Economic Evaluations of Multicomponent Disease Management Programs with Markov Models: A Systematic Review

Florian Kirsch 1,2,*

1Munich School of Management and Munich Center of Health Sciences, Ludwig-Maximilians-University, Munich, Germany; 2Institute for Health Economics and Health Care Management, Helmholtz Zentrum München, Member of the German Center for Diabetes Research (DZD), Neuherberg, Germany

ABSTRACT

Background: Disease management programs (DMPs) for chronic diseases are being increasingly implemented worldwide. Objectives: To present a systematic overview of the economic effects of DMPs with Markov models. The quality of the models is assessed, the method by which the DMP intervention is incorporated into the model is examined, and the differences in the structure and data used in the models are considered. Methods: A literature search was conducted; the Preferred Reporting Items for Systematic Reviews and Meta-Analyses statement was followed to ensure systematic selection of the articles. Study characteristics e.g. results, the intensity of the DMP and usual care, model design, time horizon, discount rates, utility measures, and cost-of-illusion were extracted from the reviewed studies. Model quality was assessed by two researchers with two different appraisals: one proposed by Philips et al. (Good practice guidelines for decision-analytic modelling in health technology assessment: a review and consolidation of quality assessment. Pharmacoconomics 2006;24:355-71) and the other proposed by Caro et al. (Questionnaire to assess relevance and credibility of modeling studies for informing health care decision making: an ISPOR-AMCP-NPC Good Practice Task Force report. Value Health 2014;17:174-82). Results: A total of 16 studies (9 on chronic heart disease, 2 on asthma, and 5 on diabetes) met the inclusion criteria. Five studies reported cost savings and 11 studies reported additional costs. In the quality, the overall score of the models ranged from 39% to 65%, it ranged from 34% to 52%. Eleven models integrated effectiveness derived from a clinical trial or a meta-analysis of complete DMPs and only five models combined intervention effects from different sources into a DMP. The main limitations of the models are bad reporting practice and the variation in the selection of input parameters. Conclusions: Eleven of the 14 studies reported cost-effectiveness results of less than $30,000 per quality-adjusted life-year and the remaining two studies less than $30,000 per life-year gained. Nevertheless, if the reporting and selection of data problems are addressed, then Markov models should provide more reliable information for decision makers, because understanding under what circumstances a DMP is cost-effective is an important determinant of efficient resource allocation. Keywords: chronic disease, cost-effectiveness, DMP, Markov model, review.

Copyright © 2016, International Society for Pharmacoconomics and Outcomes Research (ISPOR). Published by Elsevier Inc.
quality of the models is assessed, and how the DMP intervention was incorporated into the model and how it fitted are examined. Finally, differences in the structure and data of the models on outcomes are estimated.

**Methods**

**Data Sources and Searches**

We followed the instructions of the standards of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (see Appendix I in Supplemental Materials found at http://dx.doi.org/10.1016/j.jval.2016.07.004) [11]. A systematic literature search was conducted on June 26, 2015 (see Appendix II in Supplemental Materials found at http://dx.doi.org/10.1016/j.jval.2016.07.004). The databases PubMed, Embase, Business Source Complete, and EconLit were screened for articles using the following search terms: disease management/disease management programme/management program; decision analytic/model/Markov; chronic disease/COPD/asthma/breast cancer/diabetes/coronary/heart. After removing duplicates, abstracts were screened. For inclusion, the following a priori defined criteria had to be fulfilled: 1) effects and costs were considered, 2) publication year was after 1995, 3) a Markov model was used, 4) articles were in either English or German, and 5) physicians and patients played an active role in the DMP process. Subsequently, all articles that were considered potentially eligible by at least one reviewer were subject to full-text analysis. The reference lists of these studies were searched to find additional relevant literature. Any disagreements on inclusion of studies were solved by consensus.

**Data Extraction**

Data extraction was performed by one investigator. The fulfillment of the requirements of the Disease Management Association of America on full-service DMPs was considered. As far as we know, there is no standard or validated method to grade the intensity of a DMP or of usual care. Therefore, we took a pragmatic approach. An intervention that met all six requirements was assessed as a high-intensity treatment, one that met five was medium intensity, and one that met fewer than five was low intensity. The degree of care in the control group was divided into low-, medium-, and high-intensity care. High intensity was achieved if a management plan and patient education were provided, medium intensity if one of these was provided, and low intensity if neither was provided.

Differences in costing year were addressed by using the gross domestic product index of the Organisation for Economic Co-operation and Development [12]. First, the costs were inflated to the price year 2011 in the original country and then converted into US $ to US purchasing power parities.

**Quality Assessment**

Model quality was assessed with two appraisals: one proposed by Philips et al. [13] in 2006 and the other proposed by Caro et al. [14] in 2014. One amendment was made to the framework of the appraisal of Philips et al. [13]: quality items concerning costs from the 2004 appraisal [15] were added. Because this review is not written from the perspective of a specific decision maker, the first part of the Caro et al. [14] appraisal, which addressed the extent to which the results of the model apply to the setting of interest to the decision maker, was not considered further. All items in the quality appraisals evaluated as not available were rated as not fulfilled for the descriptive analysis of the quality appraisals. The quality assessment was performed by two researchers, and any disagreements on the rating of items were solved by consensus.

**Results**

**Literature Search**

The literature search (Fig. 1) yielded a total of 3180 citations. After removing 799 duplications, 2381 abstracts were screened and the full-text articles of 66 citations plus 39 citations from reference lists were reviewed. A total of 16 studies met the inclusion criteria: 9 studies focused on chronic heart disease, 2 on asthma, and 5 on diabetes.

Additional study characteristics can be found in Tables 1 to 3, and the costs and utility values incorporated into the model are provided in Tables I and II in Supplemental Materials found at http://dx.doi.org/10.1016/j.jval.2016.07.004.

**Economic Results**

For DMPs in chronic heart disease, the results ranged from cost savings of $657 and an increase of 0.0051 quality-adjusted life-years (QALYs) [16,17] to additional costs of $4,607 per life-year gained (LYG) [18] and $146,544 per QALY [19]. Three studies [16–18] reported cost savings and six studies [19–24] reported additional costs.

The two studies for asthma reported cost savings of $798 and a gain of 0.62 QALYs [22], or additional costs of $3635 per QALY [25]. The remaining five studies for diabetes reported results from cost savings, which were not specified further in one study [26], to additional costs, of up to $21,701 per LYG [27] and $85,087 per QALY [24].

