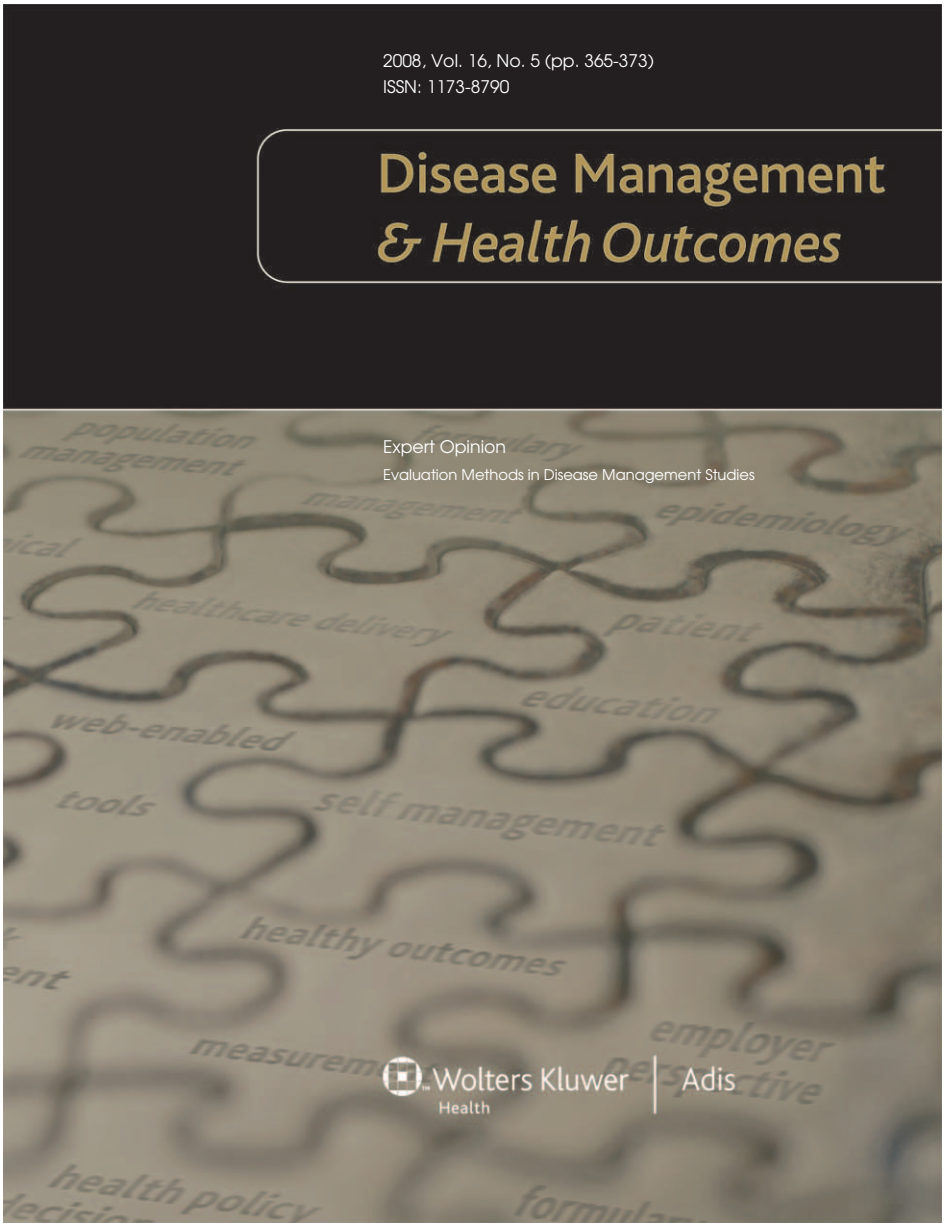

EXPERT OPINION

Dis Manage Health Outcomes 2008; 16 (5): 365-373
1173-8790/08/0005-0365/$48.00/0

Evaluation Methods in Disease Management Studies

2004–07

Thomas Wilson,[1,2] *Martin MacDowell*,[1,3] *Patricia Salber*,[1,4] *Gary Montrose*[1,5] and *Carolyn Hamm*[1,6]

1 Population Health Impact Institute, Loveland, Ohio, USA
2 Trajectory Healthcare, LLC, Loveland, Ohio, USA
3 UIC College of Medicine at Rockford, Rockford, Illinois, USA
4 Universal American Corp., Houston, Texas, USA
5 Montrose Healthcare Strategies, LLC, Denver, Colorado, USA
6 Walter Reed Army Medical Center, Washington, DC, USA

Abstract

A variety of program evaluation designs are available to assess the impact of disease or care management programs, which can make it difficult to compare outcomes of different interventions. The need to compare programs has resulted in consideration of standardizing evaluations of disease management programs; however, recommendations on the conduct of such evaluations have not been widely adopted. The purpose of this article is to examine the consistency of study characteristics of disease management peer reviewed evaluations over a 3-year period (1 January 2004–28 February 2007) and to suggest questions that must be answered to ensure basic transparency of methods and metrics.

