

Wirksamkeit von Disease-Management-Programmen in Deutschland – Probleme der medizinischen Evaluationsforschung anhand eines Studienprotokolls

Martin Beyer¹, Jochen Gensichen¹, Joachim Szecsenyi², Michel Wensing^{2,3} und Ferdinand M. Gerlach¹

¹ Institut für Allgemeinmedizin, Johann Wolfgang Goethe-Universität

² Abteilung Allgemeinmedizin und Versorgungsforschung, Universität Heidelberg

³ Center for Quality of Care Research (WOK), University of Nijmegen, Niederlande

Zusammenfassung

In Deutschland wurden seit 2002 einige der weltgrößten Disease-Management-Programme (DMP) – zunächst für die Indikationen Diabetes Typ II und koronare Herzerkrankung – mit geschätzt rund 5 Mio. berechtigten Versicherten und (2005) 2,04 Mio. eingeschriebenen Teilnehmern aufgelegt. Die Programme waren vor Beginn nicht pilotiert worden, und auch von der gesetzlich vorgeschriebenen Regelevaluation ist kein belastbarer Wirksamkeitsnachweis zu erwarten. Der Aufsatz beschreibt das Studienprotokoll für eine randomisiert-kontrollierte Evaluationsstudie und diskutiert als methodische Probleme vor allem die Formulierung von primären Zielkriterien, die Gewährleistung einer ‚naturalistischen‘ Interventionsintensität und die Bildung der Kontrollgruppe unter Bedingungen landesweiter Implementierung. Die vorgeschlagene randomisierte Studie konnte unter bestehenden Bedingungen nicht durchgeführt werden, anhand ihres

Designs lassen sich aber Anforderungen an die Programmevaluation formulieren.

Sachwörter: Disease-Management-Programme – Evaluation – Programmevaluation – Randomisierte kontrollierte Studien

Effectiveness of German disease management programs – problems of clinical evaluation research in the light of a study protocol

Summary

Since 2002 some of the world's largest disease management programs have been

launched in Germany – initially for type II diabetes and coronary heart disease – with approximately 5 million eligible insured persons and (in 2005) 2.04 million registered participants. The programs were not piloted prior to their start, and the prescribed statutory evaluation cannot be expected to result in an effectiveness validation. The article describes the development of a study protocol for a randomized controlled trial. It also discusses methodological problems, in particular the definition of primary target criteria, the guarantee of a 'naturalistic' intensity of intervention, and the creation of a control group during nation-wide implementation. Due to lacking support from the stakeholders the proposed randomized trial could not be implemented. Its design, however, allows the required standards for the program evaluation to be specified.

Key words: disease management – evaluation studies – program evaluation – randomized controlled trials

Notwendig: Summative und formative Programmevaluation

In Deutschland werden Disease-Management-Programme (DMP) für chronische Erkrankungen seit 2002 breit implementiert, ohne dass diese pilotiert wurden oder vorweg ein belastbarer Wirksamkeitsnachweis erbracht wurde. Zwar gibt es umfangreiche internationale Belege für die Wirksamkeit von Disease Management; in Deutschland besteht aber weiterhin ein dringender Evaluationsbedarf, der durch die vorgeschriebene Regel-Evaluation der Programme aus prinzipiellen Gründen nicht zu befriedigen ist. Im Auftrag des AOK-Bundesverbandes hat eine Arbeitsgruppe der allgemeinmedizinischen Universitätseinrichtungen aus Heidelberg und Kiel/Frankfurt in Zusammenarbeit mit dem DiEM (Institut für Evidenz-basierte Medizin, Köln) bereits 2004 ein Protokoll für eine randomisiert-kontrollierte Studie zu den DMP Diabetes Typ II und koronare Herzkrankheit erarbeitet, das in der Folge von den Spitzenverbänden der gesetzlichen Krankenkassen jedoch nicht unterstützt bzw. umgesetzt wurde. Anhand des Studienprotokolls werden in diesem Aufsatz Probleme und Möglichkeiten – auch im Hinblick auf weitere DMP – diskutiert, zumal die Programme in der gesundheitspolitischen Debatte weiterhin umstritten sind.