**Quality Assessment**

In the Philips et al. [13] quality appraisal, the overall score results for chronic heart disease ranged from 39% [18] to 65% [23], for asthma from 53% [25] to 58% [22], and for diabetes from 45% [27] to 53% [28] (see Table III in Supplemental Materials found at http://dx.doi.org/10.1016/j.jval.2016.07.004, and also Fig. 2).

For the Caro et al. [14] quality appraisal, the overall scores for chronic heart disease ranged from 34% [18] to 52% [29,21,23,29], for asthma from 31% [25] to 38% [22], and for diabetes from 34% [28] to 51% [30] (see Table IV in Supplemental Materials found at http://dx.doi.org/10.1016/j.jval.2016.07.004, and also Fig. 3).

The performance of the models for the subdimensions and each item for the Philips et al. [13] and Caro et al. [14] quality appraisals can be found in Tables III and IV in Supplemental Materials and in Figures 2 and 3. Chronic heart disease models performed the best, with average overall scores of 55% and 48% in the Philips et al. [13] and Caro et al. [14] quality appraisals, respectively, compared with 56% and 36% for asthma and 50% and 43% for diabetes. In the Philips et al. [13] quality appraisal, the chronic heart disease models performed the best in the dimension structure, with an average overall score of 65% versus 61% for asthma and 62% for diabetes.

There was only a slightly positive trend in the quality of results over time, and the average overall score from models published from 2010 onward [16,17,19,26,31,32] was 57% versus 51% from models published before 2010 [18,20–25,27,28,30] in the Philips et al. [13] quality appraisal and was 47% versus 45% in the Caro et al. [14] quality appraisal. To see whether the models were ranked in the same order in both quality appraisals, the Spearman rank correlation coefficient was calculated as 0.262 with a t value of 1.016. This low value shows that the correlation between the two quality appraisals is weak, although the positive value
suggests that a high score in one tends to lead to a high score in the other quality appraisal.

Incorporation of DMPs into the Models

Nearly all models [16–18,20–23,25,26,29,30] integrated effectiveness derived from a clinical trial or a meta-analysis of complete DMPs. Only five models [19,24,27,28,32], two for chronic heart disease [19,32] and three for diabetes [24,27,28], used different sources and combined the different interventions into a DMP.

Gandjour [32] constructed a model for hypertension, and the DMP reduced the nonadherence rate in patients and in physicians, leading to a relative risk reduction in stroke, myocardial infarction (MI), and death [33]. For MI, Ito et al. [19] evaluated improvements in medication adherence from informational, behavioral, and complex interventions and from combinations of these interventions. The effectiveness rates for each single intervention were derived from systematic reviews and meta-analyses [34–36], and for the combined interventions the effectiveness rates were based on the authors’ assumptions that were not described further in the reference.

The remaining three models were for diabetes. Huang et al. [28] used different sources for the effects of angiotensin-converting-enzyme (ACE) inhibitors [37,38] and aspirin [39], and stated that they assumed that the joint effect of aspirin and ACE inhibitors on cardiovascular effects was multiplicative. Gilmer et al. [24] reported that a clinical database stored the medical and epidemiological data, which were not described further in the study. The data were directly implemented in the model to calculate clinical outcomes. The interaction of the effects, however, remained unclear. Palmer et al. [27] incorporated probability reductions for ACE inhibitors and intensive therapy from different sources and added several further interventions to usual care and to intensive therapy, and thus it is not clear whether the additional effects were additive or multiplicative.

Structure

The bubble charts of the Markov models can be found in Appendix III in Supplemental Materials found at http://dx.doi.org/10.1016/j.jval.2016.07.004.

Chronic heart disease

Chronic heart disease encompasses many different diseases, including chronic heart failure, hypertension, congestive heart failure (CHF), MI, and coronary heart disease (CHD).

The chronic heart failure models ranged from five states [21] to over six [18] and eight states [29]. Starting at index hospitalization [21,29] or no previous hospitalization [18], further hospitalizations were possible before death occurred.

The two hypertension models in that review were published by Gandjour [23,32] and both had five states (no cardiovascular disease, renal failure, MI, stroke, and death).

Only the model reported by Miller et al. [20] dealt with CHF and had four states (New York Heart Association [NYHA] classification I, II, III and IV, and death).
<table>
<thead>
<tr>
<th>Author</th>
<th>Country</th>
<th>Disease</th>
<th>Population identification process</th>
<th>Evidence-based practice guidelines</th>
<th>Collaborative practice models</th>
<th>Patient self-management education</th>
<th>Process and outcomes measurement, evaluation, and management</th>
<th>Routine reporting/feedback loop</th>
<th>Intervention group</th>
<th>Control group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moertl et al. [29]</td>
<td>Austria and Canada</td>
<td>Chronic HF</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>6/6 high intense</td>
<td>Medium intense (management plan)</td>
<td>Medium intense (patient education)</td>
</tr>
<tr>
<td>Chan et al. [18]</td>
<td>United States</td>
<td>HF</td>
<td>✓</td>
<td>x</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>5/6 medium intense</td>
<td>Medium intense (patient education)</td>
<td>Low intense</td>
</tr>
<tr>
<td>Gohler et al. [21]</td>
<td>Germany</td>
<td>Chronic HF</td>
<td>✓</td>
<td>x</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>5/6 medium intense</td>
<td>5/6 medium intense</td>
<td>Low intense</td>
</tr>
<tr>
<td>Gandjour [32]</td>
<td>Germany</td>
<td>Hypertension</td>
<td>✓</td>
<td>✓</td>
<td>x</td>
<td>✓</td>
<td>✓</td>
<td>5/6 medium intense</td>
<td>5/6 medium intense</td>
<td>Low intense</td>
</tr>
<tr>
<td>Gandjour and Stock [23]</td>
<td>Germany</td>
<td>Hypertension</td>
<td>✓</td>
<td>✓</td>
<td>x</td>
<td>✓</td>
<td>✓</td>
<td>5/6 medium intense</td>
<td>5/6 medium intense</td>
<td>Low intense</td>
</tr>
<tr>
<td>Miller et al. [20]</td>
<td>United States</td>
<td>CHF</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>6/6 high intense</td>
<td>6/6 high intense</td>
<td>Low intense</td>
</tr>
<tr>
<td>Gillespie et al. [17]</td>
<td>Ireland</td>
<td>CHD</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>6/6 high intense</td>
<td>6/6 high intense</td>
<td>Northern Ireland: medium intense (management plan); Republic of Ireland: low intense</td>
</tr>
<tr>
<td>Gillespie et al. [16]</td>
<td>Ireland</td>
<td>CHD</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>6/6 high intense</td>
<td>6/6 high intense</td>
<td>Northern Ireland: medium intense (management plan); Republic of Ireland: low intense</td>
</tr>
<tr>
<td>Ito et al. [19]</td>
<td>United States</td>
<td>MI</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>6/6 high intense</td>
<td>Low intense</td>
<td></td>
</tr>
<tr>
<td>Steuten et al. [22]</td>
<td>The Netherlands</td>
<td>Asthma</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>6/6 high intense</td>
<td>Medium intense (patient education)</td>
<td>Low intense</td>
</tr>
<tr>
<td>Gordois et al. [25]</td>
<td>Australia</td>
<td>Asthma</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>6/6 high intense</td>
<td>Low intense</td>
<td></td>
</tr>
<tr>
<td>Yu et al. [26]</td>
<td>United States</td>
<td>Type 2 diabetes</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>6/6 high intense</td>
<td>Medium intense (management plan)</td>
<td>Low intense</td>
</tr>
<tr>
<td>Huang et al. [28]</td>
<td>United States</td>
<td>Type 2 diabetes</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>6/6 high intense</td>
<td>Medium intense (management plan)</td>
<td>Low intense</td>
</tr>
</tbody>
</table>

