International Journal of Technology Assessment in Health Care, 29:1 (2013), 61–70.

doi:10.1017/S0266462312000700

APPROACHES TO CHRONIC DISEASE MANAGEMENT EVALUATION IN USE IN EUROPE: A REVIEW OF CURRENT METHODS AND PERFORMANCE MEASURES

Annalijn Conklin
MRC Epidemiology Unit, University of Cambridge

Hubertus Vrijhoef
Integrated Care, Medical Center Tilburg University

Ellen Nolte
Health and Healthcare, RAND Europe Cambridge Ltd

Objectives: An overview was produced of approaches currently used to evaluate chronic disease management in selected European countries. The study aims to describe the methods and metrics used in Europe as a first to help advance the methodological basis for their assessment.
Methods: A common template for collection of evaluation methods and performance measures was sent to key informants in twelve European countries; responses were summarized in tables based on template evaluation categories. Extracted data were descriptively analyzed.
Results: Approaches to the evaluation of chronic disease management vary widely in objectives, designs, metrics, observation period, and data collection methods. Half of the reported studies used noncontrolled designs. The majority measure clinical process measures, patient behavior and satisfaction, cost and utilization; several also used a range of structural indicators. Effects are usually observed over 1 or 3 years on patient populations with a single, commonly prevalent, chronic disease.
Conclusions: There is wide variation within and between European countries on approaches to evaluating chronic disease management in their objectives, designs, indicators, target audiences, and actors involved. This study is the first extensive, international overview of the area reported in the literature.

Keywords: Chronic disease management, Long-term care, Program evaluation, Europe, Methods

Chronic diseases place a substantial burden on individuals, their carers and society. They account for a large share of health care costs, yet care remains suboptimal (1). Structured disease management is proposed as a means to improve quality and reduce cost of health care, and to improve health outcomes. Despite intuitive appeal and growing numbers of studies, the evidence such approaches achieve these ends remains uncertain (2;3).

This work was funded under the European Commission FP7 program (grant 223277). The authors are grateful to the project partners and additional experts who contributed to data collection/verification: Matthias Brunn (Université Paris XII - Val de Marne, UPVM, Paris, France), Karine Chevreul (UPVM, Paris, France), Isabelle Durand-Zaleski (UPVM, Paris, France), Arianne Elissen (Universiteit Maastricht, UM, Maastricht, The Netherlands), Antje Erler (Johann Wolfgang Goethe Universität, GUF, Frankfurt, Germany), Fadila Farsi (Centre Anticancereux Leon Berard, CLB, Lyon, France), Béatrice Fervers (CLB, Lyon, France), Maria Flamm (Paracelsus Medizinische Privatuniversitat Salzburg, PMU, Salzburg, Austria), Anne Frølich (Københavns Universitet, UCPH, Copenhagen, Denmark), Birgit Fullerton (GUF, Frankfurt, Germany), Péter Gaál (Semmelweis University, Budapest Hungary), Ramune Jacobson (UCPH, Copenhagen, Denmark), Robert Krohn (AQUA - Institut für angewandte Qualitätsförderung und Forschung im Gesundheitswesen GmbH, AQUA, Göttingen, Germany), Hélène Labrosse (CLB, Lyon, France), Taavi Lai (University of Tartu, Tartu, Estonia), Isabelle Peytremann Bridevaux (Institut Universitaire de Médecine Sociale et Préventive, Lausanne, Switzerland), Boris Pöhlmann (AQUA, Göttingen, Germany), Walter Ricciardi (Università Cattolica Sacro Cuore, Rome, Italy), Burgi Riens (AQUA, Göttingen, Germany), Martin Roland (University of Cambridge, Cambridge, UK), Antonio Sarria Santamera (Instituto de Salud Carlos III, ISCIII, Madrid, Spain), Zuleika Saz-Parkinson (ISCIII, Madrid, Spain), and Andreas Sönnichsen (PMU, Salzburg, Austria).

What we know about the impact of approaches to manage chronic disease tends to come from small studies frequently focused on high risk patients, and often undertaken in academic settings (4). Much less is known about large-scale programs or about transfer of small-scale interventions from an original site to other locations (2;3). One important reason for limited evidence is lack of universally accepted methods to evaluate a given intervention that are both scientifically sound and operationally feasible. Such evaluation methods are, however, a precondition to select efficient and effective interventions that can address the growing burden of chronic conditions.

Countries in Europe are implementing a range of approaches, many in the form of disease management, to better meet the needs of chronically ill patients, although the nature and scope of related approaches differ (5). As all countries are facing the challenge of chronic disease, it is critical to facilitate sharing of experiences to enable cross-country learning (1). Yet, as approaches to chronic disease management vary, attempts to evaluate them are equally divergent, both within and between countries (1;5). It is, therefore, necessary to better understand the range of approaches taken to assess effects of chronic disease management so as to facilitate lesson-learning and method standard-setting.

This study aims to contribute to filling this gap by reviewing approaches to evaluate models of chronic disease management

being/are implemented in twelve European countries. Our work is part of a larger study on developing and validating disease management evaluation methods for European healthcare systems (DISMEVAL). Our key objective is to describe whether and how such approaches are being evaluated in Europe as a first step to help advance the methodological basis for their assessment.

