Determining if disease management saves money: an introduction to meta-analysis

Ariel Linden DrPH MS[1,2] and John L. Adams PhD[3]

[1]President, Linden Consulting Group, [2]Adjunct Professor, Oregon Health and Science University, School of Medicine and School of Nursing, Portland, OR, USA
[3]Senior Scientist, RAND Corporation, Santa Monica, CA, USA

Abstract

Disease management (DM) programmes have long been promoted as a major medical cost-saving mechanism, even though the scant research that exists on the topic has provided conflicting results. In a 2004 literature review, the Congressional Budget Office stated that 'there is insufficient evidence to conclude that disease management programs can generally reduce the overall cost of health care services'. To address this question more accurately, a meta-analysis was warranted. Meta-analysis is the quantitative technique used to pool the results of many studies on the same topic and summarize them statistically. This method is also quite suitable for individual DM firms to assess whether their programmes are effective at the aggregate level. This paper describes the elements of a rigorous meta-analytic process and discusses potential biases. A hypothetical DM organization is then evaluated with a specific emphasis on medical cost-savings, simulating a case in which different populations are served, evaluation methodologies are employed, and diseases are managed.

Introduction

At the bequest of congress, the Congressional Budget Office (CBO) conducted a review of the disease management (DM) literature in 2004 to address the question of whether or not DM programmes can reduce the overall cost of health care and how such the programmes might apply to Medicare. The report concluded that, while DM appears to have clinical value for patients and results in high satisfaction levels, 'there is insufficient evidence to conclude that disease management programs can generally reduce the overall cost of health care services' [1]. These findings prompted an immediate response from several DM industry advocates [2–5], suggesting that the review was not robust and did not include several recent peer-reviewed papers which, in fact, did demonstrate cost-savings.

Given that DM has been promoted as a major medical cost-saving mechanism and, subsequently, has witnessed tremendous growth, the furor raised over this report is understandable. In 2003, 58% of self-insured employers offered DM programmes to their employees – an upsurge from 41% in 2002 [6]. Increasingly, states are implementing DM for their Medicaid populations and Medicare is currently initiating the Chronic Care Improvement Program as a component of the Medicare Prescription Drug, Improvement, and Modernization Act of 2003. Specialty DM companies' annual revenues have increased from $85 million in 1997 to more than $600 million in 2002 [7]. Profits for American Healthways, one of the largest DM firms in the United States,

rose nearly 59% in the second quarter of 2005 alone [8]. Thus, the industry has a vested interest in demonstrating to current and prospective clients that medical cost-savings as a result of DM are achievable.

Had the CBO approached this assignment in a more rigorous manner, the results may have been viewed as being more credible. Given these circumstances, the more appropriate approach would have been to conduct a systematic review including a meta-analysis. A systematic review entails following a structured process to synthesize information from several independent research studies to determine where findings are consistent and outcomes can be considered generalizable. Meta-analysis is the quantitative technique used to pool the results of many studies on the same topic and summarize them statistically. These two terms are used synonymously and interchangeably as both are typically performed together (however, they are mutually exclusive and can be performed independently).

The most fundamental reason for using a meta-analytic approach to evaluate the effectiveness of DM in reducing costs is that the current body of literature is inconclusive. Whereas some studies show positive cost-saving results, there is a similar number of studies that do not. Moreover, many studies have inherent methodological flaws, either in their design or in the interpretation of their results [9–13]. Consequently, one cannot necessarily rely on the findings of any single independent study or even several studies to provide guidance as to whether DM is a cost-effective addition to the usual care process.

In addition to the CBO report, two other reviews have addressed the economic value of DM programmes, with neither showing much benefit [14,15]. To date, only one rigorous meta-analysis on this topic has been conducted. Krause investigated the economic effectiveness of DM [direct economic outcomes were defined as the measurement of the number, average or rate of hospitalization, outpatient visits, emergency department (ED) visits and medical costs] and found a statistically significant overall weighted effect size of 0.311 [95% confidence intervals (CI) = 0.272–0.350] indicating that the DM programmes reviewed were economically effective [16]. The disparity between the findings of the review articles and the meta-analysis serves to only further confuse the issue of whether DM has any cost-saving potential.

A similar problem exists at the individual DM level for firms which implement programmes in disparate populations or utilize a variety of evaluation methodologies to assess the effectiveness of their programmes. For example, it is not uncommon for a DM organization to provide services to an array of customers such as commercial health plans, Medicaid and Medicare. Some health plans may require an actuarial approach to evaluate the programme's effectiveness, Medicare may request a randomized controlled trial (RCT), and some states may insist upon an econometric approach for their Medicaid programme evaluation. Given the differences among populations served and the evaluation tools used, it is probable that some programmes may be found to be more effective than others. In this situation, a meta-analysis could be considered the most appropriate method for determining if the programme is effective overall.

The purpose of this paper is to introduce readers to the meta-analytic approach in assessing DM programme effectiveness, with a specific emphasis on medical cost-savings. The elements of a rigorous systematic review process will be presented and potential biases will be discussed. Finally, using the meta-analytic techniques presented, a hypothetical DM organization will be evaluated, simulating a case in which different populations are served (e.g. Medicaid, Medicare and commercial health plans), evaluation methodologies are employed (pre-post total population approach, matched case-controls and the RCT), and diseases or conditions are managed [diabetes, asthma, congestive heart failure (CHF) and coronary artery disease (CAD)].