Continued on next page.
The two models for coronary heart disease (CHD) were reported by Gillespie et al. [16,17] and had four states (CHD, coronary events, post-MI, and death) and an identical structure. The MI model developed by Ito et al. [19] included four states (post-MI, post-MI and stroke, post-MI and CHF, and second post-MI and stroke).

The structure of each model seemed adequate for reflecting the disease and its progression. Although further states and comorbidities could have been considered, the models would have been more complicated.

Asthma
For asthma, the model by Gordois et al. [25] included four states (mild, moderate, and severe asthma, and death) and the model by Steuten et al. [22] included five states (successful control, suboptimal control, primary care–managed exacerbation, hospital-managed exacerbation, and death from all causes).

Diabetes
Models for diabetes covered type 1 or type 2 diabetes or both. The number of states varied from 4 [30] to 15 [24]. For one model, the number of states was not reported [28]. The model by Mason et al. [30] had 4 states (healthy, stroke, MI, and death) and the one by Gilmer et al. [24] had 15 states (angina, MI, CHF, stroke, peripheral vascular disease, diabetic retinopathy, macula edema, cataract, hypoglycemia, ketoacidosis, lactic acidosis, nephropathy, neuropathy, foot ulcers and amputation, and nonspecific mortality). Overall, the diabetes models were similar, even though the models considered cardiovascular comorbidities of diabetes with methods that were more or less complex and comprehensive, and with different numbers of possible disease states.

Data
DMP and usual care
Yearly costs for DMPs for chronic heart disease ranged from $229 [23] to $3610 [20], for asthma from $136 [22] to $198 [25], and for diabetes from $450 [28] to $2554 [24] per patient per year.

The description of the overall package included in the DMP and the usual care was sometimes detailed but mostly superficial. Nevertheless, for chronic heart disease, five DMPs [16,17,19,20,29] were rated as high intensity and four DMPs [18,21,23,32] as medium intensity; for asthma, both DMPs [22,25] were rated as high intensity; and for diabetes, four DMPs [24,26,28,30] were rated as high intensity and one DMP [27] as low intensity.

In addition, the intensity of the usual care was categorized, and the level of care in all three indication areas ranged from low intensity [16,17,19–21,23–26,32] to medium intensity [18,22,27–30].

Methodological inconsistencies in the classifications make the estimation of true differences difficult. All medium-intensity DMPs [18,21,23,32], all for chronic heart disease, resulted in additional costs ranging from $1,091 [23] to $19,379 [32] per QALY gained. The low-intensity DMP [27] conducted in diabetes, depending on the combination of interventions, was partially dominant. Some of the high-intensity DMPs [15,17,22,26] were dominant (two [16,17] for chronic heart disease, one [22] for asthma, and one [26] for diabetes), and others [19,20,24,25,28–30] resulted in additional costs per QALY (three [19,20,29] for chronic heart disease ranging from $13,673 [29] to $146,544, one [25] for asthma of $3,635, and three [24,28,30] for diabetes ranging from $4,971 [30] to $85,087 [24] per QALY).