METHODS

Data Collection

We developed a common template to collect data on the health system and policy context; type and format of approaches to managing chronic disease; and approaches to evaluation. The latter covered evaluation aim(s), audience, actors involved, budget, frequency, design, length, data source, indicators of program effect, and feedback mechanisms in place. The template used simple checkboxes and open-ended questions, and included a glossary of term definitions where appropriate and relevant (Supplementary Table 1, can be viewed online at www.journals.cambridge.org/thc2013078).

Our study was restricted to approaches aimed at managing people with established chronic disease and considered a broad range of possible models such as care pathway, case management, chronic care model, coordinated care, multidisciplinary team(s), nurse-led clinic, and/or provider network, among others. As many European countries are experimenting with/ implementing diverse approaches, country selection was necessarily pragmatic. It was guided by three main criteria to capture: (i) the range of approaches to funding and governing health care across Europe; (ii) the range of stages in economic development; and (iii) geographical spread. We thus identified thirteen countries in Europe, of which twelve were included in this review: Austria, Denmark, England, Estonia, France, Germany, Hungary, Italy, Latvia, Lithuania, The Netherlands, and French-Switzerland (Lausanne).

Key informants in the selected countries collected data using our common template and were asked to adopt an evidence-based approach by making use of the best data available (i.e., research projects, policy documents, routine statistics, stakeholder interviews, etc.). Given the variety of approaches in each country, informants were further asked to present a "sample" considered representative of a given health system in terms of the type and setting of delivery model, providers involved, key strategies used, and the population covered. Key informants were identified primarily as partners in the DISMEVAL project but also through the authors' existing professional networks. They had to demonstrate expertise in noncommunicable disease(s) and/or the country's health system context as shown by relevant academic publications and/or advisory roles in relevant governmental bodies.

Principal data collection occurred from June 2009 to December 2009. Subsequent follow-up with each country's key informant(s) to ensure accuracy and to update information where necessary and appropriate occurred in November and December 2010 and again in January and February 2011. To the extent possible, key informant data were supplemented by separate searches of relevant databases and Web sites for countries which the authors had sufficient language skills. A limited amount of additional information was retrieved through such searches which we re-validated with country experts.

Data Extraction and Analysis

Data were extracted along two dimensions: (i) the chronic disease management approach by country and (ii) evaluation categories given in the expert-completed template. Data on performance measures were extracted as reported, following given response categories: cost; utilization; structure (registry, reminder, other); and process (referral rates, monitoring, clinical, knowledge, other). However, open responses to "other" effect indicators were clustered (e.g., self-management, quality of life, health status), when possible, to ensure consistency and enable comparison. We assumed no control group or external benchmark was used in a documented evaluation when respondents did not specify whether or what kind of comparator was used for a design category selected.

Data synthesis used content analysis to categorize and determine numbers and frequencies, and narrative summary to identify commonalities and differences between evaluation approaches (6). Data thus extracted were twice verified with key informants for accuracy and completeness.

RESULTS

Most chronic disease management approaches reviewed here underwent some form of evaluation, or plans existed to do so; a single approach often had multiple evaluations. A couple notable exceptions were Germany where one approach had ten documented evaluations; and Hungary where half the approaches reviewed had none planned. In Latvia and Lithuania, we found no documented evidence of evaluations as experts identified an absence of chronic disease management approaches for study.

Reported evaluations varied considerably in nature and scope, with differences in objectives, designs, performance metrics, length, and data sources. The range of approaches to evaluation was characterized by a mix of controlled and noncontrolled studies measuring predominantly clinical process measures, patient behavior and satisfaction, cost, and utilization. Effects were usually observed over 1 or 3 years on a population of patients with a single, commonly prevalent, chronic condition (Table 1, and Supplementary Table 2, which can be viewed online at www.journals.cambridge.org/thc2013078, for more detail).

Interventions and Populations

The majority of evaluations were of approaches targeting diabetes and/or cardiovascular disease(s) which are prevalent

https://doi.org/10.1017/S0266462312000700 Published online by Cambridge University Press

Table 1. Key Features of Evaluations of Chronic Disease Management Approaches Surveyed in Europe