Study designs vary considerably among the current literature on evaluations of disease management interventions involving US health plans (25 studies). The mechanism for defining the intervention populations were not consistent, even among interventions focused on a single disease, and evaluations employed both administrative and clinical data. The current literature included both randomized (n = 10) and non-randomized studies (n = 15). The referent population varied among the non-randomized studies, and included data from the pre-intervention period and both concurrent and historical control groups. The outcome metrics used in the evaluations included mortality and readmission rates, as well as time to readmission and various cost parameters. The majority of reviewed studies corrected for the confounding variables of age and sex, and a high proportion corrected for a range of other confounding factors.

In conclusion, the evaluations of disease management programs in the literature cannot be considered standardized. To increase the transparency and validity of disease management intervention evaluations, we recommend consideration of five basic questions regarding intervention descriptions, intervention population, referent population, outcomes metrics, and confounding variables. Standardization on such basic parameters is a necessary step towards being able to assess the quality and validity of evaluations. Such standardization is essential for comparing the effectiveness of alternative programs, and to enable data-driven value-based purchasing decisions.

A 2003 article commissioned by the Disease Management Association of America (DMAA) listed seven different study design types that could be used to examine the impact of disease management: follow-up, case-control, cross-sectional, ecological, quasi-experimental, benchmark, and post-only.[1] In 2005, the

DMAA *Outcomes Guide*[2] listed 12 different 'study design characteristics':

1. randomized controlled trial (RCT)
2. pre-post with the patient as their own control
3. pre-post historical control design with concurrent control group
4. pre-post historical control with actuarial adjustment
5. pre-post historical control with no actuarial adjustment
6. pre-post multiple baseline (with randomized controls)
7. benchmark study
8. pre-post time series
9. post-only control group
10. post-only/no control group
11. cross-sectional (no comparison)
12. simple time-series (e.g. run charts).

In a 2007 book published by the DMAA,[3] multiple methods for evaluating disease management programs were described, based upon the disciplines of health services research, epidemiology, econometrics, and actuarial sciences.

These articles highlighted the multiplicity of design options available to assess the impact of disease management. Many industry experts are concerned with this lack of standardization and are working to develop a standard method to enable disease management programs to be compared side by side.[4] The purpose of this article is to examine the consistency of disease management peer reviewed evaluations over a 3-year period. This is an essential initial step in considering the variations present in the published literature. The goal is to develop a framework for assessing disease management evaluations that is robust to the variety of disease management interventions and evaluation methods. Valid evaluation findings should be used when comparing program effectiveness and outcomes.

This push toward methods standardization was started in earnest in 2003 when the disease management company American Healthways (now known as Healthways) and Johns Hopkins University suggested that a "pre-post study was sufficient at this time" (the inclusion of an actuarial referent was recommended in the appendix).[5] This model was never widely adopted[6] and other approaches have been recommended. The most recent attempt was by the DMAA itself, which, in December 2006, recommended "use of pre-post study design that incorporates, whenever possible, an equivalent, concurrent comparison group" and suggested several more detailed recommendations including "risk adjustment" and the use of "non-chronic trend" as a reference.[7] This book was updated with more detail and clarification in 2007[8] and 2008.[4]

In 2006, a study was commissioned by CorSolutions, a disease management company, and performed by the RAND Corporation. It recommended "in the absence of a control group … the use of benchmark data as rough estimate of secular trend and statistical estimates to the degree possible."[9] Also in 2006, Sexner et al.[10] recommended a "retrospective cohort pre-post baseline population based savings analysis" where results are "compared with the trend of the non-diseased population … the most appropriate available comparison group." In 2006, Linden[11] recommended that the condition-specific admission rate over an extended baseline be used as a reference in the evaluation of disease management interventions measuring condition-specific admissions. In December 2007, the American Academy of Actuaries recommended that "causality" was not the goal, but "correlation" or association was "acceptable to most buyers," and recommended an actuarial method, with a strong stress on equivalence.[12] These reviews did not examine the level of standardization of evaluation today. Other reviews of the disease management literature have been conducted, but these have primarily focused on the value of disease management, not the methods used to determine that value.[13-20]

There is a need to review the current level of standardization among disease management impact studies. The purpose of this paper is to assess the processes and methods used to assess outcomes in peer-reviewed, US-based disease management evaluation studies performed with the involvement of a health plan and published between 2004 and 2007. The specific outcomes/effects of the disease management programs will not be examined.

To this end, we performed a search in MEDLINE using the key search criteria 'insurance claims' or 'administrative claims' or 'chronic illness care' or 'chronic care model' or 'population-based management' or 'disease-management' or 'disease management' or 'care management' and 'return on investment' or 'ROI' or 'economic' or 'financial' or 'savings' amongst articles published between 1 January 2004 and 31 December 2006. Abstracts were reviewed by three panelists (P. Salber, M. MacDowell, and T. Wilson) and articles were excluded if they did not involve a health plan, focused only on prescription-based disease management, featured interventions that were managed by employers, or described studies that were not performed in the US. This resulted in the selection of 23 articles.[21-43] Two additional articles that met the criteria above and were published between 1 January 2007 and 28 February 2007, were added to the review at a later date.[44,45] Each article was reviewed by M. MacDowell, C. Hamm, and T. Wilson.