Time horizon
The time horizons used by the models for chronic heart disease were 15 years [18], 20 years [29], 35 years [23], 50 years [16], and 55
<table>
<thead>
<tr>
<th>Author</th>
<th>Effectiveness of intervention</th>
<th>Cost of intervention (2011 US $ prices)</th>
<th>Perspective</th>
<th>Model design</th>
<th>Study population</th>
<th>Study design intervention data derived</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moertl et al. [29]</td>
<td>Different transition probabilities for Rehospitalization: UC: 1.2, nurse-led MC: 0.8; BMC: 0.4</td>
<td>Austria/Canada: UC: $124/$217</td>
<td>Third-party payer</td>
<td>Eight-state Markov model: No additional hospitalization, first additional hospitalization, status after 1 additional hospitalization, second additional hospitalization, status after 2 additional hospitalizations, third or more additional hospitalization, status after at least 3 additional hospitalizations, and death</td>
<td>278 adults discharged from HF hospitalization</td>
<td>Clinical trial</td>
</tr>
<tr>
<td>Chan et al. [18]</td>
<td>Different transition probabilities for UC/disease management for Death hospitalized: 0.203/0.192</td>
<td>$864 in the first year, then the intervention stops</td>
<td>Not in text</td>
<td>Six-state Markov model: No previous hospitalization, 1 previous hospitalizations, 2 previous hospitalizations, 3 previous hospitalizations, 4+ previous hospitalizations, and death</td>
<td>1000 patients with HF: ejection fractions of ≤35% on ACE inhibitors and other conventional drugs</td>
<td>Clinical trial</td>
</tr>
<tr>
<td>Göhler et al. [21]</td>
<td>Transition probabilities Control group: Index hospitalization: hospitalized death 0.010, nonhospitalized death 0.168, rehospitalization 0.168; Rehospitalization 1: hospitalized death 0.075; nonhospitalized death 0.012; rehospitalization 0.213; Rehospitalization 2: hospitalized death 0.085; nonhospitalized death 0.013; rehospitalization 0.268; Rehospitalization 3+: hospitalized death 0.095; nonhospitalized death 0.015; rehospitalization 0.334 Intervention group: Relative risk for all-cause mortality: 0.81 Relative risk for all-cause rehospitalization: 0.84</td>
<td>$573 initiation costs</td>
<td>Societal perspective</td>
<td>Five-state Markov model: Index hospitalization, rehospitalization 1 to 3+, and death</td>
<td>Adults with CHF (mean age 67 y) on the basis of German data in the EuroHeart Failure Survey</td>
<td>Meta-analysis</td>
</tr>
<tr>
<td>Gandjour [32]</td>
<td>Noncompliance reduction: patients 20% and physicians 9% [19,50] Relative risk of adherent patients for: Stroke: 0.71 MI: 0.79 Death: 0.90</td>
<td>$1231 per patient per year</td>
<td>Third-party payer</td>
<td>Five-state Markov model: No cardiovascular disease, MI, stroke (including transient ischemia), renal failure, and death</td>
<td>Adults from 40 to 69 y with hypertension and without cardiovascular disease at model start</td>
<td>Clinical trial</td>
</tr>
<tr>
<td>Gandjour and Stock [23]</td>
<td>Relative risk of patients in the disease management group for: Stroke: 0.71 MI: 0.79 Death: 0.90</td>
<td>$229 per patient per year</td>
<td>Third-party payer</td>
<td>Five-state Markov model: No cardiovascular disease, MI, stroke (including transient ischemia), renal failure, and death</td>
<td>Hypothetical cohort of hypertensive male and female patients without cardiovascular disease at different</td>
<td>Mainly meta-analyses</td>
</tr>
<tr>
<td>Study</td>
<td>Approach</td>
<td>Patient Population</td>
<td>Transition Probabilities</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------------</td>
<td>--------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| Miller et al. [20]    | Different transition probabilities for control group/disease management group from Month 0-6: NYHA I: to I 57.9%/67.4%; to II 36.8%/29.2%; to III & IV 5.3%/3.4%; NYHA II: to I 18.3%/21.6%; to II 66.4%/64.8%; to III & IV 15.4%/13.6%; NYHA III & IV: to I 5.1%/7.4%; to II 48.7%/35.8%; to III & IV 46.2%/56.8% | 1069 adult patients with echocardiographic evidence of CHF (751 of them suffered from systolic heart failure) | Clinical trial
| Gillespie et al. [17] | Different transition probabilities for control group/intervention group for coronary events (stable angina, unstable angina, MI and CHD death), post-MI, and death. No numbers stated in the text or the technical appendix | 838 adults with CHD excluding patients with angina only and diabetes | Clinical trial
| Gillespie et al. [16] | Different transition probabilities for control group/intervention group for coronary events (stable angina, unstable angina, MI and CHD death), post-MI, and death. No numbers stated in the text or the technical appendix | 903 adult patients with CHD (documented MI, CABG, angioplasty, or angina) | Clinical trial
| Ito et al. [19]       | Relative risk of post-MI patients by number of drugs                      | Hypothetical cohort of patients 65 y of age who were discharged alive after MI and were prescribed all four classes of guideline-recommended post-MI therapy | Clinical trial
| Steuten et al. [22]   | Different transition probabilities UC/disease management from: Successful control: to successful control 0.956/0.972; to suboptimal control 0.024/0.016; to primary care exacerbation 0.17/0.10; and hospital exacerbation 0.003/0.002 Suboptimal control: to successful control 0.100/0.300; to suboptimal control 0.885/0.691; to primary care exacerbation 0.014/0.008; and hospital exacerbation 0.001/0.001 Primary care exacerbation: to successful control 0.031/0.146; to suboptimal control 0.225/0.156; to primary care exacerbation 0.576/0.568; and hospital exacerbation 0.168/0.130 | 658 adults with asthma | Clinical trial

Disease management: 
- $368; mailed education $10; polypills $2477; pharmacotherapy $259 (2012) to $211 (2018)

