Published in final edited form as:
J Eval Clin Pract. 2011 June ; 17(3): 429–434. doi:10.1111/j.1365-2753.2010.01444.x.

Combining adjusted and unadjusted findings in mixed research synthesis

Corrine I. Voils, PhD[1], **Jamie L. Crandell, PhD**[2], **YunKyung Chang, PhD RN**[3], **Jennifer Leeman, DrPH Mdiv**[4], and **Margarete Sandelowski, PhD RN FAAN**[5]

[1] Research Health Science Specialist, Durham Veterans Affairs Medical Center and Associate Professor of Medicine, Duke University Medical Center, Durham, North Carolina, USA

[2] Research Assistant Professor, University of North Carolina at Chapel Hill School of Nursing & Department of Biostatistics, Chapel Hill, North Carolina, USA

[3] Social Research Specialist, University of North Carolina at Chapel Hill School of Nursing, Chapel Hill, North Carolina, USA

[4] Research Associate Professor, University of North Carolina at Chapel Hill School of Nursing, Chapel Hill, North Carolina, USA

[5] Boshamer Distinguished Professor, University of North Carolina at Chapel Hill School of Nursing, Chapel Hill, North Carolina, USA

Abstract

Rationale, aims and objectives—Finding ways to incorporate disparate types of evidence into research syntheses has the potential to build a better evidence base for clinical practice and policy. Yet conducting such mixed research synthesis studies is challenging. Researchers have to determine whether and how to use adjusted and unadjusted quantitative findings in combination with each other and with qualitative findings.

Methods—Among quantitative findings, adjustment for confounding, either via study design or statistical analysis, can be a considerable source of heterogeneity. Yet there is no consensus about the best way to synthesize findings resulting from different methods for addressing confounding. When synthesizing qualitative and quantitative findings, additional considerations include determining whether findings are amenable to synthesis by aggregation or configuration, which, in turn, depends on the degree of interpretive transformation of findings.

Results—Qualitative survey findings appear similar in form to unadjusted or minimally adjusted quantitative findings and, when addressing the same relationship, can be summed. More interpreted qualitative findings appear similar in form to adjusted findings found in, for example, structural equation models specifying the relationship among a host of latent variables. An option for synthesis of conceptually similar models is reciprocal translation.

Conclusions—These decisions will ultimately be judged on the meaningfulness of their results to practice or policy.

Keywords

adjustment; confounding; meta-analysis; qualitative research; quantitative research; research synthesis

Correspondence: Dr Corrine I. Voils, HSR&D, 508 Fulton St. (152), Durham, NC 27705, USA, voils001@mc.duke.edu.

Introduction

Combining adjusted and unadjusted findings in mixed research synthesis

Scores of articles and book chapters have been written advocating that systematic reviews include more types of evidence [1,2]. This is in contrast to the modus operandi of the past, in which the criteria for inclusion were so restrictive that far more findings were excluded than included [3,4]. To include more of the available evidence, researchers have developed a range of methods to conduct mixed research synthesis, in which the findings of qualitative, quantitative observational and experimental studies are summed up, integrated, or otherwise assembled via qualitative and/or quantitative methods to serve as a basis for practice or policy [5]. Although advances have been made in conceiving and executing mixed research synthesis studies [2,6–9], they continue to be challenging because they compel researchers to accommodate the range of methodological differences that define them [10].

Among the many decisions that researchers conducting these types of systematic reviews must make – and that have yet to be addressed in the specific context of mixed research synthesis – is whether and how to use adjusted and unadjusted quantitative findings. This decision must take into account not only how similar all of the quantitative findings are, but also how similar the quantitative findings are to the qualitative findings. Relatively little has been written about the issues involved in synthesizing adjusted and unadjusted quantitative findings, let alone with qualitative findings. Accordingly, the goal of this article is to describe those issues. We first describe various ways that quantitative findings are presented in reports, with adjustment achieved via study design or statistical methods, and challenges to combining those findings. We then address additional considerations for combining adjusted and unadjusted quantitative findings with qualitative findings.