History of meta-analysis

Combining results of different studies to determine the weighted-average effect is by no means a recently developed concept. Roger Cotes, an English mathematician, averaged measurements made by different astronomers as early as the 18th century. Karl Pearson is usually credited as being the first researcher to use formal methods for pooling data from independent studies when he averaged the correlation between mortality and inoculation against enteric fever across five communities to determine the preventive effects of the inoculation in 1904 [17]. In 1932, Ronald Fisher described a method for combining probabilities from tests of significance across separate studies, which he included in the fourth edition of his manuscript *Statistical Methods for Research Workers* [18]. A few years later, Yates and Cochran [19] developed statistical methods for grouping outcomes of agricultural experiments.

It was not until 1955 that the first application of meta-analysis appeared in the medical literature [20]. In this study, Beecher investigated the impact of the placebo effect across various diagnoses. Interestingly enough, 35% of patients on the placebo appeared to improve. This topic is still the focus of much scrutiny as illustrated in a 1998 meta-analysis by Kirsch and Saperstein [21], which found that 25% of patients responded favourably (as indicated by reduced depression) to the placebo as compared with patients on an actual antidepressant.

The term 'meta-analysis' was actually coined by Gene Glass in 1976 [22] while describing a novel approach for weighting the effect size of individual studies in assessing a cumulative impact of a given treatment. Glass developed these techniques while determining the effectiveness of psychotherapy. This marked a turning point in the use of meta-analysis and it soon gained widespread adoption in many areas of research, particularly in the field of medicine.

Perhaps the most significant contribution in this area was the development of the Cochrane Collaboration. In a 1979 essay [23], Archie Cochrane, a trained doctor and well-respected health services researcher explored the medical field to invest in the initiation of new RCT and maintain regular systematic reviews. He specifically challenged the obstetrics specialty, suggesting that they were lax in their research endeavours. Over the next 10 years, several specialists in the field of obstetrics took up the challenge, resulting in a book of systematic reviews of topics relating to pregnancy and childbirth [24]. The Cochrane Collaboration was established in 1993, resulting in the coordination of groups of individuals worldwide to prepare and maintain systematic reviews of specific areas of health. An outgrowth of the collaboration was the development of the Cochrane Library, which provides access to all new and revised systematic reviews. In the most recent quarterly update (issue 3, published on 20 July 2005), 4041 systematic reviews were available in the Cochrane Database [25].

In summary, meta-analysis has a long and distinguished history. Over the past 30 years, it has evolved to become an important research method in the field of health care, providing important information about interventions that are effective, ineffective, or even harmful. Consequently, the results of meta-analyses serve as the foundation for evidence-based medicine.

Structured approach to conducting a meta-analysis

Conventional literature reviews, such as the CBO report described earlier, are subjective and thus likely to be faulted for their inherent biases. Common concerns include the selection of studies that support popular opinion or the author's position on the topic. For example, if the author has published manuscripts in the specific area under review, we may find that the review contains studies with similar or supported findings. Additionally, direct or indirect funding for the review by a stakeholder in the results may lead to biased reporting. Finally, it is not enough to tally the number of studies with positive results and the number with negative results to determine which outcome is supported by the weight of the evidence. Indeed, factors such as study design, sample size and effect size must be considered. Without a structured or systematic approach, the conclusions drawn from the review may be
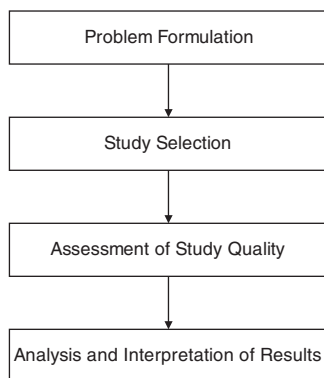
Figure 1 A systematic approach for conducting a rigorous meta-analysis of the literature [23].

unfounded, even if the individual studies on which the review relies are valid and reliable.

The Cochrane Collaboration has developed (and updates regularly) a reviewers' handbook [26] that provides a comprehensive guide for researchers interested in conducting meta-analyses. While several of the sections present specific guidelines for submitting reviews to the Cochrane Library, the overall manual provides excellent instruction in the development of a thoughtful and robust systematic review. The major steps in the meta-analytic process are shown in Fig. 1 and include: (1) Problem Formulation; (2) Study Selection; (3) Assessment of Study Quality; and (4) Analysis and Interpretation of Results. As illustrated, these explicit procedures are similar to those needed for conducting any individual research study, thereby allowing other researchers to replicate the exact protocols in their own activities.

Problem formulation

A clearly defined research question is integral to any investigation, but to meta-analysis in particular. Individual studies will naturally vary in their participant's characteristics, interventions, outcomes and study designs. A meta-analyst will have to either narrow the focus of the review or perform subgroup analyses if the data permit. For example, the CBO report asked the question: 'Do DM programs reduce the overall cost of health care services?' To address this question, the CBO focused on three diseases with high prevalence in the Medicare population: CHF, CAD and diabetes. However, the analysis did not qualify the types of interventions provided, and the outcomes chosen were much broader than just cost-savings (which were measured differently in each study as well). Moreover, given that the question was whether or not DM is applicable to the Medicare population, particular emphasis should have been placed on participants in that age group rather than making generalizations across a wide range of ages without the benefit of a statistical analysis.