	Number of European disease management approaches with documented evaluations										
	Austria	Denmark	England	Estonia	France	Germany	Hungary	Italy	Netherlands	Switzerland (Lausanne)	*TOTAL*
Disease management approaches reviewed	5	5	4	6	8	6	6	4	4	3	51
Target group for approaches reviewed:											
Single chronic condition – CVD, DM2, cancer, etc.	4	5	2	6	7	1	4	4	3	2	38
Generalist	1	–	1	–	–	3	2	–	–	1	8
Multi-morbidity	–	–	1	–	–	2	–	–	–	–	3
Other (e.g., falls, substance misuse, dyslipidemia, etc.)	–	1	2	–	1	–	–	1	1	–	6
Status of evaluations:											
Completed / in progress	5	2	7	9	10	19[a]	4	3	5	3	67
Continuous	–	–	–	–	3	4[a]	–	1	–	–	8
Planned	2	3	–	–	2	1	–	–	–	–	8
None[b]	1	–	–	–	–	–	3	–	–	–	4
Length of observation:											
12 months or less	3	4	1	7	6	8	2	2	1	1	35
Between 12 and 36 months	–	1	5	–	8	8	–	1	4	1	28
More than 36 months	–	–	–	–	–	5	1	–	–	–	6
Length not reported or information not available	4	–	1	2	1	3	1	1	–	1	14
Types of actors involved in evaluations											
Internal	5	–	1	5	5	5	2	1	1	1	26
External	1	2	6	–	8	19	2	3	4	2	47
Both	–	3	–	4	–	–	–	–	–	–	7
Not reported or information not available	1	–	–	–	2	–	–	–	–	–	3
Evaluations aimed[c] to:											
monitor implementation/ quality/ participation	1	5	2	–	2	1	–	2	2	1	16
assess efficacy/ (cost-)effectiveness/ impact	3	4	6	4	1	14	2	3	4	2	43
identify feasibility of improvement/ scale-up	4	–	–	1	2	2	–	1	–	–	10
determine return on investment/ value for money	–	–	–	–	6	2	2	1	1	1	13
examine processes/ patient experiences / acceptability	–	–	2	8	4	6	–	2	3	–	25
explore methodological influences on effects	–	–	–	–	–	2	–	1	–	1	4
Evaluation designs:											
Experimental or longitudinal with internal controls	1	–	1	–	–	9	–	1	1	–	13
Cross-sectional (pre/post or post-only) / observational	4	1	2	3	10	14	3	3	4	3	47
Mixed qualitative and observational methods	1	1	3	5	1	–	1	–	–	–	12
Qualitative	–	–	2	1	1	–	–	1	–	–	5
Return on investment	–	–	–	–	–	1	–	–	–	–	1
Information not available	1	3	–	–	4	–	–	–	–	–	8
Metrics used to measure program effects:											
Structure	1	–	3	7	8[d]	2	1	3	1	–	26
Clinical process	3	5[d]	–	6	5[d]	14	4	4[d]	5	2	48
Organizational process	3	3[d]	3	6	8[d]	3	–	3[d]	5	2	36
Intermediate clinical outcomes	1	2	2	–	4[d]	6	–	1	1	–	17
Utilization	3	2	5	6	–	11	4	2	4	3	40
Cost	–	1	3	7	3[d]	11	3	1	4	3	36
Longer-term outcome (health status)	5	4	2	1	–	3	2	1	1	2	21
Other	1	2	6	7	8[d]	17	1	1	5	3	51

https://doi.org/10.1017/S0266462312000700 Published online by Cambridge University Press

Table 1. Continued.

	Number of European disease management approaches with documented evaluations										
	Austria	Denmark	England	Estonia	France	Germany	Hungary	Italy	Netherlands	Switzerland (Lausanne)	*TOTAL*
Data sources:											
Newly collected	3	2	6	9	6	18	2	4	5	2	57
Routine	2	4	4	7	10	10	4	2	3	3	49
Not reported or information not available	2	1	–	–	4	1	–	–	–	–	8

[a]One approach in Germany (Disease Management Programme, DMP) has nine completed and one continuous evaluations.
[b]For approaches with no planned evaluation, questions of evaluation aim, methods, and metrics were not applicable.
[c]Evaluations are counted for each separate aim.
[d]Measures relate to established targets.

conditions with well documented clinical guidelines (7). Other targeted conditions included cancer, respiratory disease, stroke, frailty, and dementia. Studies of approaches targeting less common conditions (e.g., Multiple Sclerosis, Parkinson's disease, and Schizophrenia) were found in one country reviewed. More generalist models of chronic disease care were seldom evaluated.

Evaluation Aims and Design

Evaluations most commonly aimed to assess performance (and/or process) of a given approach in terms of established targets, quality, effectiveness (clinical and/or cost), satisfaction, or adherence. Precise and measurable information about established targets against which measured effects might be compared (e.g., 40 percent participation by new breast cancer patients and inclusion of a minimum of 150 patients per year), was rarely specified.

The design of evaluations focused on outcomes varied widely within and between countries. We identified randomized controlled trials; controlled longitudinal evaluations (prospective and retrospective); controlled pre-post or post-only studies; noncontrolled pre-post or post-only assessments with benchmark comparisons to an external reference; cross-sectional studies with or without a control group; and noncontrolled noncomparative pre-post or post-only assessments (Table 2). Many program evaluation designs were described as being qualitative, and, therefore, would not involve controls; while several observational or post-only designs combined qualitative methods with a satisfaction survey (without counterfactual comparison) to assess outcomes.

We found fewer evaluations of only the process of a given approach, but a similar range of designs: qualitative, cross-sectional, post-only, longitudinal (with and without control), and mixed-methods case study. Some process evaluations were reported to have confidential protocols. We also found several formative evaluations aiming to study feasibility of a given approach for pilot implementation and/or scale-up.