The goal was to describe the level of standardization and variability of the study characteristics listed below. The consensus opinion regarding the evaluation design characteristics was reached by three reviewers (M. MacDowell, C. Hamm, and T. Wilson).

1. Evaluation Design Characteristics

Descriptions of the following characteristics must be present in order to achieve transparency of methods and metrics.

1.1 Intervention

This is what is done to the population. This is not an element of the evaluation design; it is rather the thing the study is evaluating.

1.2 Intervention Population

This is the group of people in which the disease management program is implemented.

1.3 Outcomes Metrics

These are the targeted measures that the intervention is designed to impact. Although the term includes the word 'outcomes,' these measures may or may not be caused by the intervention. Indeed, a basic requirement of an evaluation is the use of these in the baseline period, to compare with the post period in the intervention population only, to the pre-period in the external referent (if available).

1.4 Referent

This is what the intervention population is compared with and, ideally, is designed to estimate values for the outcome metrics in the intervention population in the absence of the intervention. Broadly defined, referents include control groups, expert opinion, the defined time period for the 'pre-intervention' period in a pre/post-intervention study (involving either the same individuals or a different group of people), historical control groups from a time period prior to that of the intervention population, and the concurrent period to the intervention population.

The referent can be chosen by an experimental method (e.g. randomization) or a non-experimental method (e.g. people who opt out, staggered roll-out, an employer group that did not participate in disease management, a matched control group, a previous period in time for a patient population). We have used the term 'referent,' rather than the more common 'reference population' as there are some referents in place that are not populations, but rather benchmarks (as recommended by RAND), articles, or non-standard comparisons (e.g. the pre- intervention period in a pre-post study) or articles, or non-standard comparisons (e.g. the pre-period in a pre-post evaluation design).

1.5 Confounding Factors

These are factors, which may or may not be measured, that could be independently related to the outcome metric but are not part of the intervention. Common confounding variables are age and sex. These factors can confound a causal relationship. Here is a typical example: if the age distribution in the intervention population was 40% elderly, while in the concurrent referent the age distribution was 60% elderly, the better outcomes in the intervention population could be totally or partially a result of the younger population in the intervention group, rather than the intervention itself. Other potential confounding factors include differences between the intervention population and the referent in benefit design, disease severity, the presence of co-morbid conditions, the propensity of participants to take care of their health, and provider mix. Importantly, it is unlikely that all confounding factors can be taken into account in any study. Thus, positive (or negative) results from disease management studies should never be thought of as 'proof.' Confounders are variables that are only useful when there is a referent other than the pre-intervention period in a simple pre-post study in a single cohort of patients.

Each evaluation design characteristic was examined by three reviewers (M. MacDowell, C. Hamm, and T. Wilson).

2. Characteristics of the Reviewed Disease Management Evaluations

2.1 Interventions

The majority of the 25 articles that met our selection criteria for disease management interventions were 'telephonic nurse management,' 'monthly group meetings,' 'senior life management,' 'care advocacy,' 'immunization reminders,' 'home-based teaching,' 'interventions to enhance fitness,' 'interventions to enhance care management,' and 'collaborative care.' See table I for a complete description of interventions.

2.2 Intervention Population Definition:

There was a large number of different disease/condition states represented in the 25 articles. These included congestive heart failure (CHF; n = 7), diabetes mellitus (n = 2), asthma (n = 3),

Table I. Description of cited articles reviewed (as described by the authors of the articles)