A more comprehensive approach to address the research question of whether DM programmes are economically effective was demonstrated in a recent meta-analysis by Krause [16]. In this systematic review, the author (1) selected heart disease, asthma and diabetes as the diseases under study; (2) classified interventions as self-managed, nurse-managed or team-managed; (3)

incorporated levels of disease severity; and (4) included all study designs but analysed separately to limit the threats to internal validity. A statistically significant overall weighted effect size of 0.311 was reported (95% CI = 0.272–0.350) indicating that in the 67 studies reviewed, DM programmes were economically effective. After adjusting for disease severity, no statistically different effect sizes were noted for study design, disease type or intervention modality.

Study selection

The selection process of suitable studies for a meta-analysis is quite involved. A computerized search of medical literature databases such as MEDLINE, HEALTHSTAR, Cochrane and EMBASE is central to this process and is typically performed by a research-trained librarian. Specific search terms are used in the medical subject heading to generate a list of relevant studies. Titles, abstracts and articles are then reviewed manually to ensure that they meet inclusion criteria (e.g. in a meta-analysis on cost-savings, medical and administrative costs as well as financial savings must be reported in order to be included in the meta-analysis). Other sources of relevant data should also be searched. For example, conference proceedings typically present abstracts of studies in progress or otherwise unpublished data. These researchers may be willing to provide their data for the meta-analysis or point the reviewer to additional sources of information.

Additional decisions that must be made at this juncture include determining if one or more reviewers will be used, if reviewers chosen will be content area experts or non-experts, if reviewers will be blinded to information that may introduce bias, and how disagreements between reviewers will be handled [23]. Finally, it is important to document all studies which are either accepted or rejected for the review, preferably via a data collection form or database.

Assessment of study quality

Perhaps the most important aspect of any meta-analysis is assessing the quality of each and every individual study eligible to be included in the review. One cannot assume that the peer-review process is a sufficient means to identify issues that impact the validity of study results. Therefore, the meta-analyst must conduct an independent review of each manuscript to determine the potential impact of biases, whether the appropriate statistical analyses were used, and if the study appears generalizable across people, settings, treatments or outcomes [27–29]. In order to minimize the likelihood of including studies with inherent threats to internal validity, most meta-analyses rely strictly on research that implemented the RCT design. However, there are several robust quasi-experimental designs [30,31] and statistical techniques [32–34] that should be considered for inclusion in a meta-analysis if used in an individual research study.

Analysis and interpretation of results

The statistical analysis component of a meta-analysis is what differentiates a systematic review from a conventional literature review. It involves standardizing the outcome measures for each individual study, weighting and pooling all outcomes to attain an

average treatment effect across all studies, and evaluating whether the results may be impacted by confounding variables or bias, via various tests and sensitivity analysis. Upon completion of this process, this information is then used to interpret the findings and draw conclusions.

Standardizing outcome measures

This procedure essentially performs comparisons of outcomes between treatment and control groups for each individual study. The difference in the outcome between the two groups is called the treatment effect and, logically enough, the magnitude of that difference is referred to as the effect size. Continuous variables are standardized by dividing the treatment effect by the standard deviation. Dichotomous variables (e.g. disease/no disease, treatment/no treatment) are reported as odds ratios or relative risk. These conversions provide a convenient method for standardizing treatment effects to make them more comparable from study to study.

Average pooled effect

Once the standardized statistic is calculated for each study, the pooled or overall treatment effect can then be estimated. Here again, a weighting must occur to account for differences in the size of various study populations. Thus, in the weighted average treatment effect equation, larger studies with smaller standard errors are given more influence than smaller studies with more variability. Depending on the amount of variability that exists between studies, the meta-analyst is offered two techniques to choose from – the 'fixed effects' model [35] or the 'random effects' model [36].

The fixed effects model assumes that no heterogeneity exists in the treatment effect across studies (other than that due to chance), and is therefore 'fixed'. In other words, different studies report similar effect sizes across similar types of persons, diseases, interventions, settings, etc. When large variability across studies cannot readily be explained, the random effects model is typically chosen. This model treats the heterogeneity in the treatment effects across studies as random, and should therefore form a distribution with the centre point indicating the average effect. This topic will be further elaborated in the following sections.

There is some controversy as to which method is superior and each technique has its limitations. The fixed effects model is criticized for assuming that variability across studies is due to chance when, in fact, any number of biases could be the source of that variation. The random effects model is criticized for theorizing that the studies were 'sampled' and therefore accurately represent a population of studies where the true treatment effect varies [37]. Given that these two techniques are premised on different assumptions, it is somewhat surprising that they obtain dissimilar results only when studies are distinctly heterogeneous [38]. Tests for heterogeneity, such as the Q non-combinability statistic, chi-squared test or I^2 statistic [39], are typically provided in meta-analysis software programmes [40]. The resulting values can aid the analyst in determining which effects model to apply to the data.