Approaches were also assessed to monitor implementation or report activity, or occasionally to identify areas for improvement of service, staff or quality. Again, diverse designs were used to audit performance: qualitative, noncontrolled cross-sectional and controlled longitudinal approaches. Some studies combined the aim of monitoring performance with aims of a process or a program evaluation (e.g., integrated clinical pathway for heart diseases, Denmark; or IGEA project, Italy).

Very few approaches were evaluated solely for economic impact, both in Germany (Table 1). However, economic effects were commonly examined among outcomes assessed in program evaluations, for example "added value for public money" (France: REVESDIAB, DIABAIX, and COPA); cost reduction (GP contracts, Germany); "financial performance and savings distribution" (Care Coordination Pilot, Hungary); cost-utility (Raffaello project, Italy); cost-effectiveness (England: Expert Patient Program, Partnerships for Older People Project; Switzerland: Diabaide; The Netherlands: Matador and integrated stroke services programs); or cost consequences (Integrated Care Pilots, England).

Over half of documented evaluation designs did not use a control group for counterfactual comparison (Table 2). When the design included a comparison strategy, different approaches were taken to estimate effects. Some used either concurrent reference populations (e.g., case-mix adjusted hospital patients, Stroke Service Delft; diabetics in the region of Therapie Aktiv) or the general population (e.g., Partnerships for Older People Project). Alternatively, several evaluations created statistical controls by either stratifying (using age, sex, and insurance status) or matching (using propensity score, risk, or morbidity, age, and sex). Other comparison strategies involved using a "benchmark" as an external reference group which also varied widely (e.g., international best practice or the literature; performance

https://doi.org/10.1017/S0266462312000700 Published online by Cambridge University Press

Table 2. Study Designs and Comparison Strategies Used in Documented Evaluations of Chronic Disease Management Approaches in Europe

Comparison strategy	Study designs[a]							
	RCT	Observational	Longitudinal cohort	Pre-post (before-and-after)	Post-only	ROI	Qualitative	Other[b]
Control group (internal comparator)	Therapie Aktiv (A); Disease management programs (D); Expert Patients Programme (ENG); Raffaello project (I)	Prosper Net integrated care contracts (D)	Disease management programs (D); GP contracts (D); Gesundes Kinzigtal integrated care contracts (D); National care standard for vascular risk[c] (NL)	Therapie Aktiv, (completed[d] and planned) (A); Evercare program (ENG); From on-demand to proactive primary care[b] (I); Multifunctional community centers (H); Delta physician network (CH)	ASALEE (F); Care Coordination Pilot (H); Stroke Service Delft (NL)			
Benchmark (external comparator)				SIKS project[b] (DK); Capital region disease management program (DK); Quality management in primary care - DM2, CVD (EE); CDM - MS (EE); DIABAIX[d] (F)	REVESDIAB (F); COPA (F); multidisciplinary team RCP (F); disease management programs (D); IGEA project (I)			Integrated Care Pilots[e] (ENG); Partnerships for Older People Project[b] (ENG)
None		Kardiomobil (A); Prosper Net integrated care contracts (D); Community nurses - rural areas (D)	Disease management programs (D); Gesundes Kinzigtal integrated care contracts (D); Primary care chain - DM2 (NL)	Treatment (& financing) protocols (H); Leonardo pilot (I); Diabaide[e] (CH); Matador[e] program (NL); Primary care chain for DM2 (NL)	Interface management Styria (A); Breast cancer clinical pathway (CH); Expert Patients Programme (ENG); REVESDIAB (F); DIABAIX (F); COPA (F); multidisciplinary team RCP (F); disease management programs (D); GP contracts (D), medical care centers (D)	Gesundes Kinzigtal integrated care contracts (D)	CDM - MS (EE); Expert Patients Programme (ENG); organization of access to supportive care (F); IGEA project (I); multidisciplinary team RCP (F)	Integrated care - stroke (A); Care Coordination Pilot (H); CDM - COPD, PD, Schizophrenia and Quality management in primary care – CVD, DM2 (EE)

[a]Evaluations of some approaches had no information on their design (planned or completed), or the protocol was explicitly confidential.
[b]Design included mixed methods.
[c]Design included cost/benefit as a key measure but design was not an economic evaluation/ return on investment (ROI).
[d]Comparator used in post-only comparison.
[e]Study included cost modeling or cost-effectiveness analysis.
A, Austria; CDM, chronic disease management at interface of primary and secondary care; CH, Switzerland; COPD, chronic obstructive pulmonary disease; CVD, cardiovascular disease; D, Germany; DK, Denmark; DM2, diabetes mellitus type 2; EE, Estonia; ENG, England; F, France; H, Hungary; I, Italy; MS, multiple sclerosis; NL, The Netherlands; PD, Parkinson's disease; RCT, randomized controlled trial; ROI, return on investment

targets or timelines; other regions, sites or practices in a country; and a regional standard or mean value) (Supplementary Table 2).

Notably, we identified only three countries where evaluations included an aim to explore methodological influences on findings (Table 1). One feasibility study explicitly sought to identify and refine evaluation techniques for future use (Leonardo pilot project, Italy); while three other evaluations included assessing the impact of selection effects on findings of effectiveness (Germany: Disease Management Program; Switzerland: Delta physician network).