„Strukturierte Behandlungsprogramme“ (Disease-Management-Programme) wurden in Deutschland mit dem „Gesetz zur Reform des Risikostrukturausgleichs in der gesetzlichen Krankenversicherung“ [1] eingeführt. Heute sind bereits 2,04 Mio. Patienten in eines von rund 6.000 (kassenspezifisch verschiedenen, aber strukturell gleichen) Programmen für Diabetes Typ I und Typ II, koronare Herzkrankheit und Brustkrebs eingeschrieben [2]. Hintergrund der Einführung von DMP waren erkannte Defizite in der Versorgung chronisch Kranker [3]; präzipitiert wurde die politische Durchsetzung aber erst durch die Einsicht, dass mit einem politisch gewollten Wettbewerb zwischen den Kassen der gesetzlichen Krankenversiche-

rung eine Entmischung der Risikostrukturen eintreten werde, gegen die ein entsprechender Ausgleich zu schaffen sei [4, 5]. Akkreditierung und Durchführung, aber auch die Bewertung von Aussichten und Ergebnissen von DMP waren dadurch von Anfang an mit der Hypothek belastet, dass es um einen finanziell sehr gewichtigen Risikostrukturausgleich ging (schon 2004 betrug die Umverteilungssumme pro Diabetes-Patient 1.290 Euro mehr als nach dem vorherigen Verfahren [6]). DMP wurden deshalb unter dem bürokratisch aufwendigen Anspruch rechtsfester Nachprüfbarkeit implementiert. Dieser Anspruch stellte jedoch ein erhebliches Implementierungshindernis dar. Nach den bisherigen Erfahrungen ist festzustellen, dass der Anreiz für die Kassen durch die Anbindung des Risikostrukturausgleichs dazu geführt hat, dass in die DMP und damit die Versorgung chronisch Kranker investiert und der Wettbewerb um die „guten Risiken“ gebremst wurde.

Sachlich fordern die in den jeweiligen Änderungsverordnungen zum Risikostrukturausgleich niedergelegten Grundstrukturen der DMP unterschiedliche, international anerkannte, evidenzbasierte Elemente der strukturierten Versorgung chronischer Erkrankungen. Aus praktischer Sicht am wichtigsten dürften sein:

- Persönliche Einschreibung des einzelnen Patienten in der (Hausarzt-) Praxis
- Evidenzbasierte Leitlinienempfehlungen für Diagnostik und Therapie
- Vereinbarung individueller Behandlungsziele oder Zielparameter
- Strukturierte Dokumentation (einzelner) Behandlungs- oder Outcomeparameter
- Definition von Schnittstellen in der Versorgung
- Anspruch auf Patientenschulung

Zwar ist die Wirksamkeit des Grundansatzes von Disease Management und der genannten Elemente international in Studien gut belegt – vgl. z. B. [7–9]. Jedoch muss bezüglich der Implementierung der DMP auf mindestens zwei Probleme hingewiesen werden. In einem jüngst erschienenen Review von

Wensing et al. wurde dargelegt, dass offenbar Strategien organisatorischer Veränderungen bei der Implementierung von DMP eine eigenständige Bedeutung für den Erfolg und die Verbesserung der Versorgung zukommt [10]. Ob durch die DMP zielführend solche Veränderungen in Deutschland angestoßen wurden, muss derzeit zweifelhaft bleiben. Zudem wurde bereits im vorausgehenden Aufsatz von Gerlach et al. [11] ausgeführt, dass DMP diese Elemente pragmatisch zusammenführen, aber jeweils nur Teilaspekte eines *Chronic Care-Modells* realisieren. Ob bzw. inwieweit bei der Implementierung der deutschen DMP diese Wirkmechanismen von strukturierter Versorgung zum Tragen kommen, sollte daher systematisch evaluiert werden.

Grundsätzlich besteht die gesetzliche Verpflichtung zu einer Regelevaluation zugelassener DMP (Ausführungsbestimmungen hat das Bundesversicherungsamt erlassen [12]). Diese Evaluation erfolgt allerdings ohne Kontrollgruppe, basiert auf (bürokratisch belasteten) Regeldokumentationen und ist streng auf den Akkreditierungszeitraum beschränkt. Primäres Ziel ist hierbei, v.a. vergleichend die Implementierungsfähigkeit und damit den *relativen Erfolg* einzelner Kassen/Verträge zu bewerten. Auswertungen sollen zentral durchgeführt werden (mit Ergänzungserhebungen über Teilstichproben zu Lebensqualität und gesundheitsökonomischen Daten), und erste Ergebnisse sind für Mitte 2006 angekündigt. Dem öffentlichen Legitimationsbedarf kommen einzelne Vertragspartner durch Zwischenauswertungen von Routinedaten (Qualitätssicherungsberichte von Kassenärztlichen Vereinigungen [z. B. 13]) und durch Patientenbefragungen [14] nach. Evaluationen ohne Kontrollgruppe, wie es sie auch bereits in den früheren Modellvorhaben zur Diabetikerbehandlung gab [z. B. 15], sind allerdings immer mit dem Mangel behaftet, dass der Wirkungsmechanismus der Intervention nicht darstellbar ist. Beobachtungsstudien, die auf der Basis von Routinedaten aus gesundheitspolitischen Großprogrammen durchgeführt werden, unterliegen in der Realität noch zahlreichen weiteren Verzer-