Forest plots

A graphical display of the data via a forest plot is an extremely useful means of visually inspecting information processed in the meta-analysis. Figure 2 presents results of a hypothetical meta-analysis. Each study included in the meta-analysis is listed along the Y-axis, together with its sample size. Each corresponding square represents the point estimate for the study with a horizontal line running through it to signify the CI. The size of each square is indicative of the weight it was given in the meta-analysis. In this example, Study C was given more weight than Study D. The overall weight-averaged effect size is plotted at the bottom of the plot and the vertical line stemming out of it provides a visual reference for the plots above. As shown, nearly all squares are on or near the overall effect size line, indicating good homogeneity in the effect sizes across the individual studies. The lines representing the CI allow for visual determination of statistical significance, with those crossing over the 0 intercept indicating no treatment effect between cases and controls. In this example, Studies C and F show no statistically significant effects while all other studies (including the overall) show statistically significant results. This graph gets its name from the fact that when all the values are plotted, the display bears resemblance to a tree [41].

Funnel plots

Figure 3 illustrates another simple yet effective graphic display of meta-analytic data called a funnel plot. In this display, each individual study's treatment effect (X-axis) is plotted against its sample size (Y-axis). One of the basic tenets of statistics is that precision will naturally improve, as expressed by a narrowing of the standard deviation around the mean, as sample size increases. In the absence of bias, the display will resemble an inverted symmetrical funnel with smaller studies (and their larger standard
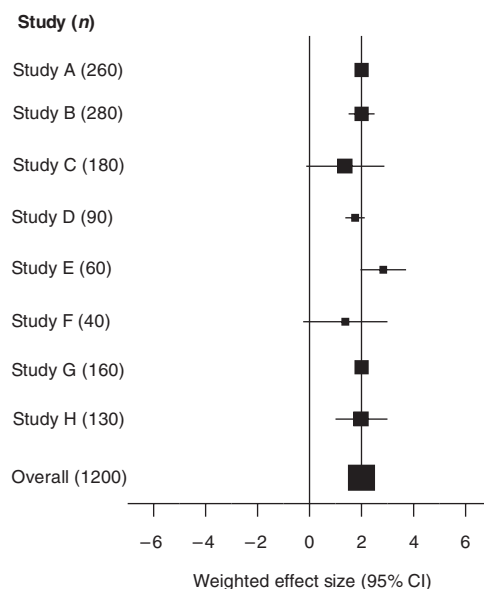
Figure 2 An example of a forest plot for presenting results from a meta-analysis (see text for explanation). CI, confidence interval.
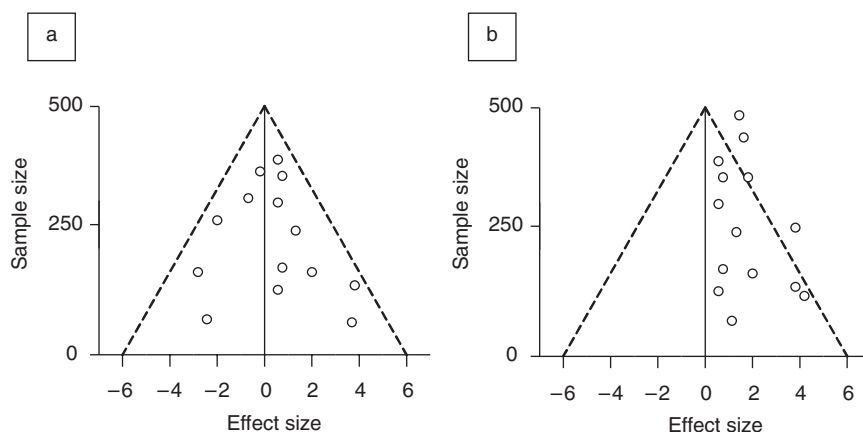
Figure 3 Illustration of a funnel plot for a meta-analysis (a) free of publication bias, and (b) exhibiting publication bias.

deviations) distributed near the bottom of the graph and larger studies (with their smaller standard deviations) clustered near the top of the graph (as illustrated in Fig. 1a) [42]. A major threat to the validity of findings in meta-analysis is that of publication bias. Publication bias occurs as a result of the tendency for investigators to submit to the peer-review process, only those studies that have achieved significant and positive outcomes [43]. If publication bias exists, most values on the funnel would be located to the right of the 0 intercept, indicating that all individual studies reported positive treatment effects (Fig. 1b).

Biases associated with individual study methodologies may also be identified in visual inspection of the funnel plot. DM evaluations typically rely on observational designs, which are susceptible to threats to internal validity. If the relationship between sample size and precision is not maintained across studies, pooled results may exhibit the same graphic form as that of publication bias and thus must be viewed with caution [26,44].

Sensitivity analysis

When used together, the forest and funnel plots assist the analyst in determining how robust the findings are in respect to various elements of the meta-analytic methodology. Some factors that could be investigated in DM outcome studies include the impact of study design, diseases or conditions, interventions, severity and choice of statistical method employed in the meta-analysis. If the sensitivity analysis does not change the results substantially, more confidence can be placed in these findings. Conversely, if the results change the conclusions drawn from the study, greater caution should be taken in interpreting the outcomes [23].