Metrics, Length of Observation, and Data Sources

We found large differences within and between countries in the indicators of program effect and the length of observation used to evaluate approaches reviewed. Many indicators described related to established targets, particularly for approaches in France, Italy and Denmark. There appeared to be diverse interpretations of a given construct for indicator classification as we found highly varied metrics reported to evaluate structure, process and program effects, not only for the open category "other" but also for structured responses. Moreover, several evaluations using a variety of metrics included some with no clear link to the study's aims or design.

Program evaluations commonly combined multiple outcome metrics, including clinical effectiveness, hospitalization/utilization rate, survival, mortality, cost, cost-effectiveness/cost-benefit, quality-adjusted life-years, disease incidence, and standardized targets. Many included intermediate clinical outcomes: namely, hemoglobin or cholesterol levels, or medical parameters for disease control (e.g., body mass index, blood pressure, and disease-specific prescriptions). Clinical process measures and/or patient satisfaction also tended to be used in the mix (Table 3 and, for more detail, Supplementary Table 3, which can be viewed online at www.journals.cambridge.org/thc2013078).

Most evaluations included one or more process measures related to clinical aspects of disease management (e.g., referral rates, laboratory tests performed, prescription rates, adherence to clinical standards, etc.) and sometimes also organizational process measures of service monitoring (e.g., number, type and waiting times for consultations). Very few evaluations measured patient or provider knowledge (Therapie Aktiv, interface management Styria, integrated care contracts and SIKS project).

Although several process and outcome indicators appeared to be commonly used to evaluate disease management approaches within and between countries, specific descriptions showed great variation in the operational definitions of common metrics. For example, "referral rates" were widely measured but specified differently: namely, frequency of recommending (the approach) to patients; number of new patients in the program (i.e., recruitment); referral to ocular fundus examination; first set of tests for 70 percent of all patients; frequency or per-member number of physician visits; etc. Similarly, various metrics were used to evaluate a program's "monitoring" processes including: consultation rate; percentage of members with check-ups; waiting times; number of patients receiving at least one action to prevent complications, or whose case was presented at least every 6 months in a multi-professional coordination meeting. Finally, although cost effects were rarely specified, descriptions also revealed diverse indicator definitions (e.g., total expenditures, operating costs, average cost per patient, revenues, per-member sick days and prescriptions costs), with the occasional reporting of resource use as a process measure (e.g., in Denmark) rather than an indicator of program effect.

When described, "other" structural metrics showed the greatest variation. Examples included: square-meters of building; equipment purchased; scope, local accessibility, or number and timelines of services provided; hotline established; annual reassessment procedures; training; participation and integration of "involved actors" (including nonmedical personnel adhering to the approach); proportion of patients assisted by a case manager; caseload; extent of IT penetration ("informatization"); proportion of indicators attributed a value for specified time frames; performance improvement (using Barthel Index and Rankin scale); the care structure; and funding received per participating patient overall and by source (data not shown). Structural measures such as registries (IGEA project and From On-Demand to Proactive Primary Care, Italy) and reminders (Leonardo project, Italy; breast cancer clinical pathway, French-Switzerland) were rarely studied.

Evaluation time frames were equally variable but tended to involve 12 and 36 months of observation. Health and/or economic impacts were commonly measured after only twelve months of observation, although there were also examples of shorter time frames of 2 to 3 months (e.g., SIKS project and integrated clinical pathway for heart diseases, Denmark; CCP, Hungary). We found examples of longer-term evaluations of clinical and financial outcomes spanning 36 months in England, France, Germany, and The Netherlands and one 60-month evaluation to assess the sustainability of multi-functional community centers in Hungary.

The majority of documented evaluations used both new data and routine sources such as medical records, laboratory tests, and provider registries. New data were predominantly collected by surveys, but other sources involved interviews, focus groups, site visits or direct observation, alongside intervention-specific data sets and literature or document review (Supplementary Table 1).

It is noteworthy to find that evaluation measures were sometimes not commensurate with the aim or design. In France, for example, six evaluations of three distinct approaches aimed to "assess value for money" but only one measured cost explicitly (some examined "budget" but only as a measure of organizational activity). We found a generalist approach in Hungary

Table 3. Indicators of Effect Used in Documented Evaluations of European Disease Management Approaches

