Control procedure [cut and paste text from the article if possible:_____
f) And Use your own words and fill out the following headings, when possible:
Heading - Nature of intervention: [textbox]
Heading- Who did what to whom: [textbox]
Heading - Number of sessions: [textbox]
Heading - Duration of follow up: [textbox]
Heading - Intensity of intervention (if mentioned): [textbox]

g) Based on the description of the Control procedure, did Control participants receive attention or intervention beyond usual care with the intention of offsetting the additional attention received by Intervention participants? Yes / No / Unclear

12. **Adherence measurements**:

a) What was (were) the measures of adherence?

[Cut and paste article's description of measure(s) of adherence- include number of measures, what measure was the primary measure (if stated), name and types of measures, descriptors of measures used (e.g., Valid, sensitive, etc.):_____

Typed answer: **Put these as headings in the text box: List all the measures (names and types) and provide a description of each measure (include if the measure was valid, sensitive, etc) ; Indicate the primary measure (if stated):**

_____

b) What was (were) the measures of adherence? Check all measures of adherence that apply:

- ☐ clinician judgment
- ☐ therapeutic response, explicitly used to measure adherence
- ☐ attendance at scheduled follow up appointments
- ☐ self report- questionnaire
- ☐ Self- report – diary
- ☐ Self-report- interview
- ☐ drug concentration in body fluid (quantitative)
- ☐ presence of drug in body fluid (qualitative)
- ☐ pill count
- ☐ electronic pill monitor
- ☐ pharmacy refill record
- ☐ direct observation
- ☐ other – name them if not listed above [textbox]

c) Who is involved in assessing this measure of adherence (if explicitly stated, otherwise this is considered to be unclear)?

check boxes - check as many as apply

- ☐ caregiver (family or friend)
- ☐ patient
- ☐ physician
- ☐ nurse
- ☐ research staff (not directly responsible for patient medical care)
- ☐ pharmacist
- ☐ unclear

☐ other [Text box to enter who the Other is (enter more than one thing in text box if multiple unlisted persons)]

d) When was the measure of adherence taken [check boxes]? (check all that apply)
☐ Baseline
☐Continuously (eg, pharmacy record; MEMS)
☐Periodically; if so:
How many times after baseline: [text box to enter number of times and at what time points it was measured]

e) Quality of measure:
i. Validity
☐0- Documented in the article to <u>not</u> be valid OR stated in the article to be documented by another source to be invalid
☐1- Validity not assessed within the article
☐2- Measure documented to be valid in comparison with a criterion standard based on data in the study or citation of another study

ii. Reliability:
☐0- Documented in the article to <u>not</u> be reliable OR stated in the article to be documented by another source to be unreliable
☐1- Reliability not assessed within the article
☐2- Measure documented to be reliable or valid by data in the study or citation of another study (reliability is implicitly high if measure is valid)

iii. Objective:
☐0- Subjective measure without appropriate blinding or with uncertain blinding to patients' treatment group
☐1- Subjective measure with appropriate blinding (method or blinded group explicitly stated)
☐2- Objective measure

iv. Unobtrusive:
☐0- Obtrusive/obvious to patients
☐1- Unclear whether the patient is aware adherence is being measured
☐2- Patient is likely to be unaware the measure is being taken and the measure does not interrupt the normal pattern of medication consumption

v. Longitudinal data:
☐ 0- Data provided by measure covers up to one week of adherence for a chronic regimen
☐ 1- Data provided by measure describes adherence for more than one week, for a chronic regimen or for the duration of the acute medication regimen (if less than one week)

13. **Measurement of Clinical Health Outcomes**:
[cut & pasted text from article- enter number of clinical outcomes what the clinical health outcomes were, , how they were measured, how measured them, any descriptors of measures (valid, etc.)_____]

Outcomes:

14. **Results for adherence measurements**:

a) How were adherence data provided?

 as continuous variables (for example, mean possession rates from pharmacy pill monitoring) for blood pressure