In summary, there are several steps involved in the analysis and interpretation of data in a meta-analysis. The treatment effect for outcomes in individual studies are first converted to a standardized metric so that an overall treatment effect can be calculated when all study outcomes are pooled. Depending upon the amount of heterogeneity between studies, either a fixed or random effects model will be chosen for providing a weighted average treatment effect. There are statistical tests available to measure whether heterogeneity and biases are of concern, and these can be visually displayed using the forest and funnel plots. Testing basic assumptions of the review process via sensitivity analysis will determine how much confidence can be placed on the findings when drawing conclusions about the meta-analysis results.

A meta-analysis of a hypothetical DM organization

The concepts and methodologies presented thus far to evaluate the overall effectiveness of an intervention or treatment described in the literature can be equally as valuable when applied at the organizational level of any DM firm. In this section, a meta-analysis will be performed using data from a hypothetical DM organization that manages multiple disease conditions (diabetes, asthma, CHF and CAD), provides services to a variety of populations (Medicaid, Medicare and commercial health plans), and evaluates programme effectiveness using several different techniques (the pre-post total population approach, matched case-controls and the RCT). These data were fabricated in a manner to emphasize a variety of issues that are of real concern in DM.

Figure 4 presents the forest plot for the simulated data used in the meta-analysis. The X-axis indicates the cost-savings per-person-per-year (PPPY). On this scale, values to the right of 0 indicate cost-savings, and values to the left of 0 (negative values) indicate an increase in costs. The left Y-axis illustrates the subgroups by type and number of studies in each category. The right Y-axis provides weighted effect sizes and 95% CI for each category.

Twenty-seven individual DM programme evaluations were included in the meta-analysis with the overall estimated pooled effect showing a PPPY savings of $1000 (CI = $700, $1300). This can be restated as: 'We are 95% confident that the average savings per person across all programs is between $700 and $1300 annually.'

When pooling results by evaluation design, results for the RCT show a cost increase (indicated by a negative value), while both pre-post and case-control designs show cost-savings. The size of the CI surrounding each value provides additional information as to the implication of these findings. For example, in comparing outcomes among various research designs, we would expect the RCT to have the smallest variability between studies. Conversely, we would expect the pre-post design to exhibit the largest variability between studies. In DM, the extent of this variability is largely dictated by the amount of control programme evaluators have over
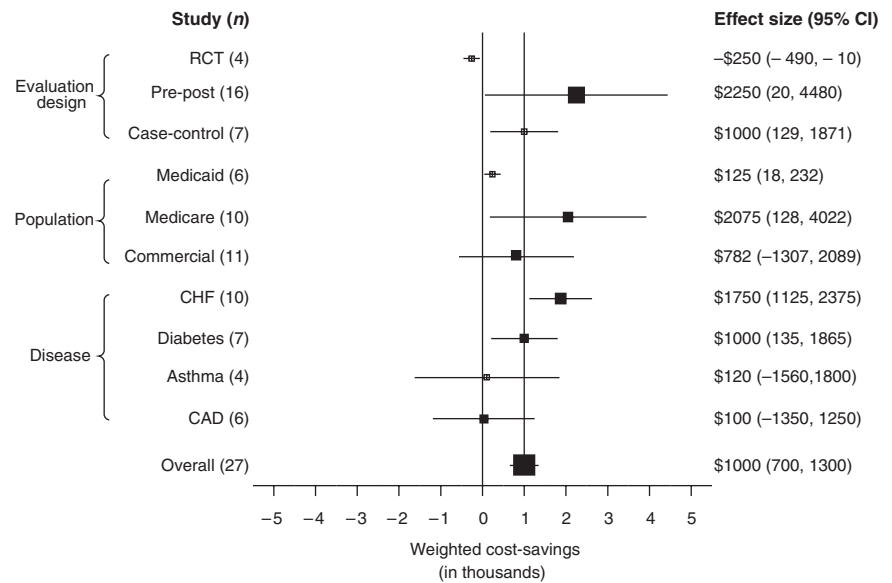
Figure 4 Forest plot of a meta-analysis of 27 hypothetical DM programme evaluations. Weighted effect sizes represent pooled average per-person-per-year savings. CAD, coronary artery disease; CHF, congestive heart failure; CI, confidence interval; DM, disease management; RCT, randomized controlled trial.

selection bias. In the RCT, assignment to treatment condition is completely controlled by the researcher. This virtually eliminates the introduction of selection bias and other related threats to validity (e.g. regression to the mean) into the study. As a consequence, treatment effects are typically small. Conversely, in the pre-post design where there is no control over programme enrolment, a myriad of biases are introduced into the study, typically leading to tremendous variability in treatment effects [26]. As case-control studies can only match on observed characteristics [30], we would expect the treatment effect and variability between studies to lie closer to the RCT than pre-post designs.