Indicators of program effect surveyed	Number of European disease management approaches with documented evaluations										
	Austria	Denmark	England	Estonia	France	Germany	Hungary	Italy	Netherlands	Switzerland (Lausanne)	*TOTAL*
Cost (specific examples varied)	–	1	3	6	3	4	2	1	4	2	26
Utilization (length of stay; relative and/or absolute number of hospital admissions)	1	–	4	6	2	3	3	2	4	2	27
Open responses (e.g., Health Status; QALY; mortality (inpatient); survival (outside hospital))	2	–	1	1	–	–	2	2	1	1	10
Structural measures:											
Registry	–	–	–	–	–	–	–	2	–	1	3
Reminder	–	–	–	–	–	–	–	1	–	–	1
Other (specific examples varied)	1	–	3	5	2	–	1	1	2	–	15
Process measures:											
Referral rates (specific examples varied)	1	1	–	6	2	3	2	2	2	2	21
Monitoring (specific examples varied)	1	–	–	6	2	1	–	3	1	2	16
Clinical (specific examples varied)	2	1	–	6	2	2	2	4	4	2	25
Knowledge (e.g., patient disease knowledge; provider awareness of intervention)	2	1	1	–	–	1	–	–	1	–	6
Other											
Self-management (e.g., smoking cessation, healthier diet, increased physical activity, etc.)	–	–	1	–	–	1	–	1	2	–	5
Satisfaction	1	1	1	5	2	3	–	2	3	2	20
Quality of life	1	1	2	–	–	1	1	–	4	1	11
Varied other indicators (e.g., transport time; reach; adverse events; drop-out rates; productivity loss; provider motivation; estimated patient demand)	2	–	2	–	–	2	1	1	–	–	8

with two documented evaluations aiming to assess financial performance and distribution of savings, yet the length of observation of each was only 2 months and designs were either a post-only with control or a noncontrolled observational with qualitative methods. In another case, the evaluation aim did not warrant excess measures: while the aim was focused on feasibility and implementation rather than efficacy and/or efficiency, new data were collected on a series of clinical process measures, patient self-management and satisfaction, as well as structural and organizational measures (Leonardo pilot project, Italy).

Actors, Audience, and Budget

Different actors carried out evaluations of approaches reviewed. In some countries, external actors conducted evaluations (e.g., integrated clinical pathway for cancer and for heart diseases, Denmark). In others, evaluations involved only internal actors (e.g., chronic disease management approaches in Estonia) or both in collaboration (e.g., Denmark: SIKS project, regional disease management programs; Estonia: quality management in primary care for diabetes and CVD). In England, France, Germany, Italy, and The Netherlands, external evaluations appeared to dominate. By contrast, most approaches reported for Austria were evaluated internally.

Evaluations were predominantly intended for healthcare funders and providers, either as the sole audience or mixed with other stakeholders. Funders included statutory insurance funds (Austria, Germany, and France). Provider audiences were generally reported to be managerial such as health management (Denmark), Medical Director Boards and Chief Executive Officers (Switzerland), and a Regional Hospital Agency (France). Sometimes the intended audience was the national and/or regional government (e.g., France; Germany; Hungary; Netherlands; Tuscany, Italy). Less common audiences included patients or patient organizations (e.g., Matador program, stroke services Delft and primary care chain for diabetes, The Netherlands), the general public (e.g., Care Coordination Pilot, Hungary), and researchers (Raffaello project, Italy).

Data on evaluation budgets were often unavailable. When described, the majority of earmarked funds were reported for external evaluations, which tended to be ad hoc or routine rather than continuous in frequency.

https://doi.org/10.1017/S0266462312000700 Published online by Cambridge University Press

DISCUSSION

This study identified wide variation within and between selected European countries on approaches to evaluating chronic disease management. Differences existed in the aims and objectives, designs, and especially metrics, as well as types of target audiences and actors involved.

Before discussing these findings, we consider some limitations to the analysis presented here. Primarily, data collected represented a cross-section of the best available documented evidence in a selection of countries, as identified by key informants invited based on known expertise. Thus, it is possible that data collection was uneven, despite using a standardized instrument. Unevenness likely reflected differences in availability of relevant data and/or different stages of development of approaches or their evaluations within a given country. Responses from a single or small group of country expert(s) might inadvertently bias some of the data collected in terms of detail and scope of reported information. Concurrently, data collection was limited to a sample of approaches in each country. Thus, the study does not provide an exhaustive inventory of all chronic disease management initiatives in included countries. Moreover, our findings cannot be taken as representative of evaluation of health interventions generally in any country reviewed here because information was limited to evaluation of chronic disease management initiatives.

One main explanation for the observed differences in evaluation approaches is the varied nature of initiatives studied and also disparate emphases on scientific or statutory evaluation and/or diverse objectives among evaluation funders. Similarly, the wide range in performance measures and varied observation periods might be explained by data availability, design chosen, competing interests (scientific vs. statutory studies), and evaluation research capacity, among other factors. Overall, the observed variation in evaluation approaches within and between countries reinforces existing calls to standardize such approaches; particularly more systematic use of a comparison strategy in designs; logical and justified linkage of aims/designs with indicators; and perhaps a shared list of operationalized definitions of common metrics. Our findings are part of a novel project aimed at informing guidance to overcome the many methodological and conceptual challenges to assessing chronic disease management and the substantial confusion in evaluation nomenclature (7–10).

Approximately half of evaluations reviewed here used nonexperimental designs without means for comparison. This lack of control group is problematic as it does not enable assessment of the counterfactual and, therefore, conclusions about whether an intervention's effects could be achieved without it. Specifically, uncontrolled designs mask innate biases such as secular trends or regression to the mean, and may result in over-/underestimation of intervention effects (8;11). However, using a "control" group to assess complex health interventions in this context is increasingly debated (3), not least because whole population initiatives beg the question of whether usual care can be a "fair" comparison. Challenges in disentangling intervention effects from broader contextual factors were found in the evaluation of the Matador diabetes management program because "usual care" was affected by wider changes to patient care being implemented across the Dutch health system (5;12). Similar problems of confounding have been reported elsewhere (13).