Adjustment for confounders as a source of heterogeneity in quantitative studies

When synthesizing quantitative studies, there will generally be variability within and among studies; this variability is referred to as heterogeneity. Heterogeneity, which may be observed or hidden (unobserved), may result from differences in study design, populations, measures, time to follow up, or other person- and study-level characteristics [11]. The source of heterogeneity discussed in this paper is adjustment for confounders. Confounders are observed or unobserved variables that are related to both the independent variable of interest and the outcome variable and, as such, can influence the magnitude of the relationship between the independent variable and outcome.

As explained further below, the effect of confounders can be minimized by study design or statistical methods. All of these methods have the potential to reduce or eliminate bias in the parameter that is being estimated (the average treatment effect or the average treatment effect on the treated), thereby approximating the true treatment effect.

Adjustment inherent in study design

Confounders can be addressed through study designs in which control for confounders is inherent. Consider the randomized study, in which all subjects have an equal probability of being assigned to one of the treatment groups (or conditions). In this design, any systematic bias due to observed or unobserved variables is eliminated. For example, consider a study evaluating the effect of an intervention on medication adherence. If patient characteristics such as sex or health status are related both to patient preference for treatment and to adherence and patients are not randomized, then the estimate of the average treatment effect on the treated will be biased. In such cases, randomizing subjects to the intervention or control group should result in an approximately equal distribution of sex and health status in each of the groups, and confounding is not a concern.

Another study design that may reduce confounding is the case–control study. In a case–control study, a group of subjects with the outcome of interest (cases) is selected first, and then a group of subjects without the outcome (controls) is selected. The controls are generally selected to be comparable with the cases, with the degree of comparability being entirely up to the investigator. At minimum, the control group will be chosen to be similar to the cases based on one or more demographic variables. More complex case–control studies will strive to make the two groups similar on a long list of potential confounders, or even to match controls to cases on all observed independent variables via propensity scores or other methods. The cases and controls are then compared on one or more exposures (i.e. variables that are hypothesized to explain differences in outcomes between cases and controls). For example, a researcher may wish to examine differences in sexual practices between individuals with and without human immunodeficiency virus (HIV). Accordingly, the cases (persons with HIV) would be selected first, followed by controls (persons without HIV). The controls may be selected such that, on average, they are comparable with the group of cases on possible confounders (e.g. age, sex, race, sexual orientation, history of intravenous drug use), or each case may be matched to one or more controls with identical or nearly identical values on the possible confounders. The two groups would then be compared on their sexual practices.

Case–control studies are often used when the outcome is rare and a large cohort would be needed to ensure the presence of cases or when a sample of cases is conveniently available to the investigator. When the degree of matching is low, a case–control study is simply a retrospective cohort study in which the cases are oversampled. For example, cases might be patients at an HIV clinic who report trouble adhering to their medication regimen, and the controls may be a sample of patients at the clinic who report no trouble adhering. This amounts to a cohort study of patients at the clinic, being careful to include those who have trouble adhering. Therefore, when there is no or very little matching on potential confounders, the results of a case–control study can be interpreted as a cohort study, which has no inherent controlling for confounders.

Case–control matching on potential confounders generally minimizes the effect of these confounders but may introduce heterogeneity in the effect of the exposure on the outcome relative to unmatched studies or studies matched on different confounders [12]. For example, if the one or more of the confounders on which matching occurs is very closely related to the exposure(s) of interest, the effect size and study power are likely to be smaller than in an unmatched study because the matching makes the exposure more comparable in the cases and controls. If the matching variables are closely related to the outcome, but not the exposure, matching can actually enhance the power to detect a treatment effect relative to an unmatched scenario [13].

Statistical methods of adjustment

Observed confounders can be addressed statistically, and this is typically done in non-randomized study designs, such as quasi-experimental or observational cohort studies, where there is no inherent control for confounders. When planning such studies, researchers determine possible confounders and assess those during data collection. Then, the effect of the confounders is addressed statistically using (1) multiple regression (in which covariates are entered into the model simultaneously with the independent variable of interest); (2) propensity scores (in which the probability of each subject having the outcome of interest based on the potential confounders is computed and controlled for in the analysis) or propensity score matching (in which the treatment and control groups are matched on propensity scores); or (3) stratified analyses (in which the population is divided into strata of the confounder, and analyses are conducted separately within each stratum). These statistical methods can also be used to analyse data from randomized or case–control studies if the

randomization or matching process fails to equate the distribution of a measured confounder between the groups.