Study	Design/nature of RCT	Main attributes of study population	Intervention	Outcome metrics	Confounding metrics
Martin et al.[21]	RCT (individual)	Medicare, CHF, etc.	Senior life/care management	SF-6, resource use measured by admission rates and bed-days per thousand per year, member satisfaction, and costs measured by paid claims	Age, sex (all others were outcomes, e.g. SF)
Tinkelman and Wilson[22]	Non-RCT	Asthma	Multiple intervention DM	Total costs	Age, sex
Galbreath et al.[23]	RCT (individual)	CHF	Multiple intervention DM	All-cause mortality, EF, % on guideline-based Rx, costs, stays	Age, sex, race, marital status, functional status, BP, medical history, pulse, pharmacotherapy
DeBusk et al.[24]	RCT (offices)	Low-risk CHF	Nurse case management, telephonic	Time to rehospitalization, reasons for re-hospitalization	Age, sex, race, marital status, education level, HF history (previous diagnosis of HF, baseline NYHA class I–II, III–IV), cause of HF (coronary disease, hypertension, valvular disease, unknown), symptoms before hospitalization (angina, atypical chest pain, exertional shortness of breath, cough, paroxysmal nocturnal dyspnea, peripheral edema, orthopnea)
Sackett et al.[25]	Non-RCT	Prenatal	Prenatal program with case management	LBW, costs	None
Villagra and Ahmed[26]	Non-RCT	Diabetes	Multiple intervention DM	Cost (inpatient, outpatient, professional services, Rx drugs, other), use (days, ER, visits, admissions), quallity indicators (HbA_{1c}, ACE, dilated retinal exam, microalbumin, lipid, tobacco use)	Age, sex
Berg et al.[27]	RCT (individual)	All members	Direct mail marketing of immunization	Influenza/pneumonia admission/ER, influenza vaccinations pneumonia vaccinations, estimated cost/savings (multiplying the estimated absolute difference in utilization and utilization or vaccination)	Age, sex, household size, member months, dual coverage, condition prevalence (eight common chronics)
Berg et al.[28]	Non-RCT	CHF	Telephonic DM	Annual cost, $US (medical, Rx). Medical service utilization (annualized rate per 1000), prescription drug use (annual days supply per person), prescription drug use baseline (% who had prescription)	Demographic (months pre, mean) and co-morbidity (% CAD, COPD, hypertension, diabetes mellitus, arthritis, dementia, depression)
Scott et al.[29]	RCT (individual)	≥11 outpatient visits	Monthly group meetings patients and doctors	Per patient clinic visits, pharmacy fills, hospital admissions, hospital observation admissions, hospital outpatient visits, professional services, emergency visits, skilled nursing facility admissions, home health visits; cost (by service line), QOL (multiple)	Age, sex, marital status, % with a chronic condition (14 listed), no. of Rx, mobility limitation, depression, % of memory problems, live alone, no basic ADL deficits, no advanced ADL deficits, no household ADL deficits, fair or poor health status

Continued next page

Table I. Contd

Study	Design/nature of RCT	Main attributes of study population	Intervention	Outcome metrics	Confounding metrics
Rost et al.[30]	RCT (groups)	Depression	Enhanced care management	Days free of depression, program and outpatient costs, patient time and transportation costs, QALYs	None listed
Catov et al.[31]	Non-RCT	Asthma	Home-based teachings	Hospital admits, ER admits	Sex, race, age, region of country, urban/rural
Hudson et al.[32]	Non-RCT	CHF	Remote monitoring	ER visits, inpatient admissions, re-admissions	None (stratified by sex and age)
van Vonno et al.[33]	Non-RCT	CHF	Multiple intervention DM	CHF-related hospital stay or outpatient visit, total cost differences pre-post (adjusted for length of program participations)	Age, sex, conditions (rheumatic, hypertensive, ischemic, pulmonary heart diseases), drugs (dioxin, diuretics, etc), used nursing home, used home care, no. of diagnosis codes
Delaronde et al.[34]	RCT	Asthma	Multiple intervention DM	Changes in asthma medication use, physician office visits, ER visits, hospitalizations, and QOL	Age, sex
Dunagan et al.[35]	RCT (individuals)	CHF	Nurse, telephonic	Mortality, re-admission or ED visit, any re-admission, primary HF readmission, any readmision or death	Age, sex, race, marital status, help living at home, health with daily living, ischemic etiology of HF, LVEF, NYHA classification, ACE regimen, discharge creatinine, Charlson score, SF-12, Back Inventory Scale, MLHFQ (physical and mental)
Hawkins et al.[36]	Non-RCT	Special needs children	Risk assessment and case management	Member satisfaction, admits, LOS, ED total costs	None
Schwerner et al.[37]	Non-RCT	14 chronic, complex conditions	Multiple intervention DM	PMPM, trend	14 chronic, complex conditions
Wise et al.[38]	Non-RCT	Population	Medical and DM	PMPM (all, IP, OP, Rx, Prof), quality (HEDIS: HbA_{1c} testing and control, eye exam, lipid exam, LDL <130, nephropathy monitoring)	Age, sex, health plan benefits (estimated), number of health conditions, prevalence of health conditions
Daly et al.[39]	RCT (individual)	Chronically critical ill >3 days mechanical ventilation, post-discharge	Case management	Hospitalization re-admission, re-admitted within 48 hours, death at re-admission, days to first re-admission, total re-hospitalization days; discharge disposition (long-term acute care, rehabilitation, nursing home, death), discharge status (with oxygen, tracheotomy, ventilator, cognitive impairment); SF-8 score physical functioning and mental functioning, patient ADL/ADL score	Age, APACHE score, Glascow, pre-hospital admission co-morbid score, medication pre-hospital score, length of hospital stay, sex, race, primary ICD-9 diagnosis