{if checked}

i. what were the rates? [narrative and "cut and paste" **tables**, including statistical results comparing intervention and control groups, by adherence measure (if applicable): [____]

ii. Was there a statistically significant effect on adherence or compliance based on differences between Intervention and Control groups, as defined by p-values provided in the results, for continuous outcomes? Yes / No / Unsure

[Text box for any explanation/comments]

 as discrete variables (for example, proportion of patients achieving a high level of adherence)

{if checked}

i. what were the rates? [narrative and "cut and paste" **tables**, including statistical results comparing intervention and control groups, by adherence measure if applicable: [____]

ii. Was there a statistically significant effect on adherence or compliance based on **proportions** of "compliant" patients, comparing Intervention and Control patients? Yes / No / Unsure

[Text box for any explanation/comments]

15. **Results for clinical health outcomes**

a) Results for clinical health outcomes - extract all clinical outcomes in detail and their variance for each study group, as well as levels of statistical significance for differences between study groups.

[Narrative and "cut and paste" tables]

b) Was there a statistically significant effect on clinical health outcomes? Yes / No /Unsure

[Text box for any explanation/comments]

c) Were clinical health outcome results reported <u>based on baseline adherence?</u>**:**

☐ 0- Baseline adherence was not measured,

☐ 1- Baseline adherence measured but clinical results were not reported according to initial adherence level,

☐ 2- Yes, clinical results were reported based on baseline adherence level or <u>if only nonadherent</u> patients are recruited, if intention to treat analysis is followed

16. **Effect on Outcome**: [**For insertion in the Cochrane Adherence and Outcomes table;** typed answer based on cut and paste and answers to questions 13 and 14 (all parts)]:

17. Other outcomes

a) Were adverse effects of the intervention assessed?
Yes / No / Unstated
If yes, specify [_____]

b) Were incremental costs/resources of the intervention assessed?
Yes / No / Unclear / Unstated
If yes, specify [_____]

18. Funding

Was a funding source(s) statement reported in the manuscript? Yes/No
If yes,

a) was there public funding? Yes/No/Unclear

b) was there private (industry) funding Yes/No/Unclear

19. **Additional comments**:
[typed answer:_____

8.0 References

1. DiMatteo MR, Giordani PJ, Lepper HS, Croghan TW. Patient adherence and medical treatment outcomes: a meta-analysis. *Medical Care*. 2002;40(9):794–811.

2. Simpson SH, Eurich DT, Majumdar SR, et al. A meta-analysis of the association between adherence to drug therapy and mortality. *BMJ (Clinical research ed.)*. 2006;333(7557):15.

3. Jin J, Sklar GE, Min V, Oh S. Factors affecting therapeutic compliance : A review from the patient ' s perspective. *Therapeutics And Clinical Risk Management*. 2008;4(1):269–286.

4. Krousel-Wood M, Thomas S, Muntner P, Morisky D. Medication adherence: a key factor in achieving blood pressure control and good clinical outcomes in hypertensive patients. *Current Opinion in Cardiology*. 2004;19(4):357–62.

5. World Health Organization. *Adherence to long-term therapies: evidence for action*. 2003:199.

6. Haynes RB, Ackloo E, Sahota N, Hp M, Yao X. Interventions for enhancing medication adherence. *Cochrane Handbook for Systematic Reviews of Interventions*. 2008;(4).

7. Gordis L. Conceptual and methodologic problems in measuring patient compliance. In: Haynes R, Sackett D, eds. *Compliance in Health Care*. Baltimore: John Hopkins University Press; 1979:23–45.

8. Vermeire E, Hearnshaw H, Van Royen P, Denekens J. Patient adherence to treatment: three decades of research. A comprehensive review. *Journal of clinical pharmacy and therapeutics*. 2001;26(5):331–42.