Synthesizing adjusted and unadjusted findings

In a report of study results, the authors may present unadjusted findings only, adjusted findings only, or both. (For the remainder of this paper, we define adjusted findings as those resulting from statistical adjustment during data analysis.) An unadjusted finding is the bivariate relationship between an independent and dependent variable that does not control for covariates or confounders, such as the relationship between intervention type and adherence. Unadjusted findings are often presented for randomized and case–control studies because randomization (in randomized studies) or matching (in case–control studies) is assumed to account for confounders, such that the difference in outcomes between treatment and control groups is due to the treatment and nothing else. Thus, unadjusted findings represent adjustment that is accounted for by study design and not by statistical methods.

In quasi-experimental and cohort studies, unadjusted findings are sometimes presented, but they are generally recognized as having high potential for bias due to confounding. Therefore, for these types of studies, adjusted findings are usually preferred. An example of an adjusted finding would be the relationship between race and medication adherence after controlling for the influence of health status on medication adherence. The covariates may be an entire set selected a priori – according to theoretical or clinical relevance – or may be a subset identified through variable selection methods [14].

There is no documented consensus about whether or how to synthesize adjusted and unadjusted findings in a research synthesis. Several options are available and used in meta-analyses of health data. The simplest option to avoid heterogeneity due to adjustment for confounders is to include only studies with unadjusted findings. This is quite common, especially in randomized studies. In non-randomized studies, covariate adjustment is likely to influence the results to a greater extent, so researchers may choose to include only findings that result from adjustment for a set of pre-specified covariates [15,16]. Although selecting studies this way minimizes heterogeneity, it also reduces the number of studies available for synthesis. In many areas of inquiry, this would result in too few findings to synthesize. In addition, the selected studies may be non-representative and bias the pooled estimate, which can lead to vastly different conclusions [17].

Another option is to include both adjusted and unadjusted results while ignoring potential heterogeneity due to adjustment. For example, in a meta-analysis examining the relationship between the duration of breastfeeding and the risk of overweight in later life, Harder *et al.* [18] included unadjusted results when they were available and adjusted results when they were not. The authors synthesized unadjusted odds ratios from six studies with adjusted odds ratios from 11 studies, pooling them without considering heterogeneity due to covariate adjustment, and finding a robust dose–response relationship between longer duration of breastfeeding and decrease in risk of overweight. Controlling for confounders in observational studies usually reduces the effect size, so heterogeneity should have been considered in this case.

Quigley [19] criticized Harder *et al.* [18] for pooling unadjusted and adjusted results from observational studies with no distinction between the two. She suggested another common technique for dealing with this potential heterogeneity: synthesizing the adjusted and unadjusted findings separately. That is one method of examining heterogeneity due to adjustment while using both unadjusted and adjusted estimates (see Pavia *et al.* [20] for another example). In her own reanalysis of Harder *et al.*'s findings using only adjusted findings, Quigley found that some trends were weaker than Harder *et al.* had concluded.

Quigley thus concluded: 'While breastfeeding appears to be associated with a reduced risk of overweight, it is unclear whether this is due to confounding' (p. 871).

In a meta-analysis examining the effect of prevention interventions on risk behaviours among people living with HIV, Crepaz *et al.* [21] conducted separate analyses on unadjusted and adjusted findings, which yielded the same conclusion, namely, that the interventions reduced unprotected sex and acquisition of sexually transmitted diseases. They concluded that 'the overall effect size was robust as it remained significant when using data … without adjustment for baseline behavior' (p. 3). Difficulty interpreting findings arises, however, when adjusted and unadjusted findings do not support the same conclusion.