Stroke: 
1 drug: 0.87
2 drugs: 0.69
3 drugs: 0.61
4 drugs: 0.47

CHF: 
1 drug: 0.90
2 drugs: 0.82
3 drugs: 0.73
4 drugs: 0.65

$3610 per patient per year Third-party payer

$273 per patient per year Third-party payer

$273 per patient per year Third-party payer
<table>
<thead>
<tr>
<th>Author</th>
<th>Effectiveness of intervention</th>
<th>Cost of intervention (2011 US $ prices)</th>
<th>Perspective</th>
<th>Model design</th>
<th>Study population</th>
<th>Study design intervention data derived</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gordois et al. [25]</td>
<td>Hospital exacerbation: to successful control 0.008/0.222; to suboptimal control 0.442/0.341; to primary care exacerbation 0.149/0.228; and hospital exacerbation 0.329/0.209</td>
<td>$991 (per patient in 5 y)</td>
<td>Third-party payer</td>
<td>Four-state Markov model: Mild, moderate, severe asthma and death</td>
<td>351 patients with asthma</td>
<td>Clinical trial</td>
</tr>
<tr>
<td>Yu et al. [26]</td>
<td>Different transition probabilities for enhanced care group and control group for nonfatal/fatal CHD and nonfatal/fatal stroke: Control group</td>
<td>Not found in text</td>
<td>Third-party payer</td>
<td>Eleven-state Markov model: Not described further</td>
<td>611 adult patients with diabetes type 2 (204 in intervention group and 407 in control group)</td>
<td>Retrospective matched-cohort study</td>
</tr>
<tr>
<td>Huang et al. [28]</td>
<td>Different transition probabilities for control group and diabetes management group. Glucose control beneficial effect on microvascular events but not on cardiovascular events (estimated with the UKPDS risk prediction). ACE inhibitors benefits not further described. Aspirin reduces the probability of CHD and increases the probability of gastrointestinal bleed. Comparison of probabilities or differences not stated in the text or the technical Appendix.</td>
<td>Year 1: $848</td>
<td>Societal perspective</td>
<td>Markov model (number of states not in text)</td>
<td>Nonpregnant adult patients (18-75 y) with diabetes</td>
<td>Serial cross-sectional study</td>
</tr>
<tr>
<td>Gilmer et al. [24]</td>
<td>Not mentioned in text. In the technical appendix, the following can be found: &quot;clinical database: it is a set of...”</td>
<td>Year 1: $2554; Year 2 and beyond: $2222</td>
<td>Third-party payer</td>
<td>Fifteen state Markov model: Angina, MI, CHF, stroke, peripheral vascular disease,</td>
<td>3893 patients with diabetes</td>
<td>Pre-post observational study</td>
</tr>
<tr>
<td>Study</td>
<td>Methodology</td>
<td>Risk Factors</td>
<td>Subject Information</td>
<td>Cost Information</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>-------------</td>
<td>--------------</td>
<td>---------------------</td>
<td>-----------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mason et al. [30]</td>
<td>Transition probabilities not mentioned in text. Only the following can be found: “The model applies Framingham risk equations each year to patients in the healthy state to calculate the risk of suffering a stroke or MI according to their age, sex, and cardiovascular risk factors.”</td>
<td>Diabetic retinopathy, macula edema, cataract, hypoglycemia, ketoacidosis, lactic acidosis, nephropathy, neuropathy, foot ulcer and amputation, and nonspecific mortality</td>
<td>1407 subjects attending the diabetes center at Hope Hospital, Salford, UK, with raised blood pressure and hyperlipidemia</td>
<td>Clinical trial</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Palmer et al. [27]</td>
<td>Reduction of relative risk for complications of diabetes for the single interventions</td>
<td></td>
<td>19-y-old</td>
<td>Clinical trial</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ACE: 55% reduction in the rate of annual increase in urinary albumin excretion</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ACE: 50% reduction in probability (macroalbuminuria to end-stage renal disease)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Intensive therapy: 76% reduction in incidence background retinopathy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Intensive therapy: 47% reduction in progression background retinopathy to proliferative retinopathy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Intensive therapy: 0% or 41% reduction in diabetes-attributed probability AMI/probability stroke/probability amputation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

ACE, angiotensin-converting-enzyme; AMI, acute myocardial infarction; BMC, intensive patient management; CABG, coronary artery bypass grafting; CHD, coronary heart disease; CHF, congestive heart failure; HF, heart failure; MC, multidisciplinary care; MI, myocardial infarction; NYHA, New York Heart Association; RR, relative risk; UC, usual care; UKPDS, U.K. Prospective Diabetes Study.
years [32] to lifetime [17,20,21]; for asthma, both models [22,25] used a time horizon of 5 years; and for diabetes, the time horizon range was from 10 years [26] and over 40 years [24] to lifetime [27,28,30].

Discount rates

For chronic heart disease, the discount rates ranged from 3% [19,20,23,32] to more than 3.5% [16,17] and to 5% [21,29]; for asthma, between 4% [22] and 5% [25]; for diabetes, between 3% [24,26–28] and 5% [26,30] for costs and effects in the base case. Only Yu et al. [26] used different discount rates for costs (5%) and effects (3%) in diabetes.

Utility weights

Fourteen [16,17,19–26,28,30–32] out of 16 models estimated QALYs as the outcome measure, and thus incorporated utility values (see Table I in Supplemental Materials).

Utility values across indication area. The models were conducted in three indication areas; therefore, many possible disease states were considered. Nonetheless, there were overlaps in chronic heart disease and diabetes, which both included stroke, CAD, MI, and renal failure. For asthma, no overlaps with diabetes or chronic heart disease were identified. The incorporated values for stroke (six models [19,23,26,28,30,32]) ranged from 0.5 [30] to 0.675 [26]; for CAD (three models) from 0.6877 [16,17] to 0.77 [26]; for MI (six models) from 0.6542 [17] to 0.88 [23,28,30,32]; and for renal failure (three models [23,28,32]), 0.61. The highest absolute spread could be found in MI, but only Gillespie et al. [17] used a utility value of 0.6542, whereas all other models incorporated a value higher than 0.77. The highest percentage spread of 35% was found in stroke.

Utility values within indication area. In chronic heart disease, Moertl et al. [29] and Göhler et al. [21] used the highest utility values in the models for chronic heart failure, followed by the value used by Ito et al. [19] in their MI model. Gandjour and Stock [23] and Gandjour [32] used the same utility values in their hypertension models, which were lower than the utility values in the chronic heart failure and MI models. Miller et al. [20] and Gillespie et al. [16,17] used the lowest utility values in their CHF and CHD models.

For asthma, Gordois et al. [25] used higher utility values than did Steuten et al. [22] on average, even when the disease states were not totally comparable.

For well-being in diabetes, Yu et al. [26] incorporated the lowest value and Huang et al. [28] and Mason et al. [30] the highest. Moreover, the incorporated values for comparable complications were similar.

The differences in incorporated utility values were greater in the models for chronic heart disease and asthma than for diabetes.

Cost of illness

Although not every model considered all kinds of costs, as far as was possible, a quantitative comparison was conducted (see Table II in Supplemental Materials).

In chronic heart disease, Moertl et al. [29] used hospitalization costs ($12,819/$19,807) more than twice those used by Chan et al. [18] ($5,759) and Göhler et al. [21] ($5,960). Gandjour and Stock [23] and Gandjour [32] used a rate for a physician visit ($66) that was twice that used by Moertl et al. [29] ($23/$35). Göhler et al. [21] used medication costs ($655) that were 2 to 3 times higher than those used in all other models [19,23,29,32].

For asthma, Steuten et al. [22] incorporated the highest costs in every category.

For diabetes, Gilmer et al. [24] used the highest medication costs ($1,872) compared with $639 [27] and $1,477 in other models [26]. Yu et al. [26] used the highest costs for CHD ($37,462) and stroke ($32,916) compared with those used by Palmer et al. [27] (stroke: $26,197) and Huang et al. [28] (stroke: year 1 survivor $26,362; year 1 nonsurvivor $41,633; CHD: $12,104).

Sensitivity of parameters

Seven [16,17,19–21,23,29] of the nine chronic heart disease models, both [22,25] the asthma models, and four [24,26,28,30] of the five diabetes models (i.e., a total of 13 out of 16 models) conducted univariate sensitivity analyses.

Funding

Independent of the indication area, industry-funded studies or studies without a conflict of interest statement [18,20,22,23] were more likely to report a worse cost-effectiveness ratio than studies supported by an institutional grant [16,17,21] or without financial support [32]. For chronic heart disease, one study [29] was industry-funded, for asthma none was, and for diabetes three [24,27,30] out of five were.