9. Quittner AL, Modi AC, Lemanek KL, Ievers-Landis CE, Rapoff M a. Evidence-based assessment of adherence to medical treatments in pediatric psychology. *Journal of pediatric psychology*. 2008;33(9):916–36; discussion 937–8.

10. Dunbar-Jacob J, Mortimer-Stephens MK. Treatment adherence in chronic disease. *Journal of clinical epidemiology*. 2001;54 Suppl 1(July):S57–60.

11. Obias-Manno D, Friedmann E, Brooks MM, et al. Adherence and arrhythmic cuppression trial (CAST) mortality in the cardiac arrhythmia. *Annals of Epidemiology*. 7(95).

12. DiMatteo MR. Variations in patients' adherence to medical recommendations: A quantitative review of 50 years of research. *Medical Care*. 2004;42(3):200–209.

13. Haynes RB, McDonald HP, Garg AX. Helping patients follow prescribed treatment. *Journal of the American Medical Association.* 2002;288(22):2880–2883.

14. Farmer KC. Methods for measuring and monitoring adherence in clinical trial and clinical practice. *Clinical therapeutics*. 1999;21(6):1074– 1090.

15. Inzucchi S. Oral antihyperglycemic therapy for type 2 diabetes. *Diabetes*. 2002;287(3).

16. Haynes RB, Mckibbon KA, Kanani R. Systematic review of randomised trials of interventions to assist patients to follow prescriptions for medications. *The Lancet*. 1996;348:1994–1997.

17. Wickersham K, Colbert A, Caruthers D, et al. Assessing fidelity to an intervention in a randomized controlled trial to improve medication adherence. *Nursing research*. 2011;60(4):264–9.

18. Gross CP, Mallory R, Heiat A, Krumholz HM. Reporting the recruitment process in clinical trials : who are these patients and how did they get there ? *Annals of Internal Medicine*. 2002;137:10–16.

19. Hannan EL. Randomized clinical trials and observational studies: guidelines for assessing respective strengths and limitations. *JACC. Cardiovascular interventions*. 2008;1(3):211–7.

20. Haynes RB, Dantes R. Patient compliance and the conduct and interpretation of therapeutic trials. *Controlled Clinical Trials*. 1987;8:12–9.

21. Moher D, Pham B, Jones a, et al. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet*. 1998;352(9128):609–13.

22. La Greca AM. Issues in adherence with pediatric regimens. *Journal of pediatric psychology*. 1990;15(4):423–36.

23. Jüni P, Altman DG, Egger M. Assessing the quality of controlled clinical trials. *BMJ.* 2001;323(July):42–46.

24. Becker MH. Patient adherence to prescribed therapies. *Medical care*. 1985;23(5):539–55.

25. Streiner DL, Norman GR. *Health measurement scales: a practical guide to their development and use*. New York: Oxford University Press; 2003:4–226.

26. Higgins J, Altman D, Sterne J. Chapter 8: Assessing risk of bias in included studies. In: In: Higgins JPT GS (editors)., ed. *Cochrane Handbook for Systematic Reviews of Interventions*. 5.1.0 ed. The Cochrane Collaboration; 2011.

27. Matza LS, Yu-Isenberg KS, Coyne KS, et al. Further testing of the reliability and validity of the ASK-20 adherence barrier questionnaire in a medical center outpatient population . *Current medical research and opinion*. 2008;24(11):3197–206.

28. Wetzels GEC, Nelemans PJ, Schouten JS a G, van Wijk BLG, Prins MH. All that glisters is not gold: a comparison of electronic monitoring versus filled prescriptions--an observational study. *BMC health services research*. 2006;6:8.

29. Choo PW, Rand CS, Inui TS, et al. Validation of patient reports, automated pharmacy records, and pill counts with electronic monitoring of adherence to antihypertensive therapy. *Medical care*. 1999;37(9):846–57.