rungen: unkontrollierbarer Fluktuation der Beobachtungseinheiten, Selektionsbias(es), Selbstrekutierungseffekten etc., die eine wissenschaftliche Aussagekraft fast unmöglich machen [16]. DMP bedürfen *als komplexe Interventionen der Programmevaluation* – einer immer noch erheblichen Herausforderung für die Versorgungsforschung [16, 17]. Damit ist erstens zwischen den Erfordernissen einer *summativen*, rein auf die Ergebnisse bezogenen, und einer *formativen*, die Implementierung begleitenden und bewertenden Studienausrichtung zu unterscheiden [18]. Es ist anzuerkennen, dass das gesetzliche Evaluationserfordernis auf eine formative Evaluation drängt, d.h. den Vertragspartnern individuell auferlegt, ihre Umsetzungsanstrengungen und -erfolge zu bewerten, wenn auch die Ergebnisse der Regelevaluation hierfür zu spät kommen. Über den prinzipiellen Erfolg (verglichen mit ‚kein DMP‘) ist damit nichts gesagt. Zweitens ist in der Evaluation komplexer Interventionen ein gestuftes Design – zwischen ‚prinzipieller Wirksamkeit der Einzelkomponenten unter Studienbedingungen‘ und ‚tatsächlicher Wirksamkeit der Gesamtmaßnahme unter realistischen Umsetzungsbedingungen‘ – zu beachten [19]. Diese unterschiedlichen Gesichtspunkte einer qualifizierten Programmevaluation werden in der öffentlichen Debatte gemeinhin übersehen.

Als Ergebnis der evaluationstheoretischen Debatte sollte ein letztlich notwendiger Methodenpluralismus [16] festgehalten werden: Die Akteure können sich durch Studien der formativen Evaluation in ihrem weiteren Handeln orientieren (und im Wettbewerb ausrichten); die Prüfung der prinzipiellen Sinnhaftigkeit von DMP („Inwieweit wird die Versorgungsqualität tatsächlich verändert?“) bedarf jedoch einer rigorosen summativen Evaluation. Diese Sinnhaftigkeit (d.h. Wirksamkeit) von DMP ist bis heute umstritten [20–22]. In Ermangelung überzeugender Studien wird immer noch bezweifelt, dass *unter den Bedingungen des deutschen Versorgungssystems* mit den Mitteln von DMP für bestimmte chronische Erkrankungen ergebnisbezogen in ‚der Fläche‘ ein *messbarer* Erfolg erreicht

werden kann. Nach dem internationalen Stand der Bewertung von Studienmethodiken kann eine solche Evidenz nur aus einer randomisierten kontrollierten Studie gewonnen werden [23].

Methode: Entwicklung von randomisiert-kontrollierten Studiendesigns für Diabetes Typ II und KHK

Im Auftrag des AOK-Bundesverbandes haben die Autoren dieses Beitrags zusammen mit Wissenschaftlern des DiEM – Institut für evidenzbasierte Medizin, Köln (P.T. Sawicki, T. Kaiser et al.) – ein Protokoll für eine randomisiert-kontrollierte Studie unter naturalistischen Bedingungen entwickelt, das für die beiden DMP Diabetes Typ II und koronare Herzkrankheit die Frage nach der Wirksamkeit klären sollte. ‚Natürlicher‘ Studienträger sollten die gesetzlichen Krankenkassen gemeinsam sein, damit Patienten unterschiedlicher gesetzlicher Kassen und für die Studienärzte ein homogenes Patientenkontinuum eingeschlossen werden können. Nach mehreren Expertengesprächen ließ sich jedoch der hierfür erforderliche Konsens, die Studie zu tragen, innerhalb der Spitzenverbände der Krankenkassen (SpIK) nicht herstellen.