In evaluating the pooled effect by the type of populations served, Medicare showed the greatest degree of cost-savings. However, commercial programmes had the next best average pooled effect the CI extended beyond the $0, indicating that some programmes resulting in cost increases. Medicaid programmes appeared to show slight but positive programme savings. These results could be explained based on the following insights: (1) Medicare recipients are older and sicker and subject to more acute exacerbations than other populations. Therefore, DM programmes have more opportunity to prevent unnecessary hospitalizations and ED visits. (2) Medicaid recipients are typically less likely to remain in the programme long enough to get the full impact of the intervention because of the loss of their Medicaid benefits. Additionally, this population tends to access ED services as a primary source of care. As a result, we would not expect DM programmes to achieve dramatic cost-savings. (3) Programmes targeting commercial health-plan members are faced with a working-aged population whose illness trajectory is harder to predict and utilizes much fewer acute services. As a result, cost-savings are less assured in this population than others.

There appears to be quite a bit of heterogeneity in pooled effect sizes between diseases managed by this company. An understanding of the nature of each disease may help explain these results: half of the patients diagnosed with CHF will die within 5 years [45] and hospitalization rates are typically quite high compared with other conditions. Acute exacerbations requiring high-

intensity services can be managed quite effectively when CHF evidence-based practice guidelines [46] are followed, leading to cost-saving opportunities. Because of the advanced age of the typical CHF patient, this plot shows a possible relationship existing between CHF and being in the Medicare population. Diabetes is a chronic disease that may take years to develop to the point where utilization of acute services becomes imminent. Even when diabetic patients are managed in accordance to guidelines [47], it may take several periods for a DM programme to realize cost-savings; Asthma is a disease in which acute exacerbations are generally rare unless environmental triggers are present or recommended therapies [48] are not being followed. In this hypothetical DM organization, most asthma programme participants are Medicaid recipients, further confounding the likelihood that these programmes will show demonstrative cost-savings. CAD is perhaps the most problematic disease for evaluating DM programme effectiveness because patients are mostly identified only after events (e.g. acute myocardial infarction) and costly services have been rendered (e.g. ambulance, ED, intensive care, hospitalization, bypass surgery or angioplasty, cardiac rehabilitation, etc.). The foremost reason for this delay is a general inability to collect detailed information necessary to assess acute myocardial infarction risk before it occurs. As a result, these DM programmes can really only focus on reducing re-hospitalizations which have less cost-saving opportunities.

Figure 5 shows a funnel plot demonstrating an asymmetrical distribution of values based on the programme evaluation type, supporting the data presented earlier in Fig. 4. As expected, the pre-post studies have the largest sample sizes of all designs, followed by case-controls and RCT. In the pre-post methodology, the entire diseased population is included in the analysis, while in the case-control design only close matches are included. Because of the stringent and resource intensive nature of the RCT, sample sizes are typically low in comparison to other designs. This funnel plot visually displays the concern of selection bias associated with the various design types. As explained earlier, in the absence of bias larger sample sizes lead to a smaller amount of variability. In
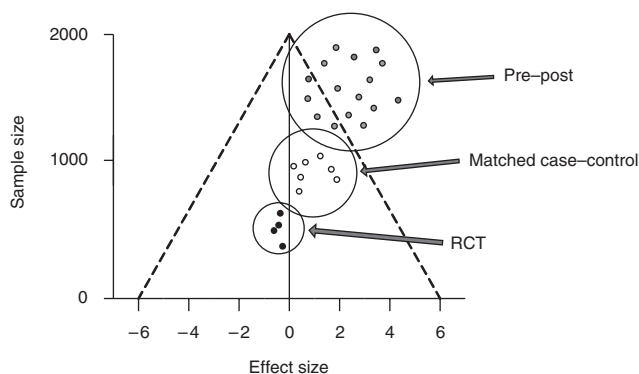
Figure 5 Funnel plot for assessing bias across individual programme evaluations. RCT, randomized controlled trial.

this example, we see the opposite occur. Taken together, these results indicate that study outcomes pooled for this meta-analysis are significantly influenced by bias.

In summary, this hypothetical DM organization shows an overall positive cost-savings of $1000 PPPY (CI = 700, 1300). The pre-post design appears to elicit the highest effect size when used to evaluate programme effectiveness, programmes directed at Medicare recipients tend to save the most money, and CHF appears to be the disease with the highest cost-saving potential. Bias is evident in comparing outcomes as a function of study design, so the overall weighted effect size must be viewed with caution. This company would be best served to begin evaluating programme effectiveness using designs with fewer threats to internal validity (case-controls, RCT, etc.). It should be noted once again that these data were fabricated in order to exemplify particular circumstances. In reality, most analyses are not as clear-cut as those presented here. However, this representation should provide readers and meta-analysts with the tools necessary to identify and explain their own results.

Conclusion

This paper introduced readers to the systematic process for conducting and interpreting results of a meta-analysis. Additionally using data from a hypothetical DM organization, a meta-analysis was performed to assess the overall weighted cost-saving achievement by the firm, as well as determining the sensitivity of results to differing evaluation designs, populations served and diseases treated.