Continued next page

Table I. Contd

Study	Design/nature of RCT	Main attributes of study population	Intervention	Outcome metrics	Confounding metrics
Sidorov[40]	Non-RCT	CHF	Multiple intervention DM	Assessment of LVH, prescribed an ACE or contraindication (LVEF 140%), prescribed a β-adrenoceptor antagonist; inpatient days, ER visits, total claims, ER claims; financial (outpatient, inpatient, CHF-related claims)	None
Grypma et al.[41]	Non-RCT	Depression	Collaborative Care Management: IMPACT	Utilization of services, annualized costs, depression scores (baseline and 6-month and percentage experiencing 50% improvement in depression) antidepressant use, treatment contacts	Age, sex
Shannon et al.[42]	RCT (individual)	High use, >65 years	CA program: referrals to home- and community-based services	Utilization in services (primary care physician, specialist, hospital admissions, hospital days, ER)	Age, sex, measures of health status (not shown)
Sweeney et al.[44]	Non-RCT	Life-limiting diagnosis	Intensive patient-centered management	Home care days, hospice days, rehabilitation days, ER visits, hospital days, hospital admissions; inpatient admissions (nausea and vomiting, anemia, fluid disorder and dehydration, fever, nutritional deficiency, metabolism, anxiety), mortality/survival	Age, sex, broad ICD-9 categories (top 6 of 16: neoplasms, circulatory, digestive system, nervous and sense, ill-defined, injury, and poisoning)
Nguyen and Ackerman[45]	Non-RCT	Diabetics	Enhanced fitness program	Hospital admissions (all, diabetes related), annual primary care visits (all, diabetes related), healthcare costs (total, primary care, inpatient costs) (note: stratifying variables: level of participation in exercise program). Note: average inpatient costs for members who had inpatient costs	Age, sex, HbA_{1c} levels, chronic disease burden, prevention score (prevention score is the sum of the number of times a subject received colon cancer screening, [fecal occult blood test or flexible sigmoidoscopy], a screening mammogram, prostate cancer screening, an influenza vaccine, or a pneumococcal vaccine during the 2 years immediately preceding the index date)
Firestone et al.[43]	Non-RCT	CAD, CHF, asthma, diabetes	Multiple intervention DM	Quality measures (LDL test, statin Rx, antihypertensive Rx) clinic visits (physician, ER, admissions, days, Rx costs)	Disease prevalence

ADL = activities of daily living; **APACHE** = Acute Physiology and Chronic Health Evaluation; **BP** = blood pressure; **CA** = care advocate; **CAD** = coronary heart disease; **CHF** = congestive heart failure; **COPD** = chronic obstructive pulmonary disease; **DM** = disease management; **ED** = emergency department; **EF** = ejection fraction; **ER** = emergency room; **HbA_{1c}** = one type of glycosylated hemoglobin; **HEDIS** = Health Plan Employer Data and Information Set; **HF** = heart failure; **ICD-9** = International Classification of Diseases, version 9; **IMPACT** = Improving Mood-Promoting Access to Collaborative Treatment; **IP** = inpatient; **LBW** = low birth weight; **LDL** = low-density lipoprotein; **LOS** = length of stay; **LVEF** = left ventricular ejection fraction; **LVH** = left ventricular hypertrophy; **MLHFQ** = Minnesota Living with Heart Failure questionnaire; **NYHA** = New York Heart Association; **OP** = outpatient; **PMPM** = per member per month; **Prof** = professional; **QALY** = quality-adjusted life-year; **QOL** = quality of life; **RCT** = randomized controlled trial; **Rx** = prescription; **SF-6** = Outcomes Study 36-Item Short-Form Health Survey.

chronic diseases (coronary artery disease, CHF, diabetes, and asthma; n = 1), total health plan population (without regard to diagnosis; n = 3), depression (n = 2), high utilizers (n = 2), severe chronic disease (n = 2), life-limiting diagnoses (n = 1), discharged with ventilator (n = 1), prenatal (n = 1), and special needs children (n = 1).

The criteria used to define the intervention population were not consistent among evaluations of programs for a single disease. For example, among studies of CHF interventions, some defined the study population using the diagnostic codes (i.e. International Classification of Diseases, version 9 [ICD-9]) from administrative data, with or without health plan membership requirements or utilization history (regarding inpatient stays, emergency room visits, etc.). In addition, some used other factors (e.g. the ability to communicate with the patient's primary care provider) as a secondary criteria for selection. The randomized trials generally used clinical parameters (e.g. left ventricular ejection fraction) to define the intervention population. There was no commonly applied standard for selecting populations.

2.3 Referent Types

Ten of the 25 studies used randomization to select the referent population; randomization was mostly at the individual level (n = 7),[21,23,27,29,34,35,39,42] while two studies randomized physician offices, rather than individuals.[24,30] One study[34] employed randomization for one sub-study and analyzed a combined population that included both the intervention group and referents for the other sub-studies.

The other 15 studies employed non-randomized designs. Two[32,36] used the pre-intervention period as the primary referent for the intervention period, three[25,41,43] used an historical referent (two of these used more than one design), ten used a concurrent referent (three of these used more than one design).[22,26,28,31,33,37,38,40,44,45] None used expert opinion; however, one study did cite an expert to adjust for benefit design differences between referent and intervention populations.