Das erarbeitete Studienprotokoll bietet nach wie vor Anlass, die Möglichkeit einer systematischen Evaluation von DMP zu demonstrieren und die methodischen Probleme von Studien in diesem Bereich zu diskutieren.

Die Studie sollte nach den allgemein anerkannten Standards randomisiert-kontrolliert durchgeführt werden. Im Rahmen des Studiendesigns waren deswegen v.a. drei schwierige Probleme zu lösen:

1. Wie können – unter den Bedingungen eines bereits angelaufenen DMP – konsistente Studienbedingungen sowie eine konstante Interventionsintensität gewährleistet werden?
2. Welche Outcomes können aufgrund der im Vergleich zum natürlichen

Verlauf der Erkrankungen kurzen Studien- und Interventionsdauer zum Zielkriterium gemacht werden?

3. Wie kann eine möglichst unbeeinflusste Kontrollgruppe gebildet werden?

Studienprotokolle wurden für die beiden DMP Diabetes Typ II und KHK erarbeitet und in ihrem Design nach Möglichkeit aufeinander abgestimmt. Die wesentlichen Merkmale sind in der Synopse (Tabelle 1) dargestellt [24].

Ergebnisse/Diskussion

1. Studiendesign/Kontrollgruppe

Die Studien wurden cluster-randomisiert mit der Studienpraxis als Randomisierungseinheit und dem eingeschriebenen Patienten als Beobachtungseinheit angelegt. Die Clusterrandomisierung [25] war notwendig, um Kontaminierungseffekte in der einzelnen Praxis zu vermeiden, da es schwer vorstellbar erscheint, dass chronisch Kranke in einer Praxis nach unterschiedlichen Vorgaben behandelt werden. Durch die Clusterrandomisierung steigt jedoch die erforderliche Probandenzahl in Abhängigkeit von der ‚Ähnlichkeit‘ der Patienten und der Interventionen innerhalb einer Einheit, die mit dem Intra-Cluster-Correlation(ICC)-Koeffizienten gemessen wird, an. Aufgrund vorgängiger Erfahrungen hatten wir in der Fallzahlberechnung einen ICC von 0.1 bzw. 0.05 angenommen. In die Berechnung geht ferner die Zahl der Randomisierungseinheiten (Praxen – unter Berücksichtigung der Zahl in einer Praxis rekrutierbarer Probanden) ein. Wir fanden ein Optimum der benötigten Probandenzahlen bei den in der Tabelle 1 angegebenen Fallzahlen; diese liegen etwa um das Doppelte höher als bei einer Randomisierung auf Einzelpatientenniveau. Dies erhöhte mithin den Studienaufwand beträchtlich. Große Schwierigkeiten bereitete die Konzeption einer Kontrollgruppe. Da DMP bundesweit und flächendeckend eingeführt werden sollten, und zudem ein Rechtsanspruch auf Teilnahme am DMP (zumindest für eligible Patienten)

Tabelle 1. Synopse der Studiendesigns der Teilstudien zu Diabetes Typ II und koronare Herzkrankheit (nach [24]).