The importance of conducting sound research in DM outcomes and publishing those results cannot be overstated. With high expectations placed on the DM industry to help manage runaway medical costs, any conclusion drawn from the research will be immediately pounced upon by either advocates or critics of DM. Given that few well-conducted and evaluated DM outcomes studies are currently found in the peer-reviewed literature, performing a meta-analysis on those manuscripts that meet the stringent criteria of a systematic review makes good sense. However, care must be taken to ensure that a sufficient number of studies are available and that the appropriate steps are taken to ensure that biases are identified and controlled for. Similar to the studies that a meta-

analysis evaluates, the process must be transparent for others to review and emulate. One of the benefits of meta-analysis is that the results can be readily updated as more individual studies become available. At the individual organization level, the meta-analytic technique is an excellent way of establishing the overall programme effect, as well as determining which elements are more sensitive than others in achieving these outcomes.

References

1. Congressional Budget Office. (2004) An Analysis of the Literature on Disease Management Programs. Washington, DC: Congressional Budget Office. Available at: http://www.cbo.gov/showdoc.cfm?index=5909&sequence=0 (accessed 13 October 2004).
2. Nussbaum S, Selecky C. (2004) Leading Voices in Disease Management Discuss CBO Report, Steps to Demystify Disease Management for Health Plans, Employers, and Government. Washington, D.C., USA: Disease Management Association of America (DMAA). Available at: http://www.dmaa.org/education_cboreport.asp (accessed 25 April 2007).
3. Diamond F. (2004) DM's cost-effectiveness doubted in CBO report. *Managed Care*, 13 (11), 31–32, 35.
4. DM advocates to CBO: the outcomes are there: industry leaders find fault with new report on savings from DM. *Disease Management Advisor*, 3 (11), (2004) 124.
5. Bayer, E., Chovan, T. & Lemieuz, J. (2005) CBO's Analysis of Potential Savings from Disease Management Programs in Medicare. Washington, D.C., USA: Center for Health Research and Policy at America's Health Insurance Plans (AHIP). Available at: http://www.ahip.org/content/default.aspx?bc=38 |72|4361 (accessed 26 February 2005).
6. Landro, L. (2004) Does disease management pay off? *Wall Street Journal*, 20, D4.
7. Foote, S. M. (2003) Population-based disease management under fee-for-service Medicare. *Health Affairs*, Web Exclusive (30 July 2003).
8. Nashville Business Journal (2005) American Healthways profits rise 59% in Q2. *Nashville Business Journal*. Available at: http://www.bizjournals.com/nashville/stories/2005/03/14/daily35.html (accessed 17 March 2005).
9. Wilson, T. & Linden, A. (2004) Measuring diabetes management. *Health Affairs*, 23 (6), 277. Re: Villagra, A. & Ahmed, T. (2004) Effectiveness of a disease management program for patients with diabetes. *Health Affairs*, 23 (4), 255–266.
10. Linden, A., Wilson, T. W. (2005) Care management for heart failure. *Annals of Internal Medicine* 142 (5), 386. Re: DeBusk, R. F., Miller, N. H., Parker, K. M., Bandura, A., Kraemer, H. C., Cher, D. J., *et al.* (2004) Care management for low-risk patients with heart failure: a randomized, controlled trial. *Annals of Internal Medicine*, 141, 606–613.
11. Wilson, T. W. & Linden, A. (2005) 'Asthma disease management effectiveness: were the controls equivalent to the cases?' *American Journal of Managed Care*, (3), 136. Re: Tinkelman, D. & Wilson, S. (2004) Asthma disease management: regression to the mean or better? *American Journal of Managed Care*, 10 (12), 948–954.
12. Linden, A. & Wilson, T. W. (2005) Letter to the editor. *Circulation* 112 (1), e11. Re: Galbreath, A. (2004) Long term healthcare and cost outcomes of disease management in a large, randomized community based population with heart failure. *Circulation*, 110 (23), 3518–3526.
13. Wilson, T. W. & Linden, A. (2005) Disease management economic value: unproven? *Health Affairs*, 24 (2), 566–567. Re: Fireman, B., Bartlett, J. & Selby, J. (2004) Can disease management reduce health care costs by improving quality? *Health Affairs*, 23 (6), 63–75.
14. Ofman, J. J., Badamgarav, E., Henning, J. M., Knight, K., Gano, A. D. Jr, Levan, R. K., Gur-Arie, S., Richards, M. S., Hasselblad, V. &

Weingarten, S. R. (2004) Does disease management improve clinical and economic outcomes in patients with chronic diseases? A systematic review. *American Journal of Medicine*, 117 (3), 182–192.

15. Goetzel, R. Z., Ozminkowski, R. J., Villagra, V. G. & Duffy, J. (2005) Return on investment in disease management: a review. *Health Care Financing Review*, 26, 1–19.

16. Krause DS. (2005) Economic effectiveness of disease management programs: a meta-analysis. *Disease Management*, 8, 114–134.

17. Pearson, K. (1904) Report on certain enteric fever inoculation statistics. *British Medical Journal*, 3, 1243–1246.

18. Fisher, R. (1932) A Statistical Methods for Research Workers, 4th edn. Edinburgh: Oliver and Boyd.

19. Yates, F. & Cochran, W. G. (1938) The analysis of groups of experiments. *Journal of Agricultural Science*, 28, 556–580.

20. Beecher, H. K. (1955) The powerful placebo. *Journal of the American Medical Association*, 159, 1602–1606.

21. Kirsch, I. & Sapirstein, G. (1998) Listening to Prozac but hearing placebo: a meta-analysis of antidepressant medication. *Prevention & Treatment* 1. Article 0002a. Available at: http://journals.apa.org/prevention/volume1/pre0010002a.html (accessed 26 February 2005).