Discussion

Economic Results

Fourteen studies [16,17,19–26,28,30–32] reported QALYs and two studies [18,27] reported LYG as outcome measures. The results ranged from cost savings of $657 and a gain of 0.0051 QALYs [16,17] to additional costs of $4,607 per LYG and $146,544 per QALY [19] for chronic heart disease; from cost savings of $798 and a gain of 0.0051 QALY [19] for chronic heart disease; from cost savings of $798 and a gain of 0.0051 QALY [19] for chronic heart disease; from cost savings of $798 and a gain of 0.0051 QALY [19] for chronic heart disease; from cost savings of $798 and a gain of 0.0051 QALY [19] for chronic heart disease; from cost savings of $798 and a gain of 0.0051 QALY [19] for chronic heart disease; from cost savings of $798 and a gain of 0.0051 QALY [19] for chronic heart disease; from cost savings of $798 and a gain of 0.0051 QALY [19] for chronic heart disease; from cost savings of $798 and a gain of 0.0051 QALY [19] for chronic heart disease.

For chronic heart disease, one study [29] was industry-funded, for asthma none was, and for diabetes three [24,27,30] out of five were.

Quality of the Models

On the basis of the reported percentages presented in the Results section, the quality of the models was far from perfect. The major weaknesses were in the consistency of the dimensions in the Philips et al. [13] quality appraisal and in the reporting of dimensions and validation in the Caro et al. [14] quality appraisal. This low percentage for consistency in the Philips et al. [13] quality appraisal can be explained by the fact that three out of five items in the consistency subdimension refer to model validation, and they were not performed or at least not reported alongside the models. Nonetheless, this is consistent with the low score the models achieved in the list reported by Caro et al. [14], because reporting of model characteristics and input parameters was partially opaque and incomplete. The weak value of the Spearman correlation coefficient may have resulted from the slight difference in focus of the two quality appraisals.
Incorporation of DMPs into the Models

In contrast to pure pharmaceutical interventions, DMPs are a broad array of different interventions, and consequently there a huge number of possible combinations for DMPs, for which all the incremental cost-effectiveness ratios should be evaluated. Markov models can obtain these ratios by assuming that combining these interventions has additive and multiplicative effects. The main approach used is relative static, similar to the economic evaluation of pure pharmaceutical products [39], which ignores the capacity Markov models provide for evaluating multicomponent interventions. Nearly all models [16–18,20–25,26,29,30] use effectiveness rates derived from complete DMPs. The advantage of this approach is that both overestimation and underestimation caused by the accumulation of independently determined effects are avoided. Nevertheless, there are also disadvantages; because only the complete intervention is evaluated, it is not clear what contribution the individual treatments in the DMP make to the overall effect. Only five models [19,24,27,28,32] combined the effects of different interventions to obtain the right DMP for their purpose. Three [24,27,28] of these studies were conducted in diabetes. Remarkably, two [24,28] of the three studies did not report transition probabilities in a transparent way, meaning that how the individual effects of interventions combined in a DMP were incorporated into the model (e.g., multiplicative or additive) could not be reconstructed. If different sources are used to assemble a DMP, the modelers should mention and justify the assumptions made about how the effects of the interventions interact with each other. This type of approach, similar to the approaches of Palmer et al. [27] for diabetes and Ito et al. [19] for chronic heart disease, would be particularly valuable, despite the uncertainty that arises from combining the interventions, because this would allow a hierarchical ranking of partial interventions and enable prioritization of medical services. No comment can be made on the reliability of the effect values because the quality of the origins of the data was not assessed. Although there were restrictions in the level of detail, the descriptions were consistent with how the interventions were incorporated into the models, irrespective of the fact that the effects were taken from a complete DMP or were combined from single interventions into a DMP.

Structure
An increasing number of health care states considered by a model reflect a higher number of considered comorbidities that can be caused by a disease. Generally, a Markov model with more possible states considering more disease-related costs, which can be avoided by successful disease control, leads to a favorable incremental cost-effectiveness ratio [39].

For chronic heart disease, the model structures were similar; therefore, a potential overestimation or underestimation of the results because of the model structure can be neglected.

For asthma, however, exacerbations are the main driver of cost of illness, and thus the inclusion [22] or exclusion [25] of exacerbations should have a high impact on results. Gordois et al. [25] did not state the cost of illness or miscellaneous costs, and the costs for ambulatory care, medication, and inpatient care were higher in the model by Steuten et al. [22]. If the unlisted cost of illness and miscellaneous costs incorporated by Gordois et al. [25] are not substantially higher than the costs used by Steuten et al. [22], it is safe to expect an underestimation of the considered disease-related costs by Gordois et al. [25].

For diabetes, the number of states varies greatly because DMPs lead to better disease control and prevention of complications. Consequently, more disease states should lead to a better incremental cost-effectiveness ratio for the intervention program, even when the basic orientation of all diabetes models (including cardiovascular diseases as comorbidities) is the same.

Data

DMP and usual care

Normally, high-intensity interventions are more effective but are also more expensive. If we ignore the poor reporting, which made classification difficult, low- [27] or high-intensity DMPs [16,17,19,20,22,24–26,28–30] are more cost-effective than medium-intensity DMPs [18,21,23,32]. This may be because DMPs are less effective or the usual care is much better for chronic heart disease, than for asthma or diabetes, because all medium-intensity DMPs were conducted for chronic heart disease.

The description of the usual care content was worse than that of the interventions. Usual care, however, is particularly important in a strongly context-dependent intervention, such as a DMP, the incremental effectiveness of which depends on background clinical practice patterns [10]. For DMPs, exactly the same rigor in describing the intervention and usual care should be applied as in pharmaceutical clinical trials, because even usual care for chronic diseases may be a multicomponent intervention.