A more sophisticated approach to addressing heterogeneity –one that allows inclusion of a greater number of studies – is meta-regression. Meta-regression is an extension of meta-analysis that allows examination of the effect of study-level characteristics on effect size estimates. The meta-regression model is a regression model with the effect size as the outcome, and study characteristics as the covariates. The estimates and tests resulting from this regression show the degree to which heterogeneity is explained by the study characteristics investigated. The fixed effects in the model can include information about adjustment for confounders in the computation of the effect size. If desired, a random effect can be added to the model to account for heterogeneity from unobserved sources. For example, following a meta-analysis of the effect of breast-feeding on obesity in childhood, Arenz and Von Kries [22] conducted a meta-regression that included number of confounding factors adjusted for (<7 vs. ≥7) as a covariate. Because the pooled effect size did not differ when this source of variance was accounted for, they concluded that the number of confounders adjusted for did not affect their meta-analytic results. In another meta-analysis, Horta *et al.* [16] theorized that age, race, or socioeconomic status could influence the relationship between maternal smoking and early weaning. In a meta-regression, they included a categorical covariate – coded as *no*, *partial* or *full adjustment* – for their variables of interest. They concluded that adjustment for confounding accounted for a sizable proportion (14.4%) of heterogeneity in the pooled effect size. When the sample size (i.e. number of studies included in the systematic review) is large enough for meta-regression, the impact of adjustment on the effect size should certainly be investigated.

When accounting for covariate adjustment in meta-regression, study design should be considered. Table 1 summarizes the features of each study design type and the corresponding properties of adjusted and unadjusted effect sizes. In randomized studies, covariate adjustment compensates for random covariate imbalance [23,24], whereas in observational studies, it compensates for non-random (systematic) imbalance. In randomized studies, both the unadjusted and adjusted effect sizes are unbiased, although the adjusted effect size is likely to be more precise. In case–control studies, covariate matching to reduce confounding may or may not occur during sample selection, and covariate adjustment may or may not occur during data analysis. Given the observational nature of case–control studies, covariate matching and adjustment in case–control studies should be treated the same as for other non-randomized studies in a meta-regression. In observational studies, the unadjusted effect size is generally biased (specifically, artificially high) as the result of confounding and will be reduced by adjustment for confounders.

Discussion

Combining interpretively diverse qualitative with quantitative adjusted and unadjusted findings

Deciding whether and how to use adjusted versus unadjusted quantitative findings in mixed research synthesis studies entails additional considerations. One such consideration is

determining how to synthesize qualitative and quantitative findings. Synthesis by aggregation entails the assimilation of findings considered to indicate the same thing about two factors (e.g. all findings indicating that lack of trust is connected to non-adherence are pooled and all findings indicating no connection between these factors are pooled) [7]. In contrast, synthesis by configuration entails the assembly or arrangement of disparate but complementary findings into coherent theoretical or narrative renderings of them [5].

Whether findings are viewed as amenable to synthesis by aggregation or configuration depends on researchers' assessments of the degree of interpretive transformation of the qualitative findings. Qualitative findings at the data-near end of the data transformation continuum are typically produced in qualitative surveys or basic qualitative descriptive studies [25] from individual or focus group interview data. Such findings are closer to the data as given by participants and are therefore the least interpreted of qualitative findings and most similar in form (in comparison with more interpreted qualitative findings) to the findings in quantitative surveys or basic descriptive/correlation studies. In contrast to data-near qualitative findings, findings on the data-far end of the continuum are highly interpreted (e.g. grounded theories, phenomenological descriptions or narrative explanations). Such findings are configurations of data from interviews, observations, documents and other sources to capture the 'multivariate … nature' (p. 129) of target events or phenomena: to grab complexity, context and causality [26].

Quantitative findings may also be differentiated by degree of interpretive transformation. Whereas the unadjusted regression model provides population estimates, adding covariates makes the model more subject-specific [27]. Thus, unadjusted findings are comparable with the least interpreted qualitative findings indicating a range of themes associated with an experience expressed in a total sample (e.g. a list of all the factors participants indicated facilitated or hindered medication adherence).

In contrast, adjusted quantitative findings are represented by regression models encompassing a single outcome and multiple covariates, mediators or moderators, and/or models of complex inter-relationships among latent variables (i.e. as in structural equation modelling). In such models, each subject's values on characteristics included in the model can be used to predict that subject's value on the outcome(s). Thus, they are comparable with the most interpreted qualitative findings – best exemplified in the grounded theory conditional matrix – which models the causes, context, contingencies, consequences, covariances and conditions surrounding a target sociocultural process.