	Teilstudie Diabetes Typ II	Teilstudie KHK
Studiendesign	Prospektive, cluster-randomisierte kontrollierte Studie (RCT); ggf. mit Kosten-Effektivitäts-Analyse	
Forschungsfrage	Führt die Behandlung von Patienten (mit der Zielerkrankung) zu einer Verbesserung der Behandlungsqualität, ausgedrückt in: Symptomatik, Erreichen von Therapiezielen, Häufigkeit von Exazerbationen/Entgleisungen, Mortalität, Risikoprofil, Folgeerkrankungen/Komplikationen, Behandlungsqualität, Lebensqualität und Behandlungsaufwand (Kostengrößen)?	
Studienpopulation	Patienten mit der Diagnose Diabetes Typ II in der ambulanten Versorgung	Patienten mit der Diagnose koronare Herzkrankheit in der ambulanten Versorgung
Einschlusskriterien (zusätzlich: Fähigkeit zur freien, informierten Studienteilnahme)	Einschreibekriterien des DMP [Nachweis diabetestypischer Symptome in Kombination mit Nüchtern-Glukose im Plasma bzw. im Serum > 7,0 mmol/l oder Nicht-Nüchtern-Glukose i. P. > 11,1, mmol/l, auch in der Vorgeschichte oder bei Abwesenheit diabetestypischer Symptome mind. zweimaliger Nachweis von Nüchtern-Glukose i. P. > 7,0 mmol/l oder von Nicht-Nüchtern-Glukose i. P. > 11,1 mmol/l oder Nachweis von Glucose i. P. > 11,1, mmol/l (> 200 mg/dl) 2 Stunden nach oraler Glukosebelastung (75 g Glukose), auch in der Vorgeschichte.]	Einschreibekriterien des DMP [akutes Koronarsyndrom (in der Vorgeschichte) oder 90%ige Wahrscheinlichkeit für KHK oder koronarangiographisch gesicherte KHK]
Statistische Hypothese	Die Teilnahme am DMP für Patienten mit Diabetes (Intervention) erreicht nach 18 Monaten Interventionszeit im Vergleich zur leitliniengestützten Standardtherapie (Kontrolle) eine zweiseitige signifikante Differenz in der Zahl der Patienten, die die vorgegebenen Blutzucker- und Blutdrucktherapieziele erreichen. Eine Verbesserung der ausreichend behandelten Patienten wird mit einer Rate von relativ 50% auf absolut 45% erreicht	Die Teilnahme am DMP für Patienten mit KHK (Intervention) erreicht nach 18 Monaten Interventionszeit im Vergleich zur leitliniengestützten Standardtherapie (Kontrolle) eine zweiseitige signifikante Differenz von 10% im Erreichen des primären (zusammengesetzten) Endpunkts.
Studienintervention	Implementation des DMP in Hausarztpraxen: <ol style="list-style-type: none"> 1. Einführungsschulung zu DMP zur Diagnostik, Kooperation und Behandlung 2. Umsetzung von mindestens zwei DMP-Tools 3. Qualitätszirkelarbeit auf der Grundlage individueller pharmakologischer Feedback-Berichte Die Interventionspatienten erhalten ein hausarztbasiertes DMP: <ol style="list-style-type: none"> 1. Individuelle Risikoeinschätzung und Therapieplanung 2. Patientenschulung 3. Diagnostik und Behandlung leitlinienbasiert 	
Primäres Studienziel/ kombinierter Endpunkt	Klinisches Ereignis: Tod, kardiovaskuläres Ereignis, Majoramputation, Visusverlust, Nierenersatztherapie, schwere Hypoglykämie oder bei Patienten ohne Ereignis: Nichterreichen der Therapieziele zum Zeitpunkt der letzten Evaluationsuntersuchung: <ul style="list-style-type: none"> – HbA1c-Wert soll nach Angaben des behandelnden Arztes gesenkt (nicht gehalten) werden, – der Blutdruckwert, der über 138 mmHg systolisch und/oder über 88 mmHg diastolisch liegt. 	Kardiovaskuläres Ereignis (Tod, Myokardinfarkt, Schlaganfall) oder bei Patienten ohne Ereignis: <ul style="list-style-type: none"> – Nichterreichen der DMP-Therapieziele (letzter Erhebungszeitpunkt) – Nichtvorliegen der indizierten Therapie zur Thrombozytenaggregationshemmung oder – Nichtvorliegen der indizierten Statintherapie – Überschreitung der Blutdruckgrenzwerte
Sekundäre Studienziele (Auswahl)	<ul style="list-style-type: none"> – jeder klinische Endpunkt (Analyse getrennt) – Erreichen jedes der beiden Therapieziele – Rate der empfohlenen Untersuchungen und Therapiemaßnahmen – Neuauftreten einzelner diabetischer Komplikationen – Lebensqualität und Therapiezufriedenheit der Patienten – Adherence zur Therapie (Patient/Arzt) 	Patientenbezogene Zielgrößen <ul style="list-style-type: none"> – Gesamt mortalität – Kardiale Mortalität – Akutes Koronarsyndrom – Myokardinfarkt – Schlaganfall – Angina Pectoris-Beschwerden – Kardiovaskuläres Risiko (Proxy)

Tabelle 1. (Fortsetzung).