2.4 Outcome Metrics

Outcome metrics included mortality, admission, re-admission, time to re-admission, and cost (total, admission, out-patient, prescription). These were often related to specific diagnoses. Other metrics included days free of depression, the percentage of low birth-weight babies among all births, quality-of-life survey outcomes, patient satisfaction survey outcomes, and Healthcare Effectiveness Data and Information Set (HEDIS) metrics. See table I for a complete description of outcome metrics employed in the studies.

2.5 Confounding Factors

Five of the articles listed no confounding variables. Four studies adjusted results for differences in the age and sex of comparator groups (all were non-randomized); the other 16 adjusted results for differences in other variables, as well as age and sex.

Among the studies with non-randomized designs, the confounding variables that were adjusted for included the number or type of co-morbidities, HEDIS metrics (that did not require chart review – as is done in the HEDIS hybrid method, which utilizes both administrative claims data and chart review to derive the numerator and denominator of metrics), and prevalence of specific national drug codes (NDC), number of prescriptions, coverage type, benefit design, and the length of time in the plan/program. Other confounding variables that were not assessed using administrative data included household size, marital status, educational level, the presence/absence of left ventricular hypertrophy, laboratory parameter values, left ventricular ejection fraction, and New York Heart Association class. See table I for a complete description of confounding factors in the studies reviewed.

3. Conclusions and Recommendations

Clearly, we do not see any trend towards the standardization of intervention populations, outcome metrics, or even a common definition of 'disease management.' within current evaluations of disease management programs. Thus, comparison of outcomes between different interventions would appear to be very difficult, if not impossible.

While we await a standard evaluation design, the best choice for value-based purchasing will be to use a basic list of criteria when assessing the methods used in a disease management evaluation. Such a list – if standardized – would at least enable the study methods to be compared as is done in 'level of evidence' scoring in the evidenced-based medicine area.[46-50] A standardized evaluation is essential for two reasons: (i) to assess level of confidence in the evaluation design (and to compare this across programs); and (ii) to assess level of confidence in the results.

Disease management evaluation studies – published or not – may include any number of technical and non-technical elements, but the foundational criteria needed to assess the validity of these studies is transparency of methods and metrics. Based upon the simple criteria we used to assess the methods used in a representa-

tive list of 25 articles, five simple questions that would provide increased transparency are listed below:

1. Is the intervention used in the disease management program clearly stated?
2. Is the population in which the intervention is implemented clearly defined?
3. Is the referent used in the disease management evaluation clearly defined?
4. Are the outcome metrics used to compare the intervention and referent group clearly defined and consistently assessed?
5. Are key confounding factors identified, measured, and properly taken into account when making a comparison between the outcome metrics obtained for the intervention and referent groups?

Other groups have suggested checklists for disease management, and we encourage the reader to review these more comprehensive lists. They range from the general to the specific: general checklists have been offered by Mathematica,[51] the National Managed Health Care Congress (now The Health/Care Evaluation workgroup),[52] the DMAA,[53] and the American Heart Association.[54,55] Specific checklists from health services research[56] and actuarial sciences[57] are also available. However, using these more comprehensive lists are not useful unless some detail is presented – made transparent – regarding the five issues we discuss.

An article is in preparation that will examine the quality and strength of these 25 studies using more sophisticated quantitative scoring criteria. This article will take the reader beyond the basics listed here to enable a more detailed look at the very important issues related to equivalence and statistical significance.

Acknowledgments

This article was funded in part by an unrestricted educational grant from the Health Industry Forum at Brandeis University provided to the Population Health Impact Institute.