Aggregating minimally interpreted qualitative findings with unadjusted quantitative findings

Minimally interpreted qualitative and quantitative findings are assimilable when they address the same relationship. For example, in the domain of medication adherence, both qualitative and quantitative descriptive studies may yield sets of factors that are associated with more or less success adhering to a medication regimen. Although the qualitative findings will typically have been produced from minimally structured and open-ended interviews allowing a wider range of responses than the highly structured and closed-ended questionnaires typically associated with quantitative research, the findings can be transformed into an $X \leftrightarrow Y$ relationship (e.g. that lack of trust in providers was a factor associated with or contributing to non-adherence). Conceptually, this is similar to the unadjusted relationship between trust and non-adherence. Accordingly, unadjusted quantitative findings will be more comparable than adjusted findings to qualitative findings in the form of surveys or basic descriptions and are therefore amenable to pooling.

In cases where numerical data are available from qualitative studies, for example, the number of subjects reporting that lack of trust in providers was a factor in their non-adherence, this number can be pooled directly with the rates/counts from unadjusted quantitative findings. In the more typical cases where no numerical data are available in reports of the qualitative studies (or from their authors), an estimate of that number can be derived and probable ranges calculated, as described in Chang *et al.* [28]. As further described in Voils *et al.* [6] these ranges can then be synthesized with the unadjusted results from the quantitative studies to estimate the probability that, for a given research participant, lack of trust in provider was associated with decreased adherence. Alternatively, the numerical results (unadjusted and/or adjusted) of the quantitative papers can be converted into themes, with each report categorized according to the presence or absence of themes. As described in Crandell *et al.* [29], a Bayesian data augmentation method may then be used to synthesize the findings.

Configuring highly interpreted qualitative findings with adjusted quantitative findings

Synthesizing (by aggregation) minimally interpretive qualitative and quantitative findings is possible, albeit challenging. Yet, even more challenging is the effort to synthesize highly interpreted qualitative and quantitative findings. One option is to translate these interpretations into higher order terms that encapsulate all lower order terms, described as reciprocal translation in Noblit and Hare [30], and illustrated in Sandelowski & Barroso [31] and Pound *et al.* [32]. For example, a qualitative study may indicate that having children will promote adherence in HIV-positive women when women perceive children as reasons to live and fight the infection, but promote non-adherence when childcare distracts from self-care. In quantitative terms, this would be conceptually comparable with a model showing a moderation relationship (which is higher order than a main effect) whereby the association between having children and adherence is moderated by the circumstances surrounding motherhood. Similarly, a grounded theory specifying the dynamics of a decision-making process may be conceptually similar to a structural equation model of this process specifying the relationship among latent variables. (Indeed, a structural equation model may be used statistically to test hypotheses specified by such a grounded theory).

Yet, the likelihood of having two or more models (produced from qualitative and/or quantitative findings) conceptually similar enough to be translated into each others' terms will be low. Indeed, in the case of quantitatively produced models, researchers typically test them in a piecemeal fashion, as when the relationship between attitudes towards a behaviour and behavioural intentions is examined rather than the entire Theory of Planned Behavior [33]. Accordingly, a more likely occurrence will be that parts of different models will conceptually overlap. In that event, it may be possible to configure a model at a level of abstraction that will encompass both overlapping and unique components of these diverse models.

Conclusion

Methods that allow synthesis of evidence across a range of sources have the potential to allow inclusion of a greater number of findings, thereby building a richer base of evidence to guide policy and practice. Mixed research synthesis studies require accommodating the methodological diversity that defines them in ways that preserve the informational value of these studies while adhering to sound methodological principles. Determining whether and how to use adjusted and unadjusted quantitative findings with each other and with qualitative findings requires researchers to make judgments about the comparability in both content and form of findings. These decisions will ultimately be judged on the meaningfulness of their results to practice or policy.

Acknowledgments

This article was supported by a National Institute of Nursing Research, National Institutes of Health (5R01NR004907, June 3, 2005-March 31, 2010) grant ('Integrating qualitative & quantitative research findings'), and with resources and facilities at the Veterans Affairs Medical Center in Durham, NC. Views expressed in this article are those of the authors and do not necessarily represent the Department of Veterans Affairs.