Teilstudie Diabetes Typ II		Teilstudie KHK
		Klinisch-kardiales Staging: Hypertonie, Angina Pectoris (CCS Klassifikation), Herzinsuffizienz (NYHA), bei Postinfarktpatienten (LV-Ejektionsfraktion) Diabetes m. Gesundheitsverhalten – Patientenadherence – Aktivitätsstatus – Diätstatus – Body Mass Index – Lebensqualität – Patientenzufriedenheit mit der Versorgung: EUROPEP – Depressivität Arztbezogene Adherence zu den DMP-Vorgaben
über einen gesundheitsökonomischen Studienteil war noch nicht abschließend entschieden		
Probandenzahlen	N = etwa 2.000 Bei alpha = 0,05, beta = 0,20: um den erwarteten Unterschied zweiseitig signifikant darzustellen, sind bei einem ICC = 0.05 etwa 100 Praxen mit je 20 Patienten notwendig	N = 2.140 Bei alpha = 0,05, beta = 0,20: um einen Unterschied von 10% zweiseitig signifikant darzustellen, sind bei einem ICC = 0.1 107 Praxen mit je 20 Patienten notwendig Mit dieser Probandenzahl hätte auch eine absolut 6%ige Senkung des kardiovaskulären Zehnjahresrisikos mit dem PROCAM-Score (bei einem ICC von hier 0.02) als signifikant nachgewiesen werden können (Sekundärauswertung)

bestand, konnte (beim Diabetes-DMP) nicht auf Kontrollregionen ohne DMP-Einführung zurückgegriffen werden. Da nicht an dem DMP teilnehmende Vertragsärzte sich vermutlich von den DMP-Ärzten unterscheiden würden, sollte die Kontrollgruppe auch nicht aus den Nicht-Teilnehmern rekrutiert werden. Schließlich wurde die Lösung diskutiert, Praxen, die in die Kontrollgruppe randomisiert werden sollten, von den Behandlungsaufträgen im Rahmen des DMP freizustellen (und entsprechend in der Vergütung zu entschädigen). Zwar hätte hierdurch in der Kontrollgruppe nicht der status quo ante (vor der Implementierung der DMP) repräsentiert werden können. Es ist jedoch zu vermuten, dass angesichts der Tatsache, dass die (ordnungsgemäße) Durchführung eines DMP von der einzelnen Praxis erhebliche Implementierungsanstrengungen erfordert, die Kontrollpraxen den Stand der ‚usual care‘

hinreichend getreu wiedergegeben hätten. Im Fall der Durchführung der Studie wäre diese Lösung gewählt worden. Aus ethischer Sicht wäre diese Lösung vermutlich nicht zu beanstanden, weil Ärzte und Patienten in der Kontrollgruppe frei gewesen wären, individuelle Behandlungsentscheidungen zu treffen.

2. Naturalistisches vs. experimentelles Design – ‚Interventionsintensität‘

Lange diskutiert – auch im Expertenkreis der SpiK – wurden zwei Fragen: Erstens, ob durch die Studienteilnahme nicht ein Selektionseffekt in Richtung auf für DMP besonders motivierte und leistungsfähige Praxen zustande kommen würde. Und zweitens, ob im Rahmen einer kontrollierten Studie nicht ein Beobachtungseffekt einträte im Hinblick auf eine unnatürlich sorgfälti-

ge und ‚wirklichkeitsfremde‘ Durchführung eines DMP.

Der erstgenannte Selektionseffekt wäre in einem Kontrollgruppendesign weitgehend aufgefangen worden. Bezüglich des Beobachtungseffekts ist zu bedenken, dass jede kontrollierte Studie eine experimentelle Eingrenzung und damit Beeinflussung der Beobachtungssituation beinhaltet (sogenannter „Hawthorne-Effekt“). In einer Studie zur Programmevaluation bedeutet dies zunächst, dass eine Homogenität der Intervention (bzw. Programmkonstanz) durch im Protokoll beschriebene Schritte der Vorbereitung und des Monitorings der Studienpraxen zu definieren war. Diese war an die Vorgaben des DMP bzw. der örtlichen DMP-Verträge angelehnt und hätte eine vollständige, aber nicht besonders unterstützte Umsetzung der Vorgaben in den Studienpraxen bedeutet. Zwar ist zu vermuten, dass bei der flächendeckenden Einfüh-

rung eines DMP in einzelnen Bereichen eine suboptimale Implementierung erfolgt ist oder erfolgen wird (mit dem Effekt reduzierter Effektivität). Es hätte jedoch keinen Sinn gehabt, ein solches ‚unteroptimales‘ DMP in der Studie zu untersuchen. Schwachstellen in der Implementierung können und sollen vielmehr mit der Regevaluation aufgedeckt werden [12].