30. Norman GR, Streiner DL. *Biostatistics: The bare essentials*. Shelton: People's Medical Publishing House; 2008:55.

31. McHorney C a. The adherence estimator: a brief, proximal screener for patient propensity to adhere to prescription medications for chronic disease. *Current medical research and opinion*. 2009;25(1):215–38.

32. Tan MY, Magarey JM, Chee SS, Lee LF, Tan MH. A brief structured education programme enhances self-care practices and improves glycaemic control in Malaysians with poorly controlled diabetes. 2011;26(5):896–907.

33. Hederos C-A, Janson S, Hedlin G. Six-year follow-up of an intervention to improve the management of preschool children with asthma. *Acta paediatrica (Oslo, Norway : 1992)*. 2009;98(12):1939–44.

34. Henry A, Batey RG. Enhancing compliance not a prerequisite for effective eradication of Helicobacter pylori: the HelP Study. *The American journal of gastroenterology*. 1999;94(3):811–5.

35. Peveler R, George C, Kinmonth A-L, Campbell M, Thompson C. Effect of antidepressant drug counselling and information leaflets on adherence to drug treatment in primary care: randomised controlled trial. *BMJ*. 1999;319:612–615.

36. Smith H, Hankins M, Hodson A, George C. Measuring the adherence to medication of elderly patients with heart failure: is there a gold standard? *International journal of cardiology*. 2010;145(1):122–3.

37. Santa Helena ETD, Nemes MIB, Eluf-Neto J. Development and validation of a multidimensional questionnaire assessing non-adherence to medicines. *Revista de saúde pública*. 2008;42(4):764–7.

38. van den Boogaard J, Lyimo R a, Boeree MJ, Kibiki GS, Aarnoutse RE. Electronic monitoring of treatment adherence and validation of alternative adherence measures in tuberculosis patients: a pilot study. *Bulletin of the World Health Organization*. 2011;89(9):632–9.

39. Rudd P. In search of the gold standard for compliance measurement. *Archives of internal medicine1*. 1979;139(6):627–628.

40. Copher R, Buzinec P, Zarotsky V, et al. Physician perception of patient adherence compared to patient adherence of osteoporosis medications from pharmacy claims. *Current medical research and opinion*. 2010;26(4):777–85.

41. Rickles NM, Svarstad BL, Statz-paynter JL, Taylor LV, Kobak KA. Pharmacist Telemonitoring of Antidepressant Use: Effects on Pharmacist – Patient Collaboration. *Journal of the American Pharmacists Association*. 2005;45(3):344– 353.

42. Vermeire E, Wens J, Van Royen P, et al. Interventions for improving adherence to treatment recommendations in people with type 2 diabetes mellitus. *Cochrane Library*. 2009;(1).

43. Vogel A. Forget Your Meds ? *Research Horizons*. 2008:36–37.

44. Fielder a R, Irwin M, Auld R, et al. Compliance in amblyopia therapy: objective monitoring of occlusion. *The British journal of ophthalmology*. 1995;79(6):585–9.

45. Rolley JX, Davidson PM, Dennison CR, et al. Medication adherence self-report instruments: implications for practice and research. *The Journal of Cardiovascular Nursing*. 2008;23(6):497–505.

46. Zeller A, Schroeder K, Peters TJ. Electronic pillboxes (MEMS) to assess the relationship between medication adherence and blood pressure control in primary care. *Scandinavian journal of primary health care*. 2007;25(4):202–7.

47. Steiner JF, Prochazka a V. The assessment of refill compliance using pharmacy records: methods, validity, and applications. *Journal of clinical epidemiology*. 1997;50(1):105–16.

48. Bell KJL, Kirby a., Hayen a., Irwig L, Glasziou P. Monitoring adherence to drug treatment by using change in cholesterol concentration: secondary analysis of trial data. *BMJ*. 2011;342(jan21 1):d12–d12.

49. Gonzalez JS, Schneider HE. Methodological issues in the assessment of diabetes treatment adherence. *Current Diabetes Reports*. 2011;11(6):472–9.