22. Glass, G. V. (1976) Primary, secondary, and meta-analysis. *Educational Researcher*, 5, 3–8.

23. Cochrane, A. L. (1979) 1931–71: a critical review, with particular reference to the medical profession. In Medicines for the Year 2000, pp. 1–11. London: Office of Health Economics.

24. Chalmers, I., Enkin, M. & Keirse, M. (eds) (1989) Effective Care in Pregnancy and Childbirth. Oxford: Oxford University Press.

25. The Cochrane Library (2005) Issue 3. Available at: http://www.3.interscience.wiley.com/cgi-bin/mrwhome/106568753/HOME (accessed 3 March 2005).

26. Alderson, P., Green, S. & Higgins, J. P. T. (eds) (2004) Cochrane Reviewers' Handbook 4.2.2 [updated March 2004]. In The Cochrane Library, Issue 1, 2004. Chichester: John Wiley & Sons.

27. Linden A, Roberts N. (2005) A users guide to the disease management literature: recommendations for reporting and assessing program outcomes. *American Journal of Managed Care*, 11 (2), 81–90.

28. Linden A, Adams J, Roberts N. (2003) An assessment of the total population approach for evaluating disease management program effectiveness. *Disease Management*, 6 (2), 93–102.

29. Linden A, Adams J, Roberts N. (2004) The generalizability of disease management program results: getting from here to there. *Managed Care Interface*, 17 (7), 38–45.

30. Linden A, Adams J, Roberts N. (2006) Evaluating disease management program effectiveness: an introduction to the regression-discontinuity design. *Journal of Evaluation in Clinical Practice*, 12 (2), 124–131.

31. Linden A, Adams J, Roberts N. (2003) Evaluating disease management program effectiveness: an introduction to time series analysis. *Disease Management*, 6 (4), 243–255.

32. Linden A, Adams J. (2006) Evaluating disease management program effectiveness: an introduction to instrumental variables. *Journal of Evaluation in Clinical Practice*, 12 (2), 148–154.

33. Linden A, Adams J, Roberts N. (2005) Using propensity scores to construct comparable control groups for disease management program evaluation. *Disease Manage Health Outcomes*, 13 (2), 107–127.

34. Linden, A., Adams, J. & Roberts, N. (2006) Strengthening the case for disease management effectiveness: unhiding the hidden bias. *Journal of Evaluation in Clinical Practice*, 12 (2), 140–147.

35. Yusuf, S., Peto, R., Lewis, J., Collins, R. & Sleight, P. (1985) Beta blockade during and after myocardial infarction: an overview of the randomized trials. *Progress in Cardiovascular Diseases*, 27 (5), 335–371.

36. DerSimonian, R. & Laird, N. (1986) Meta-analysis in clinical trials. *Controlled Clinical Trials*, 7, 177–188.

37. Greenland, S. (1987) Quantitative methods in the review of epidemiologic literature. *Epidemiological Review*, 9, 1–30.

38. Berlin, J. A., Laird, N. M., Sacks, H. S. & Chalmers, T. C. (1989) A comparison of statistical methods for combining event rates from clinical trials. *Statistics in Medicine* 2, 141–151.

39. Gavaghan, D. J., Moore, R. A. & McQuay, H. J. (2000) An evaluation of homogeneity tests in meta-analyses in pain using simulations of individual patient data. *Pain*, 85, 415–424.

40. StatsDirect LTD. (2005) *StatsDirect statistical software.* England. Available at: http://www.statsdirect.com (accessed 17 August 2005).

41. Lewis, S. & Clarke, M. (2001) Forest plots: trying to see the wood and the trees. *British Medical Journal* 322, 1479–1480.

42. Egger, M., Smith, G. D., Schneider, M. & Minder, C. (1997) Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, 315, 629–634.

43. Light, R. J. & Pillemer, D. B. (1984) Summing up. The Science of Reviewing Research. Cambridge, MA: Harvard University Press.

44. Egger, M., Schneider, M. & Smith, G. D. (1998) Meta-analysis spurious precision? Meta-analsysis of observational studies. *British Medical Journal*, 316, 140–144.

45. National Heart, Lung, and Blood Institute (NHLBI). (2004) Morbidity & Mortality: 2004 Chart Book on Cardiovascular, Lung, and Blood Diseases. Bethesda, MD: US Department of Health and Human Services, Public Health Service, National Institutes of Health.

46. Hunt, S. A., Baker, D. W., Chin, M. H. *et al.* (2001) ACC/AHA guidelines for the evaluation and management of chronic heart failure in the adult: executive summary. A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee to revise the 1995 Guidelines for the Evaluation and Management of Heart Failure). *Journal of the American College of Cardiology*, 38, 2102–2112.

47. American Diabetes Association. (2005,) Clinical practice recommendations 2005. *Diabetes Care*, 28 (Suppl. 1), S1–79.

48. National Education Program. (2002) Expert Panel Report II. Guidelines for the Diagnosis and Management of Asthma. Bethesda, MD: US Department of Health and Human Services.