Time horizon

The time horizon is particularly crucial for chronic conditions, because the economic benefits of DMPs of the reduction in resource utilization may not become evident in a short time frame. Some findings suggest that it takes at least 3 to 5 years, because of time lags in reaching full implementation, to obtain full program effectiveness [40]. Therefore, the short time horizons considered in asthma may have a negative impact on the cost-effectiveness ratio. Nevertheless, DMPs are ongoing interventions in which disease control takes priority. In contrast, in a once-only intervention, such as a pharmaceutical smoking cessation intervention, the positive impact of the intervention takes time to become evident, but no further costs arise. The positive impact of a longer time horizon on cost-effectiveness for a DMP should be smaller. Only the study by Gandjour and Stock [23] estimated the cost-effectiveness of a DMP in hypertension for time horizons of 35, 45, and 55 years. The results differed only marginally for low-risk patients with a starting age between 40 and 49 years at $14,864 per QALY (maximum time horizon of 55 years), a starting age of 50 to 59 years at $13,132 per QALY (maximum time horizon of 45 years), and a starting age of 60 to 69 years at $14,239 per QALY (maximum time horizon of 35 years). The variance for high-risk patients was even smaller and was within a range of $180 for the different age groups and time horizons. The time horizon is mostly lifetime and varies more between the indication areas than within an indication area, and its overall effect on the results should be small.

Discount rates

Even the small spread of 2% (or 3–5%) in discount rates between the models can have a large effect. High discount rates for future gains in health-related quality of life and costs negatively affect cost-effectiveness; for DMPs, future savings in cost of illness are expected. In contrast, the intervention costs are generally small compared with the potential savings of the intervention.

Utility weights

It is hard to find adequate utility measures to assemble models, and models that are more complex face higher risks in incorporating data, which may not fit. Some methodological problems were detected in the studies, and we illustrate them with examples for clarity. First, outdated data were used. For example,
<table>
<thead>
<tr>
<th>Author</th>
<th>Cycle length</th>
<th>Discount rate costs (%)</th>
<th>Discount rate effects (%)</th>
<th>Time horizon</th>
<th>Price year</th>
<th>Results in US $ (prices of 2011)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moertl et al. [29]</td>
<td>1 mo</td>
<td>5</td>
<td>5</td>
<td>20 y</td>
<td>2010</td>
<td>Austria/Canada: UC: $44,941/$47,909 and 2.36 QALYs; MC: $48,106/$51,039 and 3.04 QALYs; BMC: $43,752/$46,430 and 3.20 QALYs</td>
</tr>
<tr>
<td>Chan et al. [18]</td>
<td>1 mo</td>
<td>3</td>
<td>3</td>
<td>15 y</td>
<td>2005</td>
<td>Baseline-risk patients improved life expectancy by 0.42 life-years for an additional cost of about $4,607</td>
</tr>
<tr>
<td>Göhler et al. [21]</td>
<td>1 mo</td>
<td>5</td>
<td>5</td>
<td>Lifetime (until all patients who had not already died reached an age of 120 y)</td>
<td>2007</td>
<td>$12,145 per QALY</td>
</tr>
<tr>
<td>Gandjour [32]</td>
<td>Not in text</td>
<td>3</td>
<td>3</td>
<td>55 y (maximum)</td>
<td>2004</td>
<td>$19,379 per QALY</td>
</tr>
<tr>
<td>Gandjour and Stock [23]</td>
<td>1 y</td>
<td>3</td>
<td>3</td>
<td>35, 45, and 55 y</td>
<td>2004</td>
<td>Low-risk patients: 40–49 y $14,864 per QALY; 50–59 y $1,153 per QALY; 60–69 y $1,091 per QALY</td>
</tr>
<tr>
<td>Miller et al. [20]</td>
<td>6 mo</td>
<td>3</td>
<td>3</td>
<td>Lifetime (not described further)</td>
<td>2003</td>
<td>$53,373 per QALY</td>
</tr>
<tr>
<td>Gillespie et al. [17]</td>
<td>1 y</td>
<td>3.5</td>
<td>3.5</td>
<td>Lifetime (not described further)</td>
<td>2006</td>
<td>Cost savings of $657 and health gain of 0.0051 QALYs</td>
</tr>
<tr>
<td>Gillespie et al. [16]</td>
<td>1 y</td>
<td>3.5</td>
<td>3.5</td>
<td>50 y (maximum)</td>
<td>2006</td>
<td>Cost savings of $657 and increase of 0.0051 QALYs</td>
</tr>
<tr>
<td>Ito et al. [19]</td>
<td>3 mo</td>
<td>3</td>
<td>3</td>
<td>Lifetime (not described further)</td>
<td>2010</td>
<td>Mailed education plus DMP: $76,987; DMP: $71,414; DMP plus polypill: $137,256; polypill plus mailed education: $116,616; polypill plus DMP: $146,544</td>
</tr>
<tr>
<td>Steuten et al. [22]</td>
<td>2 wk</td>
<td>4</td>
<td>4</td>
<td>5 y</td>
<td>2004</td>
<td>With annual review: $3,635 per QALY</td>
</tr>
<tr>
<td>Gordois et al. [25]</td>
<td>6 mo</td>
<td>5</td>
<td>5</td>
<td>5 y</td>
<td>2006</td>
<td>Without annual review: $2,194 per QALY</td>
</tr>
<tr>
<td>Yu et al. [26]</td>
<td>1 y</td>
<td>3</td>
<td>3</td>
<td>10 y</td>
<td>2011</td>
<td>Cost savings of $8,788 and 0.498 QALYs gained</td>
</tr>
<tr>
<td>Huang et al. [28]</td>
<td>1 y</td>
<td>3</td>
<td>3</td>
<td>Lifetime (not described further)</td>
<td>2004</td>
<td>$39,786 per QALY</td>
</tr>
<tr>
<td>Gilmer et al. [24]</td>
<td>Not in text</td>
<td>3</td>
<td>3</td>
<td>Lifetime (40 y)</td>
<td>2003</td>
<td>Uninsured cohort: $12,400 per QALY; CMS cohort: $30,060 per QALY; Medi-Cal cohort: $34,952</td>
</tr>
<tr>
<td>Mason et al. [30]</td>
<td>1 y</td>
<td>5</td>
<td>5</td>
<td>Lifetime (not described further)</td>
<td>2003</td>
<td>Commercial cohort: $85,087; Blood pressure control: $4,971 per QALY; Lipid control: $24,672 per QALY</td>
</tr>
<tr>
<td>Palmer et al. [27]</td>
<td>1 y</td>
<td>3</td>
<td>3</td>
<td>Lifetime (not described further)</td>
<td>1996</td>
<td>All conventional insulin strategies with an add-on are dominant (cost saving); Intensive insulin therapy with add-ons causes additional costs per life-years gained from $19,945 to $21,701</td>
</tr>
</tbody>
</table>