50. Ho PM, Bryson CL, Rumsfeld JS. Medication adherence: its importance in cardiovascular outcomes. *Circulation*. 2009;119(23):3028–35.

51. Paterson DL, Potoski B, Capitano B. Measurement of adherence to antiretroviral medications. *Journal of acquired immune deficiency syndromes*. 2002;31:S103–6.

52. Kenna L a, Labbé L, Barrett JS, Pfister M. Modeling and simulation of adherence: approaches and applications in therapeutics. *The AAPS journal*. 2005;7(2):E390–407.

53. Haberer JE, Robbins GK, Ybarra M, et al. Real-time electronic adherence monitoring is feasible, comparable to unannounced pill counts, and acceptable. *AIDS and behavior*. 2012;16(2):375–82.

54. Stephenson BJ, Rowe BH, Haynes RB, Macharia WM, Leon G. Is this patient taking the treatment as prescribed? *JAMA*. 1993;269(21):2779– 2781.

55. Zeller A, Schroeder K, Peters TJ. An adherence self-report questionnaire facilitated the differentiation between nonadherence and nonresponse to antihypertensive treatment. *Journal of clinical epidemiology*. 2008;61(3):282–8.

56. Santschi V, Wuerzner G, Schneider M-P, Bugnon O, Burnier M. Clinical evaluation of IDAS II, a new electronic device enabling drug adherence monitoring. *European journal of clinical pharmacology*. 2007;63(12):1179–84.

57. Nakonezny P a, Byerly MJ, Rush a J. Electronic monitoring of antipsychotic medication adherence in outpatients with schizophrenia or schizoaffective disorder: an empirical evaluation of its reliability and predictive validity. *Psychiatry research*. 2008;157(1-3):259–63.

58. Podsadecki TJ, Vrijens BC, Tousset EP, Rode R a, Hanna GJ. “White coat compliance” limits the reliability of therapeutic drug monitoring in HIV-1-infected patients. *HIV clinical trials*. 2008;9(4):238–46.

59. Cronin TH, Kahook MY, Lathrop KL, Noecker RJ. Accuracy and performance of a commercially available Dosing Aid. *The British journal of ophthalmology*. 2007;91(4):497–9.

60. Perrin K, Williams M, Wijesinghe M, et al. Randomized controlled trial of adherence with single or combination inhaled corticosteroid/long-acting beta-agonist inhaler therapy in asthma. *The Journal of allergy and clinical immunology*. 2010;126(3):505–10.

61. Kimmel SE, Troxel AB, Loewenstein G, et al. Randomized trial of lottery-based incentives to improve warfarin adherence. *American heart journal*. 2012;164(2):268–74.

62. Andrade ASA, Mcgruder HF, Wu AW, et al. A Programmable Prompting Device Improves Adherence to Highly Active Antiretroviral Therapy in HIV-Infected Subjects with Memory Impairment. *HIV/AIDS.* 2005;21205:875–882.

63. Jank S, Bertsche T, Schellberg D, Herzog W, Haefeli WE. The A14-scale: development and evaluation of a questionnaire for assessment of adherence and individual barriers. *Pharmacy world & science : PWS*. 2009;31(4):426–31.

64. Bell DJ, Kapitao Y, Med D, et al. Adherence to Antiretroviral Therapy in Patients Receiving Free Treatment From a Government Hospital. 2007;45(5):2005–2008.

65. Boer IMD, Prins JM, Sprangers MAG, Nieuwkerk PT. Using Different Calculations of Pharmacy Refill Adherence to Predict Virological Failure Among HIV-Infected Patients. 2010;55(5):635–640.

66. Cadarette SM, Burden AM. Measuring and improving adherence to osteoporosis pharmacotherapy. *Current opinion in rheumatology*. 2010;22(4):397–403.