In practice, controlled experiments are rarely feasible for routine operations in disease management because randomization is not possible (or desirable) for reasons of cost, ethical considerations, generalizability, and practical difficulties with accurate design implementation (14–16). Recently, evaluation of complex interventions with high contextual influence such as chronic disease management has moved toward "realistic evaluation" involving pluralistic quasi-experimental methods, but our review found the (relatively limited) use of qualitative methods in disease management evaluation requires further research (8). It also remains a question whether, and to what extent, different nonexperimental or pluralistic quasi-experimental designs change an evaluation's conclusions and perceptions of success.

Many researchers have called for standardization of disease management evaluation methods and metrics as this enables comparative evidence on improvements to care quality and cost; supports decision making through the exchange of experiences and weighing-up available evidence in light of methodological and practical constraints; and also allows better communication and translation of research findings using consistent nomenclature and terminology (2;9;17;18). Despite considerable guidance on how to systematically evaluate complex interventions to manage chronic disease(s), existing analytic frameworks serve diverse purposes, including: identifying underlying mechanisms of a multi-component program (10) or key aspects of subcomponents (19); guiding indicator choice (20); and developing standardized performance measurements (21).

Although widely accepted evaluation standards remain elusive (22), our findings underscore the importance of standardization as a research agenda. The majority of evaluations, despite many rigorous designs, tended to have a poor "fit" between aims and indicators of effect. For example, evaluations aiming to measure feasibility used only patient satisfaction and no additional indicators for structure, cost, or effort required (e.g., teams of self-employed providers (ASALÉE), France; community nurses, Germany); while others aiming to evaluate staff development did not include training and learning, provider satisfaction, or staff knowledge among performance measures (e.g., quality management in primary care, Estonia). Evaluations of multi-aimed approaches require a more "balanced scorecard" using a range of indicators covering diverse perspectives—financial, service providers, users/patients, innovation, and learning (23). Moreover, as evaluations rely on measurement validity, studies in Europe would benefit from shared

https://doi.org/10.1017/S0266462312000700 Published online by Cambridge University Press

interpretations of constructs of performance measures rather than using a similar indicator (e.g., number and type of consultations) differently to measure aspects of process (i.e., referral rate and monitoring).

The frequently unclear link between indicators of effect, observation period, and evaluation aims and scope raises questions about appropriateness of measured performance. Numerous evaluations included cost measurement, yet were generally 12 months or less. By contrast, only the 60-month evaluation of multi-functional community centers (Hungary) would permit accurate assessment of long-term health benefits on mortality (or cost savings), because it takes at least 3 to 5 years for health management initiatives to reach full implementation and identify "true" program effects (3;11;24). Furthermore, very few evaluations measured patient or provider knowledge, yet both are highly relevant for chronic disease management especially if approaches seek to empower patients. Finally, patient behavior change as an intended program effect was sometimes assessed by "provider adherence to protocol".

Because indicator selection must be informed by a theory of change that makes a logical connection between planned activities and intended outcomes, a logic model is recommended as a necessary component for future evaluations commissioned in this area, with a requirement for sufficient detail on specific and measurable information (10;17).

It is worth noting, moreover, that evaluations need to overcome barriers to producing findings that might then inform policy making. One barrier to successful evaluation of chronic disease management is availability of integrated information systems which many key informants noted were absent in their countries. Another infrastructure-related barrier is lack of dedicated (financial) resources, which we found for many evaluations, particularly when continuous. This might be expected because evaluation and monitoring are often not a standard line item in most organizations' budgets. Further research on how to facilitate disease management evaluation in informing decision making might draw on the wider health sciences research literature on knowledge translation (25).

CONCLUSIONS

This study is the first extensive, international overview describing the methods and metrics used to evaluate chronic disease management in twelve European countries. Approaches to evaluation vary in objectives, designs, indicators, target audiences, and actors involved. Although many evaluations included a comparison strategy, over half did not and thus efforts to standardize approaches should involve as a minimum the use of some form of reference group. The persistent diversity in nomenclature, particularly around "common" evaluation measures, calls for a shared dictionary of terms and operational definitions. Most importantly, so we know we are measuring what matters in health technology assessment of disease management, research, and policy must encourage the application of a transparent theory of change in future—an analytic framework that clearly links indicators of effect directly to the aims and design of an evaluation.

SUPPLEMENTARY MATERIAL

Supplementary Table 1:
www.journals.cambridge.org/thc2013078
Supplementary Table 2:
www.journals.cambridge.org/thc2013078
Supplementary Table 3:
www.journals.cambridge.org/thc2013078

CONTACT INFORMATION

Annalijn Conklin, MSc, MPH, (annalijn.conklin@mrc-epid.cam.ac.uk), Analyst, RAND Europe, Cambridge, UK; University of Cambridge, Cambridge, UK
Ellen Nolte, MPH, PhD, Director of Health and Healthcare, RAND Europe, Cambridge, UK
Hubertus Vrijhoef, PhD, Professor of Integrated Care, Medical Center Tilburg University, Tilburg, The Netherlands

CONFLICTS OF INTEREST

The authors report a grant and travel support to their institution from the European Commission under the 7th Framework Programme, grant agreement 223277.