Versteht man die hier dargestellte Protokollentwicklung zugleich als Fallstudie zur Evaluation im deutschen Gesundheitswesen, muss vermerkt werden, dass sich in dieser Frage massive Interessenunterschiede zwischen den einzelnen Kassenarten widerspiegeln.

Inzwischen wurde an der Universität Heidelberg eine randomisierte Studie (ELSID) entwickelt, in der dieses Problem in den Mittelpunkt gestellt wird. Mit Förderung durch den AOK-BV wird diese auch tatsächlich durchgeführt. In einem dreiarmigen Design werden hier die Routineumsetzung des DMP Diabetes mit einer (z. B. durch Qualitätszirkelarbeit und Praxisbesuche) optimierten Umsetzung, sowie beide Interventionsgruppen mit einer Kontrollgruppe ohne DMP verglichen [26]. Hierdurch kann insbesondere die weiterführende Frage abgebildet werden, welche Unterstützung Leistungserbringer benötigen [10], um ein solches Programm strukturierter Versorgung erfolgreich umsetzen zu können.

3. Bestimmung von primären Zielkriterien – kombinierte Endpunkte

Das methodisch interessanteste, zugleich schwierigste Problem ergab sich bei der Definition der primären Zielkriterien der Studie, da die Zielkriterien in einer kontrollierten Studie von größter Bedeutung für das gesamte Studiendesign (Dauer, Fallzahlen usw.) sind. Eine kontrollierte Studie hat stets eine begrenzte Studiendauer, die im Fall der DMP besonders klein im Vergleich zu der mutmaßlichen Zeitdauer ist, in der die wesentlichen Behandlungseffekte (bzw. primären Endpunkte) der strukturierten Chronikerversorgung eintreten. So wird als einer der entscheidenden

Effekte einer DMP-gemäßen Behandlung von Diabetikern und Patienten mit koronarer Herzkrankheit die Vermeidung gravierender Komplikationen (Erblindung, Amputationen, kardiovaskuläre Ereignisse) bzw. Reduktion der Mortalität erwartet. Diese stellen sich allerdings erst über jahrelange Zeiträume ein und können daher während der Studiendauer nur eingeschränkt beobachtet werden.

Prinzipiell gab es drei Möglichkeiten, primäre Zielkriterien zu definieren:

- a) **Primärkriterien im Sinne ‚harter‘ klinischer Endpunkte** (bzw. major outcomes): Tod, kardiovaskuläre Ereignisse, gravierende Komplikationen. Zwar kann keine klinische Studie an diesen Endpunkten vorbeigehen – in jedem Fall hätte ein Monitoring von Mortalität und Morbidität schon aus ethischen Gründen stattfinden müssen; die erwartete Differenz wäre aufgrund des vermuteten Zusammenhangs zwischen Behandlungsqualität und Ereigniseintritt im Beobachtungszeitraum allerdings nur gering gewesen.
- b) **Primärkriterien ausschließlich aus dem Bereich der Prozessparameter bzw. Surrogatparameter.** Da die Studie von vornherein nicht als Implementierungsstudie, sondern als Wirksamkeitsstudie im Hinblick auf Patientenoutcomes geplant war, konnte es nicht ausreichend sein, Zielparame-ter aus dem Bereich der Behandlungsvorgaben (z. B. Zahl der Fußinspektionen oder Fundoskopien) zu wählen. Das Monitoring dieser Vorgaben erfolgt ohnehin im Rahmen der Regevaluation. Für z. B. das DMP Diabetes boten sich vereinbarte und überprüfte Behandlungsziele (Blutdruckeinstellung, HbA1c-Werte, Vermeidung von Stoffwechsellagen) an. Diesen Zielparame-tern kann zwar eine Aussagekraft für den Gesundheitszustand und die gesundheitliche Prognose der Patienten zugesprochen werden; die Enge und Evidenzbasiertheit des Zusammenhangs zum Eintritt von Komplikationen und major events ist allerdings nicht durchgängig belegt [27, 24,

S. 20 ff.]. Parameter aus diesem Bereich haben jedoch den Vorteil der leichten Erfassbarkeit ohne zusätzlichen Studienaufwand bzw. zusätzliche Beeinflussung der Behandlungssituation.