67. Machado CJ, Drew M, Guimarães C. Utilização dos registros de dispensação da farmácia como indicador da não-adesão à terapia anti-retroviral em indivíduos infectados pelo HIV Pharmacy records as an indicator of non-adherence to antiretroviral therapy by HIV-infected patients. 2009;25(3):495–506.

68. Lee JK, Grace KA, Foster TG, et al. How should we measure medication adherence in clinical trials and practice ? *Therapeutics And Clinical Risk Management*. 2007;3(4):685–690.

69. Goldman JD, Cantrell RA, Mulenga LB, et al. Simple adherence assessments to predict virologic failure among HIV-infect adults with disconcordant immunologic and clinical responses to antiretroviral therapy. *AIDS research and human retroviruses*. 2008;24(8):1031– 1035.

70. Kouanfack C, Laurent C, Peytavin G, et al. Adherence to Antiretroviral Therapy Assessed by Drug Level Monitoring and Self-Report in Cameroon. 2008;48(2):216–219.

71. Karter AJ, Parker MM, Moffet HH, et al. New prescription medication gaps: a comprehensive measure of adherence to new prescriptions. *Health services research*. 2009;44(5 Pt 1):1640–61.

72. Cassidy CM, Rabinovitch M, Schmitz N, Joober R, Malla A. A comparison study of multiple measures of adherence to antipsychotic medication in first-episode psychosis. *Journal of clinical psychopharmacology*. 2010;30(1):64–7.

73. Kalichman S, Amaral C, Swetzes C, et al. A simple single item rating scale to measure medication adherence: further evidence for convergent validity. *J Int Assoc Physicians AIDS care (Chic)*. 2009;8(6):367–374.

74. Kalichman SC, Amaral CM, Cherry C, et al. Monitoring medication adherence by unannounced pill counts conducted by telephone: reliability and criterion-related validity. *HIV clinical trials*. 2008;9(5):298–308.

75. Hawks RL, Chiang CN. Urine Testing for Drugs of Abuse. *US Department of Health and Human Services*. 1986:115.

76. Grosset K a, Reid JL, Grosset DG. Medicine-taking behavior: implications of suboptimal compliance in Parkinson's disease. *Movement disorder*. 2005;20(11):1397–404.

77. Sirey JA, Bruce ML, Kales HC. Improving Antidepressant Adherence and Depression Outcomes in Primary Care: The Treatment Initiation and Participation Program. *American journal geriatric psychiatry*. 2011;18(6):554–562.

78. Mannheimer S, Thackeray L, Huppler Hullsiek K, et al. A randomized comparison of two instruments for measuring self-reported antiretroviral adherence. *AIDS care*. 2008;20(2):161–9.

79. Lai PSM, Chua SS, Chew YY, Chan SP. Effects of pharmaceutical care on adherence and persistence to bisphosphonates in postmenopausal osteoporotic women. *Journal of clinical pharmacy and therapeutics*. 2011;36(5):557–67.

80. Mullan RJ, Montori VM, Shah ND, et al. The Diabetes Mellitus Medication Choice Decision Aid. *Archives of internal medicine*. 2009;169(17):1560–1568.

81. Ndubuka NO, Ehlers VJ. Adult patients' adherence to anti-retroviral treatment: a survey correlating pharmacy refill records and pill counts with immunological and virological indices. *International journal of nursing studies*. 2011;48(11):1323–9.

82. Jónsdóttir H, Opjordsmoen S, Birkenaes AB, et al. Medication adherence in outpatients with severe mental disorders: relation between self-reports and serum level. *Journal of clinical psychopharmacology*. 2010;30(2):169–75.

83. Purcell DW, Latka MH, Metsch LR, et al. Results from a randomized controlled trial of a peer-mentoring intervention to reduce HIV transmission and increase access to care and adherence to HIV medications among HIV-seropositive injection drug users. *Journal of acquired immune deficiency syndromes (1999)*. 2007;46 Suppl 2:S35–47.