References

1. MacDowell M, Wilson TW. Framework for assessing causality in disease management programs [commissioned paper by Disease Management Association of America]. Washington, DC: DMAA, 2003
2. Outcomes Consolidation Steering Committee. Disease management program guide: including principles for assessing disease management programs. Washington, DC: DMAA, 2005: 33
3. Lindner A, Wilson T, Berg G, et al. Perspectives in disease management evaluation: health services research, epidemiology, econometrics, and actuarial. In: Hay J, Rindress D, editors. Measuring disease management. Yardley (PA): DMAA, 2007
4. Disease Management Association of America. Outcomes guidelines report. Vol. 3. Washington, DC: DMAA, 2008
5. American Healthways and Johns Hopkins University. Standard outcomes metrics and evaluation methodology of disease management programs. 2nd Annual Disease Management Outcomes Summit; 2002 Nov 7–10; Palm Springs (CA). Nashville (TN): American Healthways Inc., 2003
6. Glabman M. 'Take my word for it': the enduring dispute over measuring DM's economic value. Managed Care April 2006 [online]. Available from URL: http://managedcaremag.com/archives/0604/0604.dmvalue.html [Accessed 2008 Nov 27]
7. Disease Management Association of America. Outcomes guidelines report. Vol I. Washington, DC: DMAA, 2006
8. Disease Management Association of America. Outcomes guidelines report. Vol II. Washington, DC: DMAA, 2007
9. Mattke S, Bergamo G, Balakrishnan A, et al. Measuring and reporting the performance of disease management programs. Santa Monica (CA): RAND, 2006 Aug
10. Sexner S, Baker K, Gold D. Guidelines for analysis of economic returns from health management programs: the art of health promotion. Art Health Prom 2006 Jul/Aug: 1-17
11. Linden A. What will it take for DM to demonstrate a return on investment? New perspectives on an old theme. Am J Manag Care 2006; 12: 217-22
12. American Academy of Actuaries' Disease Management Work Group. Disease management: a public policy practice note. Washington, DC: American Academy of Actuaries, 2007
13. Weingarten SR, Henning JM, Badamgarav E, et al. Interventions used in disease management programmes for patients with chronic illness: which ones work? Meta-analysis of published reports. BMJ 2002; 325: 1-8
14. Ofman JJ, Badamgarav E, Henning JM, et al. Does disease management improve clinical and economic outcomes in patients with chronic diseases? A systematic review. Am J Med 2004; 117: 182-92
15. Holtz-Eakin D. An analysis of the literature on disease management programs. Congressional Budget Office, 2004 Oct [online]. Available from URL: http://www.cbo.gov/ftpdocs/59xx/doc5909/10-13-DiseaseMngmnt.pdf [Accessed 2008 Oct 28]
16. Krause D. Economic effectiveness of disease management: a metaanalysis. Dis Manage 2005; 8: 114-34
17. Goetzel RZ, Ozminkowski RJ, Villagra VG, et al. Return on investment in disease management: a review. Health Care Financ Rev 2005; 26: 1-19
18. Wilson TW, MacDowell M, Montrose G. The prerequisite importance of judging the potential usefulness of workplace health promotion studies: beyond "Published in a Peer-Reviewed Journal" and "Strength of Study Design". J Health Productivity 2006; 1: 23-8
19. Pelletier KR. A review and analysis of the clinical and cost-effectiveness studies of comprehensive health promotion and disease management programs at the worksite: update VI 2000–2004. J Occup Environ Med 2005 Oct; 47 (10): 1051-8
20. Mattke S, Seid M, Sai M. Evidence for the effect of disease management: is $1 billion a year a good investment? Am J Manag Care 2007; 13: 670-6
21. Martin DC, Berger ML, Anstatt DT, et al. A randomized controlled open trial of population-based disease and case management in a Medicare Plus Choice health maintenance organization. Prev Chronic Dis 2004; 1 (4): A05 [online]. Available from URL: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1277945 [Accessed 2008 Nov 24]
22. Tinkelman D, Wilson S. Asthma disease management: regression to the mean or better? Am J Manag Care 2004 Dec; 10 (12): 948-54
23. Galbreath AD, Krasuski RA, Smith B, et al. Long-term healthcare and cost outcomes of disease management in a large, randomized, community-based population with heart failure. Circulation 2004; 110 (23): 3518-26
24. DeBusk RF, Miller NH, Parker KM, et al. Care management for low-risk patients with heart failure: a randomized, controlled trial. Ann Intern Med 2004; 141 (8): 606-13
25. Sackett K, Pope RK, Erdley WS. Demonstrating a positive return on investment for a prenatal program at a managed care organization: an economic analysis. J Perinat Neonatal Nurs 2004 Apr-Jun; 18 (2): 117-27
26. Villagra VG, Ahmed T. Effectiveness of a disease management program for patients with diabetes. Health Aff 2004; 23 (4): 255-66

27. Berg GD, Thomas E, Silverstein S, et al. Reducing medical service utilization by encouraging vaccines: randomized controlled trial. Am J Prev Med 2004 Nov; 27 (4): 284-8

28. Berg GD, Wadhwa S, Johnson AE. A matched-cohort study of health services utilization and financial outcomes for a heart failure disease-management program in elderly patients. J Am Geriatr Soc 2004 Oct; 52 (10): 1655-61

29. Scott JC, Conner DA, Venohr I, et al. Effectiveness of a group outpatient visit model for chronically ill older health maintenance organization members: a 2-year randomized trial of the cooperative health care clinic. J Am Geriatr Soc 2004 Sep; 52 (9): 1463-70

30. Rost K, Pyne JM, Dickinson LM, et al. Cost-effectiveness of enhancing primary care depression management on an ongoing basis. Ann Fam Med 2005 Jan-Feb; 3 (1): 7-14

31. Catov JM, Marsh GM, Youk AO, et al. Asthma home teaching: two evaluation approaches. Dis Manag 2005; 8 (3): 178-87

32. Hudson LR, Hamar GB, Orr P, et al. Remote physiological monitoring: clinical, financial, and behavioral outcomes in a heart failure population. Dis Manag 2005; 8 (6): 372-81

33. vanVonno CJ, Ozminkowski RJ, Smith MW, et al. What can a pilot congestive heart failure disease management program tell us about likely return on investment? A case study from a program offered to federal employees. Dis Manag 2005 Dec; 8 (6): 346-60