References

1. Pawson, R. Evidence-Based Policy: A Realist Perspective. London: Sage; 2006.
2. Pope, C.; Mays, N.; Popay, J. Synthesizing Qualitative and Quantitative Health Evidence. Berkshire: Open University Press; 2007.
3. MacLure M. Clarity bordering on stupidity: where's the quality in systematic review? Journal of Education Policy. 2005; 20:393–416.
4. Sandelowski M. Reading, writing, and systematic review. Journal of Advanced Nursing. 2008; 64:104–110. [PubMed: 18721156]
5. Sandelowski M, Voils CI, Barroso J. Defining and designing mixed research synthesis studies. Research in the Schools. 2006; 13:29–40. [PubMed: 20098638]
6. Voils CI, Hassselblad V, Crandell J, Chang Y, Lee E, Sandelowski M. A Bayesian method for the synthesis of evidence from qualitative and quantitative reports: the example of antiretroviral medication adherence. Journal of Health Services Research & Policy. 2009; 14 (4):226–233. [PubMed: 19770121]
7. Voils CI, Sandelowski M, Barroso J, Hasselblad V. Making sense of qualitative and quantitative findings in mixed research synthesis studies. Field Methods. 2008; 20 (1):3–25. [PubMed: 18677415]
8. Dixon-Woods M, Agarwal S, Jones D, Young B, Sutton A. Synthesising qualitative and quantitative evidence: a review of possible methods. Journal of Health Services Research & Policy. 2005; 10 (1):45–53. [PubMed: 15667704]
9. Boaz A, Ashby D, Denyer D, Egan M, Harden A, Jones DR, Pawson R, Tranfield D. A multitude of syntheses: a comparison of five approaches from diverse policy fields. Evidence & Policy. 2006; 2 (4):479–502.
10. Sandelowski M, Voils CI, Barroso J. Comparability work and the management of difference in research synthesis studies. Social Science & Medicine. 2007; 64 (1):236–247. [PubMed: 17029691]
11. Cooper, H. Synthesizing Research: A Guide for Literature Reviews. 3. Thousand Oaks, CA: Sage; 1998.
12. Kim I, Cohen ND, Carroll RJ. Effect heterogeneity by a matching covariate in matched case-control studies: a method for graphs-based representation. American Journal of Epidemiology. 2002; 156 (5):463–470. [PubMed: 12196316]
13. Hansson L, Khamis HJ. Matched samples logistic regression in case-control studies with missing values: when to break the matches. Statistical Methods in Medical Research. 2008; 17 (6):595–607. [PubMed: 18375456]
14. Cohen, J. Statistical Power Analysis for the Behavioral Sciences. 2. Hillsdale, NJ: Erlbaum; 1988.
15. Arenz S, Ruckerl R, Koletzko B, von Kries R. Breast-feeding and childhood obesity – a systematic review. International Journal of Obesity and Related Metabolic Disorders. 2004; 28 (10):1247–1256. [PubMed: 15314625]
16. Horta BL, Kramer MS, Platt RW. Maternal smoking and the risk of early weaning: a meta-analysis. American Journal of Public Health. 2001; 91 (2):304–307. [PubMed: 11211645]
17. Colditz GA, Burdick E, Mosteller F. Heterogeneity in meta-analysis of data from epidemiologic studies: a commentary. American Journal of Epidemiology. 1995; 142 (4):371–382. [PubMed: 7625401]
18. Harder T, Bergmann R, Kallischnigg G, Plagemann A. Duration of breastfeeding and risk of overweight: a meta-analysis. American Journal of Epidemiology. 2005; 162 (5):397–403. [PubMed: 16076830]