84. Pearson CR, Simoni JM, Hoff P, Kurth AE, Martin DP. Assessing antiretroviral adherence via electronic drug monitoring and self-report: an examination of key methodological issues. *AIDS and behavior*. 2007;11(2):161–73.

85. Mannheimer SMD, Thackeray L, Hullsiek KH, Chesney M. AIDS Care : Psychological and Socio-medical Aspects of AIDS / HIV A randomized comparison of two instruments for measuring self-reported antiretroviral adherence. 2008;(August 2012):37–41.

86. Rodríguez Chamorro MÁ, García-Jiménez E, Amariles P, Rodríguez Chamorro A, José Faus M. Revisión de tests de medición del cumplimiento terapéutico utilizados en la práctica clínica. *Atención Primaria*. 2008;40(8):413–417.

87. Knobel H, Alonso J, Casado JL, et al. Validation of a simplified medication adherence questionnaire in a large cohort of HIV-infected patients: the GEEMA Study. *AIDS (London, England)*. 2002;16(4):605–13.

88. Stephenson BJ, Rowe BH, Brian R, Macharia WM, Leon G. Is This Patient Taking the Treatment as Prescribed? 7–9.

89. Haynes R, Taylor D, Sackett D, et al. Can simple clinical measurements detect patient noncompliance? *Hypertension*. 1980;2:757–764.

90. Gilbert J, Evans CE, Haynes RB, Tugwell P. Predicting compliance with a regiment of digoxin therapy in family practice. *CMA journal*. 1980;123:119–122.

91. Morisky DE, Green LW, Levine DM. Concurrent and predictive validity of a self-reported measure of medication adherence. *Medical care*. 1986;24(1):67–74.

92. Klein A, Otto G, Kramer I. Impact of a pharmaceutical care program on liver transplant patients' compliance with immunosuppressive medication: a prospective, randomized, controlled trial using electronic monitoring. *Transplantation*. 2009;87:839–847.

93. Krapek K, King K, Warren SS, et al. Medication adherence and associated hemoglobin A1c in type 2 diabetes. *The Annals of pharmacotherapy*. 2004;38(9):1357–62.

94. Koschack J, Marx G, Schnakenberg J, Kochen MM, Himmel W. Comparison of two self-rating instruments for medication adherence assessment in hypertension revealed insufficient psychometric properties. *Journal of clinical epidemiology*. 2010;63(3):299–306.

95. Kane S, Becker B, Harmsen WS, et al. Use of a screening tool to determine nonadherent behavior in inflammatory bowel disease. *The American journal of gastroenterology*. 2012;107(2):154–60.

96. Tzeng JI, Chang C-C, Chang H-J, Lin C-C. Assessing analgesic regimen adherence with the Morisky Medication Adherence Measure for Taiwanese patients with cancer pain. *Journal of pain and symptom management*. 2008;36(2):157–66.

97. Rickles NM, Svarstad BL. Relationships between multiple self-reported nonadherence measures and pharmacy records. *Research in social & administrative pharmacy : RSAP*. 2007;3(4):363–77.

98. Svarstad BL, Chewning B a, Sleath BL, Claesson C. The Brief Medication Questionnaire: a tool for screening patient adherence and barriers to adherence. *Patient education and counseling*. 1999;37(2):113–24.

99. Koneru S, Shishov M, Ware A, et al. Effectively measuring adherence to medications for systemic lupus erythematosus in a clinical setting. *Arthritis and rheumatism*. 2007;57(6):1000–6.

100. Adams J, Scott J. Predicting medication adherence in severe mental disorders. *Acta psychiatrica Scandinavica*. 2000;101(2):119–24.

101. Weiden P, Rapkin B, Mott T, et al. Rating of medication influences (ROMI) scale in schizophrenia. *Schizophrenia bulletin*. 1994;20(2):297–310.

102. Sleath B, Blalock SJ, Stone JL, et al. Validation of a short version of the glaucoma medication self-efficacy questionnaire. *The British journal of ophthalmology*. 2012;96(2):258–62.