34. Delaronde S, Peruccio DL, Bauer BJ. Improving asthma treatment in a managed care population. Am J Manag Care 2005 Jun; 11 (6): 361-8

35. Dunagan WC, Littenberg B, Ewald GA, et al. Randomized trial of a nurse-administered, telephone-based disease management program for patients with heart failure. J Card Fail 2005 Jun; 11 (5): 358-65

36. Hawkins MR, Diehl-Svrjcek B, Dunbar LJ. Caring for children with special healthcare needs in the managed care environment. Lippincotts Case Manag 2006 Jul-Aug; 11 (4): 216-23

37. Schwerner H, Mellody T, Goldstein AB, et al. Evaluating the impact of a disease management program for chronic complex conditions at two large northeast health plans using a control group methodology. Dis Manag 2006 Feb; 9 (1): 34-44

38. Wise CG, Bahl V, Mitchell R, et al. Population-based medical and disease management: an evaluation of cost and quality. Dis Manag 2006 Feb; 9 (1): 45-55

39. Daly BJ, Douglas SL, Kelley CG, et al. Trial of a disease management program to reduce hospital readmissions of the chronically critically ill. Chest 2005 Aug; 128 (2): 507-17

40. Sidorov J. Reduced health care costs associated with disease management for chronic heart failure: a study using three methods to examine the financial impact of a heart failure disease management program among medicare advantage enrollees. J Card Fail 2006 Oct; 12 (8): 594-600

41. Grypma L, Haverkamp R, Little S, et al. Taking an evidence-based model of depression care from research to practice: making lemonade out of depression. Gen Hosp Psychiatry 2006 Mar-Apr; 28 (2): 101-7

42. Shannon GR, Wilber KH, Allen D. Reductions in costly healthcare service utilization: findings from the Care Advocate Program. J Am Geriatr Soc 2006 Jul; 54 (7): 1102-7

43. Firestone B, Bartlett J, Selby J. Can disease management reduce health care costs by improving quality? Health Aff 2004; 23 (6): 63-75

44. Sweeney L, Halpert A, Waranoff J. Patient-centered management of complex patients can reduce costs without shortening life. Am J Manag Care 2007; 13: 84-92

45. Nguyen HQ, Ackerman R. Impact of a managed-Medicare physical activity benefit on health care utilization and costs in older adults with diabetes. Diabetes Care Jan 2007; 30 (1): 43-8

46. Centre for Evidence-Based Medicine. Levels of evidence [online]. Available from URL: http://www.cebm.net/index.aspx?o=1025 [Accessed 2008 Nov 27]

47. Agency for Healthcare Research and Quality. US Preventive Services Task Force [online]. Available from URL: http://www.ahrq.gov/clinic/uspstf/uspstopics.htm [Accessed 2008 Nov 27]

48. Consolidated Standards of Reporting Trials (CONSORT) [online]. Available from URL: http://www.consort-statement.org/?o=1011 [Accessed 2008 Nov 27]

49. The Cochrane Collaboration [online]. Available from URL: http://www.cochrane.org/ [Accessed 2008 Nov 27]

50. Des Jarlais DC, Lyles C, Crepaz N, the TREND Group. Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: the TREND statement. Am J Public Health 2004; 94 (3): 361-6

51. Chen A, Brown R, Archibald N, et al. Best practices in coordinated care. Mathematica 2000 [online]. Available from URL: http://www.mathematica-mpr.com/publications/PDFs/bestpractices.pdf [Accessed 2008 Nov 27]

52. Wilson TW, Gruen J, Thar W, et al. Assessing return on investment of defined-population disease management interventions. Jt Comm J Qual Saf 2004; 30 (11): 614-21

53. Fitzner K, Siderov J, Fetterolf D, et al. Principles for assessing disease management outcomes. Dis Manage 2004; 7: 191-201

54. Faxon DP, Schwamm LH, Pasternak RC, et al. American Heart Association's Expert Panel on Disease Management. Improving quality of care through disease management: principles and recommendations from the American Heart Association's Expert Panel on Disease Management. Circulation 2004 Jun 1; 109 (21): 2651-4

55. Krumholz HM, Brindis RG, Brush JE, et al. Standards for statistical models used for public reporting of health outcomes. Circulation 2006; 113: 456-62

56. Linden A, Roberts N. A user's guide to the disease management literature: recommendations for reporting and assessing program outcomes. Am J Manage Care 2005; 11 (2): 81-90

57. Duncan I, Owen R, Dove H. Testing actuarial methods for evaluating disease management savings outcomes. Society of Actuaries, 2006 [online]. Available from URL: http://www.soa.org/research/files/pdf/Paper_8_TestingActuarial%20 Methods.pdf [Accessed 2008 Nov 27]

Correspondence: Dr *Thomas Wilson*, Population Health Impact Institute, 10663 Loveland-Madeira Rd. #210, Loveland, OH 45140, USA.
E-mail: twilson@phiinstitute.org