19. Quigley MA. Re: 'Duration of breastfeeding and risk of overweight: a meta-analysis'. American Journal of Epidemiology. 2006; 163(9):870–872. author reply 2–3. [PubMed: 16554341]
20. Pavia M, Pileggi C, Nobile CG, Angelillo IF. Association between fruit and vegetable consumption and oral cancer: a meta-analysis of observational studies. The American Journal of Clinical Nutrition. 2006; 83 (5):1126–1134. [PubMed: 16685056]
21. Crepaz N, Lyles CM, Wolitski RJ, Passin WF, Rama SM, Herbst JH, Purcell DW, Malow RM, Stall R. HIV/AIDS Prevention Research Synthesis (PRS) Team. Do prevention interventions reduce HIV risk behaviours among people living with HIV? A meta-analytic review of controlled trials. AIDS (London, England). 2006; 20 (2):143–157.
22. Arenz S, Von Kries R. Protective effect of breast-feeding against obesity in childhood: can a meta-analysis of published observational studies help to validate the hypothesis? Advances in Experimental Medicine and Biology. 2009; 639:145–152. [PubMed: 19227541]
23. Senn S. Methods for assessing difference between groups in change when initial measurement is subject to intra-individual variation. Statistics in Medicine. 1994; 13 (21):2280–2285. [PubMed: 7846427]
24. Pocock SJ, Assmann SE, Enos LE, Kasten LE. Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems. Statistics in Medicine. 2002; 21 (19):2917–2930. [PubMed: 12325108]
25. Sandelowski, M.; Barroso, J. Handbook for Synthesizing Qualitative Research. New York: Springer; 2007.
26. Tufte, E. Beautiful Evidence. Cheshire, CT: Graphics Press; 2006.
27. Hauck WW, Anderson S, Marcus SM. Should we adjust for covariates in nonlinear regression analyses of randomized trials? Controlled Clinical Trials. 1998; 19 (3):249–256. [PubMed: 9620808]
28. Chang Y, Voils C, Sandelowski M, Hasselblad V, Crandell J. Transforming verbal counts in reports of qualitative descriptive studies into numbers. Western Journal of Nursing Research. 2009; 31 (7):437–452. [PubMed: 19176403]
29. Crandell J, Voils CI, Chang Y, Sandelowski M. Bayesian data augmentation methods for the synthesis of qualitative and quantitative research findings. Quality and Quantity. 2010 Published online October 20, 2010.
30. Noblit, G.; Hare, R. Meta-Ethnography: Synthesizing Qualitative Studies. Newbury Park: Sage; 1988.
31. Sandelowski M, Barroso J. Motherhood in the context of maternal HIV infection. Research in Nursing & Health. 2003; 26:470–482. [PubMed: 14689463]
32. Pound P, Britten N, Morgan M, Yardley L, Pope C, Daker-White G. Resisting medicines: a synthesis of qualitative studies of medicine taking. Social Science & Medicine. 2005; 61:133–155. [PubMed: 15847968]
33. Ajzen I. The theory of planned behavior. Organizational Behavior and Human Decision Processes. 1991; 50:179–211.
34. Ford I, Norrie J. The role of covariates in estimating treatment effects and risk in long-term clinical trials. Statistics in Medicine. 2002; 21 (19):2899–2908. [PubMed: 12325106]

Table 1

Controlling for confounders in three study designs

Study design	Inherent control for confounders	Interpretation of unadjusted effect size	Interpretation of adjusted effect size
Randomized	Controls for observed and, by assumption, unobserved confounders	Effect size is unbiased estimate of unconfounded population effect size	Effect size is unbiased estimate of unconfounded population effect size Power of the test of treatment effect may increase*
Case–control	Controls for observed confounders on which groups are matched. Does not control for unmatched or unobserved confounders	Effect size is unbiased estimate of population effect size if groups are matched on all observed confounders	Effect size is unbiased if all observed confounders are adjusted for either by matching or in analysis
Quasi-experimental or observational cohort	No inherent control for confounders	Effect size is biased, and usually artificially large, except in the rare case in which there are no confounders (observed or unobserved)	Effect size is unbiased if adjustment is for all confounders and there are no unobserved confounders

* In a linear model, this increase in power is due to the potential increase in precision of the effect size estimate. In a non-linear model (e.g. logistic regression, Poisson regression, Cox models), adjustment for covariates in a randomized study cannot increase (and is likely to decrease) the precision of the treatment effect, but it can be expected to increase the treatment effect size in randomized trials, leading to increased power [27,34].