REFERENCES

1. Nolte E, McKee M. *Caring for People with Chronic Conditions: A Health System Perspective*. Maidenhead, UK: Open University Press/McGraw Hill Education; 2008.
2. Coleman K, Austin B, Brach C, Wagner E. Evidence on the chronic care model in the new millennium. *Health Aff (Millwood)*. 2009;28: 75-85.
3. Mattke S, Seid M, Ma S. Evidence for the effect of disease management: is $1 billion a year a good investment? *Am J Managed Care.* 2007;13: 670-6.
4. Bodenheimer T, Wagner E, Grumbach K. Improving primary care for patients with chronic illness: the chronic care model, Part 2. *JAMA.* 2002;288:909-14.
5. Nolte E, Knai C, McKee M. *Managing Chronic Conditions: Experience in Eight Countries*. Copenhagen, DK: World Health Organization on behalf of the European Observatory on Health Systems and Policies; 2008.
6. Dixon-Woods M, Agarwal S, Jones D, Young B, Sutton A. Synthesising qualitative and quantitative evidence. *J Health Serv Res Policy.* 2005;10:45-53.
7. Coleman K, Mattke S, Perrault P, Wagner E. Untangling practice redesign from disease management: how do we best care for the chronically ill? *Annu Rev Public Health.* 2009;30:385-408.
8. Conklin A, Nolte E. *Disease Management Evaluation: A Comprehensive Review of Current State of the Art*. Santa Monica, CA: RAND Corporation, 2011. Report No.: TR894.
9. Medical Research Council. *Developing and evaluating complex interventions: New guidance [Internet]*. London, UK: Medical Research Council; 2008. [cited 2012 September 19]. Available from: http://www.mrc.ac.uk/Utilities/Documentrecord/index.htm?d = MRC004871.

10. Lemmens KMM, Nieboer AP, van Schayck CP, Asin JD, Huijsman R. A model to evaluate quality and effectiveness of chronic disease management. *Qual Saf Health Care.* 2008;17:447-53.
11. Loveman E, Cave C, Green C, Royle P, Dunn N, Waugh N. The clinical and cost-effectiveness of patient education models for diabetes: a systematic review and economic evaluation. *Health Technol Assess.* 2003;7:1-190.
12. Tsiachristas A, Hipple-Walters B, Lemmens K, Nieboer A, Rutten-van Molken M. Towards integrated care for chronic conditions: Dutch policy developments to overcome the (financial) barriers. *Health Policy.* 2011;101:122-32.
13. Aadelen S. Methodological challenges to prospective study of an innovation: interregional nursing care management of cardiovascular patients. *J Eval Clin Pract.* 1998;4:197-223.
14. Cretin S, Shortell S, Keeler E. An evaluation of collaborative interventions to improve chronic illness care: framework and study design. *Eval Rev.* 2004;28:28-51.
15. Linden A, Adams J, Roberts N. Evaluating chronic disease management program effectiveness: an introduction to the regression discontinuity design. *J Eval Clin Pract.* 2004;12:124-31.
16. Linden A, Adams JL, Roberts N. Evaluating disease management program effectiveness: an introduction to time-series analysis. *Dis Manag.* 2003;6:243-55.
17. Eccles M, Armstrong D, Baker R, et al. An implementation research agenda. *Implement Sci.* 2009;4:18.
18. Donabedian A. *The Definition of Quality and Approaches to its Assessment*. Ann Arbor, MI: Health Administration Press; 1980.
19. Grey M, Knafl K, McCorkle R. A framework for the study of self- and family management of chronic conditions. *Nurs Outlook.* 2006;54:278-86.
20. Steuten L, Vrijhoef H, Severens H, van Merode F, Spreeuwenberg C. Are we measuring what matters in health technology assessment of chronic disease management? Systematic literature review. *Int J Technol Assess Health Care.* 2006;22:47-57.
21. Howe R. Performance measurement for case management: principles and objectives for developing standard measures. *Case Manager.* 2005;16:52-6.
22. Wilson T, MacDowell M, Salber P, Montrose G, Hamm C. Evaluation methods in chronic disease management studies 2004-07. *Dis Manage Health Outcomes.* 2008;16:365-73.
23. Kaplan R, Norton D. The balanced scorecard: measures that drive performance. *Harvard Bus Rev.* 1992:71-9.
24. Serxner S, Baker K, Gold D. Guidelines for analysis of economic return from health management programs. *Am J Health Promo.* 2006;20(Suppl):1-17.
25. Lavis J, Posada F, Haines A, Osei E. Use of research to inform public policymaking. *Lancet.* 2004;364:1615-21.

https://doi.org/10.1017/S0266462312000700 Published online by Cambridge University Press