BMC, intensive patient management; DMP, disease management program; MC, multidisciplinary care; UC, usual care; QALY, quality-adjusted life-year.
two studies [23,32] of chronic heart disease used data for stroke that were originally collected in 1976 [41]. Second, data were based on a small sample size. For instance, interviews with 74 patients with atrial fibrillation [42] were used by the same two studies [23,32]. Third, data extraction from original studies was often opaque. One study [30] of diabetes used a mean utility score of 0.5 for poststroke, although the value in the study the work referenced [43] estimated a mean utility score of 0.45 for major stroke. Fourth, data were incorporated into the models estimated with different instruments and different methods. For example, one study [28] of diabetes derived the utility values from several sources, and in addition to the different instruments (e.g., the
EuroQol five-dimensional questionnaire or Karnofsky performance status) used to obtain the utilities, different methods (standard gamble [SG] and time trade-off [TTO]) were used to evaluate the different health states. The mean quality weights for SG, TTO, and the rating scale differed considerably: the utility was the highest measured with SG, followed by TTO and visual analogue scales [44,45]. Fifth, there were inconsistencies in the transfer of utility values. In one study [27] of chronic heart disease, the transformation for the values in the baseline group was based on the short form 12 health survey tariff [46] instead of the questionnaire used in the clinical trial (six-dimensional health state short form) [47]. These findings suggest that the selection of data was poor.

Cost of illness
The highest absolute differences in costs used by the models were found in hospitalization and medication costs and in incorporated costs for comorbidities caused by the initial disease. These categories may also be the main cost drivers in the models. Otherwise, it was not possible to standardize the costs used; therefore, an adequate interpretation cannot be made. Standardized reporting would lead to greater transparency and make the useful estimation of input parameters possible.

Sensitivity of parameters
The results of the 13 univariate sensitivity analyses indicated that a higher potential bias should decrease the differences between the models in the effectiveness of the intervention/transition probabilities, time horizons, utility values, discount rates, simulated population, costs of the DMP, and duration of the treatment effect.

Funding
It was difficult to assess whether funders tried to influence study results or at least tried to prevent the publication of unfavorable results. Nevertheless, a closer look at the four industry-funded publications [24,27,29,30] and the four studies without a conflict of interest statement [18,20,22,23] indicates that the industry-funded studies were more likely to report worse cost-effectiveness ratios. It seems that the industry's interest is the greatest in diabetes, where 60% of the models were industry-funded compared with 11% for chronic heart disease and 0% for asthma. Overall, the interest of the industry in DMPs may not be as pronounced as in pure pharmaco-economic evaluations.

Limitations
The search was limited to the databases PubMed, MEDLINE, EconLit, and Embase. Only studies in English and German were considered. Data extraction and determining the intensity rating of the DMP and usual care were conducted by only one reviewer; therefore, errors cannot be completely ruled out. The classification of the DMPs and usual care as high, medium, and low intensity was pragmatic and did have a sound theoretical basis. The quality assessment with two different quality appraisals was conducted by two researchers; thus, single items may have been interpreted differently from the initial meaning according to the similar background of the researchers. To make the study results comparable, the gross domestic product index of the Organisation for Economic Co-operation and Development was used, and all prices were inflated to the year 2011, although this might not reflect the price increase in the health care sector adequately. The results in this review are based on published full-text articles, online supplements, and appendices without additional communication with authors or requested technical modeling reports. Furthermore, the Markov models were not available to gain deeper insight. Nonetheless, we believe that our work provides a valuable review of recent economic models for DMPs that can be used as reference for future research.

Comparison with Literature
Only two studies [10,48] reviewed the economic evaluations of DMPs with decision-analytic modeling studies, and both were conducted for chronic heart disease. Maru et al. [10] reviewed 10 models, 8 of which were Markov models. Four studies reported an incremental cost-effectiveness ratio of more than $50,000, six less than $50,000, and six were dominant. For study quality assessment, two more generic tools were used, namely, the grading system for the quality of cost-effectiveness studies [49] and the Consolidated Health Economic Evaluation Reporting Standards checklist [50]. Six [16,17,20,26,29,30] of these models met the inclusion criteria for this review and were included. Göhler et al. [48] reviewed the modeling approach of 34 models estimating the cost-effectiveness of health technologies for CHF, but neglected the cost-effectiveness itself. For quality assessment, they used their own framework to assess the studies systematically and synthesize the methodological approaches. Three [18,20,21] of these models evaluated disease management interventions and were included in this review. The number of publications, however, is still low; no models for DMPs in chronic obstructive pulmonary disease or breast cancer were found. We hope that modelers will contribute to fill these gaps in research soon.

Conclusions
Fourteen studies reported QALYs and two studies reported LYG as outcome measures. For QALYs, 11 out of 14 studies reported at least one result for a DMP less than $50,000 per QALY (seven [16,17,21,23,29,32] [18] out of nine studies of chronic heart disease, two [22,25] studies of asthma, and all three [24,28,30] of diabetes). For LYG, both studies reported costs less than $30,000 per LYG [26,27] both of diabetes. According to the quality assessment, however, weaknesses in the models were found, namely, reporting practice and selection of input data. Although journals have space and word limits, the reporting practice could be standardized easily by using technical and online appendices [10]. In addition, a systematic review of the literature should be conducted, at least for the most important parameters.

The choice of the quality appraisal used to assess the quality of studies and models in a review may influence its results, shown here by the low value of the Spearman rank correlation coefficient, indicating no correlation between the two instruments.

Because most of the models merely extrapolate the results of a clinical trial, the models are relatively static and the incremental effects of individual components of the DMPs cannot be evaluated economically.

In all indication areas, the disease and its progression are represented well by the model structures. Although there are variations between the models within an indication, these are small and unlikely to influence the results substantially.

For the incorporated data in the models, the picture is different. Utility weights, time horizons, costs, and the control group treatment program or its description vary considerably and may affect the results in either direction.

If these problems are addressed, then Markov models should be more suitable to evaluate economic effects of multicomponent interventions, and provide helpful information for decision makers.
Acknowledgment

I thank S. Gilowsky, who independently undertook the literature review and the quality assessment for validation.

Supplemental Materials

Supplemental material accompanying this article can be found in the online version as a hyperlink at http://dx.doi.org/10.1016/j.jval.2016.07.004 or, if a hard copy of article, at www.valueinhealthjournal.com/issues (select volume, issue, and article).

REFERENCES