103. Gozum S, Hacihasanoglu R. Reliability and validity of the Turkish adaptation of medication adherence self-efficacy scale in hypertensive patients. *European journal of cardiovascular nursing : journal of the Working Group on Cardiovascular Nursing of the European Society of Cardiology*. 2009;8(2):129–36.

104. Fernandez S, Chaplin W, Schoenthaler AM, Ogedegbe G. Revision and validation of the medication adherence self-efficacy scale (MASES) in hypertensive African Americans. *Journal of behavioral medicine*. 2008;31(6):453–62.

105. Wetzels G, Nelemans P, van Wijk B, et al. Determinants of poor adherence in hypertensive patients: development and validation of the "Maastricht Utrecht Adherence in Hypertension (MUAH)-questionnaire". *Patient education and counseling*. 2006;64(1-3):151–8.

106. Lupón J, González B, Mas D, et al. Patients' self-care improvement with nurse education intervention in Spain assessed by the European Heart Failure Self-Care Behaviour Scale. *European journal of cardiovascular nursing : journal of the Working Group on Cardiovascular Nursing of the European Society of Cardiology*. 2008;7(1):16–20.

107. Jaarsma T, Arestedt KF, Mårtensson J, Dracup K, Strömberg A. The European heart failure self-care behaviour scale revised into a nine-item scale (EHFScB-9): a reliable and valid international instrument. *European journal of heart failure*. 2009;11(1):99–105.

108. Song Y, Han H-R, Song H-J, et al. Psychometric Evaluation of Hill-Bone Medication Adherence Subscale. *Asian Nursing Research*. 2011;5(3):183–188.

109. Thompson K, Kulkarni J, Sergejew a a. Reliability and validity of a new Medication Adherence Rating Scale (MARS) for the psychoses. *Schizophrenia research*. 2000;42(3):241–7.

110. Kritikos V, Armour CL, Bosnic-Anticevich SZ. Interactive small-group asthma education in the community pharmacy setting: a pilot study. *The Journal of asthma : official journal of the Association for the Care of Asthma*. 2007;44(1):57–64.

111. Farooq S, Nazar Z, Irfan M, et al. Schizophrenia medication adherence in a resource-poor setting: randomised controlled trial of supervised treatment in out-patients for schizophrenia (STOPS). *The British journal of psychiatry : the journal of mental science*. 2011;199(6):467–72.

112. Herz MI, Lamberti S, Mintz J, et al. A program for relapse prevention in schizophrenia: A controlled study. *Archives of general psychiatry*. 2000;57:277–283.

113. Rodrigues A, Aparecida R, Dantas S. Adaptation and Validation of an Oral Anticoagulation Measurement of Treatment Adherence Instrument. 2010;18(3).

114. Sorensen JL, Haug N a, Delucchi KL, et al. Voucher reinforcement improves medication adherence in HIV-positive methadone patients: a randomized trial. *Drug and alcohol dependence*. 2007;88(1):54–63.

115. Chesney MA, Ickovics JR, Chambers DB. UK AIDS Care : Psychological and Socio-medical Aspects of AIDS / HIV Self-reported adherence to antiretroviral medications among participants in HIV clinical trials : The AACTG Adherence Instruments. 2010;(May):37–41.

116. Simons LE, Gilleland J, Blount RL, et al. Multidimensional Adherence Classification System: initial development with adolescent transplant recipients. *Pediatric transplantation*. 2009;13(5):590–8.

117. Toussi M, Choleau C, Reach G, et al. A novel method for measuring patients' adherence to insulin dosing guidelines: introducing indicators of adherence. *BMC medical informatics and decision making*. 2008;8:55.

118. Galloway GP, Coyle JR, Guillén JE, Flower K, Mendelson JE. A simple, novel method for assessing medication adherence: capsule photographs taken with cellular telephones. *Journal of addiction medicine*. 2011;5(3):170–4.