- c) Angesichts der grundsätzlichen Zielsetzung von strukturierten Behandlungsprogrammen für chronische Erkrankungen hätte es sich angeboten, als **primäres Zielkriterium** die **Senkung des individuellen gesundheitlichen Gesamtrisikos** (nach z. B. einem anerkannten Risikoscore) zu wählen. Für die Teilstudie zum DMP KHK wurde dies während der Protokollentwicklung auch lange erwogen. Wir konnten jedoch kaum internationale Studien finden, in denen zur Evaluation eines strukturierten Behandlungsprogramms ein solcher Ansatz bereits erprobt worden wäre. Methodische Probleme ergaben sich insbesondere daraus, dass es für das kardiovaskuläre Gesamtrisiko zwar unter primärpräventivem Blickwinkel (d. h. vor Eintritt einer koronaren Herzkrankheit bzw. eines koronaren Ereignisses) gut erprobte Risikoscores gibt. Für Patienten mit diagnostizierter KHK ist dies jedoch schwieriger, weil unterschiedliche Risikokonstellationen (z. B. in Abhängigkeit vom Koronarbefund, vom zeitlichen Abstand eines Myokardinfarktes etc.) zu beachten sind. Dennoch hätte auch hier einer der anerkannten Risikoscores eine erste Annäherung an eine Aussage über die Behandlungsqualität und prognostisch relevante Risikosenkung ermöglicht. Im Rahmen der Studie sollte daher auf der Grundlage der Daten ein KHK-spezifisches Staging entworfen werden, das auch für spätere Studien einen methodischen Fortschritt bedeutete hätte. Ein weiteres Problem ergab sich jedoch daraus, dass die Methodik zu neuartig war, und die Aussagekraft der Studie angesichts der hohen Erwartungen in der gesundheitspolitischen Diskussion an die Ergebnisse relativiert hätte. Das Protokoll hatte diesen Analyseweg dennoch für eine zusätzliche explo-

rative Sekundärauswertung vorgesehen.

Um die Aussagekraft der Studie zu optimieren und zudem in beiden Studienarmen ein gleichartiges Ergebnis zu erzielen, wurde im endgültigen Studienprotokoll für beide Arme ein *kombinierter Endpunkt* als primäres Kriterium formuliert, der sich jeweils aus harten klinischen Endpunkten (letal und gravierenden Komplikationen) sowie dem (Nicht-)Erreichen relevanter Therapieziele zusammensetzte (s. Tabelle 1). Kombinierte Endpunkte werfen eine spezifische methodische Problematik auf: sie sind in zahlreichen neueren Studien ‚in Mode gekommen‘¹; valide sind sie jedoch nur, wenn sie gleich relevante und in der Größenordnung gleich häufige (-variante) Ereignisse zusammenfassen [28]. Die Lösung, die wir im Studienprotokoll formuliert hatten, konnte nicht voll befriedigen; sie erschien allerdings vor dem Hintergrund der Evidenzlage zum Effekt von Programmen strukturierter Versorgung gerechtfertigt. Die Komponenten des kombinierten Endpunktes waren zudem als *sekundäre Zielkriterien* einzeln ausgewiesen.

Studiendesigns sollten grundsätzlich auf *ein* Primärzielkriterium ausgerichtet sein, damit die methodischen Probleme multiplen Testens vermieden werden; daher erschien die Formulierung eines kombinierten Endpunktes nötig. Im Bereich der DMP (resp. der Diabetes-Therapie) kommt jedoch hinzu, dass häufig der Zusammenhang zwischen adäquater Therapieeinstellung und dem (Nicht-)Eintritt bestimmter Komplikationen verkürzt dargestellt wird [21] und damit Effektivitätsreserven strukturierter Versorgung nicht gesehen werden [22].

Schlussfolgerungen für die Praxis

Die Notwendigkeit randomisierter kontrollierter Interventionsstudien zur Eva-

luation der deutschen DMP lässt sich in sieben Punkten begründen [24]:

1. Evaluationsstudien auf der Grundlage eines Beobachtungs- bzw. Kohortendesigns haben eine zu geringe methodische Stärke, um den eigentlich in Frage stehenden Punkt: die Wirksamkeit von DMP, zu belegen.
2. Gegenüber der Reevaluation kommen kontrollierte Studien mit – biometrisch berechenbaren – wesentlich geringeren Stichprobengrößen aus und sind daher (v. a. wenn auf die bei der Regeldokumentation nicht voraussetzbare Datenqualität Wert gelegt wird) sogar letztlich kosteneffektiver.
3. Kontrollierte Studien können auf den Gesichtspunkt der summativen Evaluation fokussieren, weil sie die Randbedingungen der Intervention (Programmkonstanz) weitgehend definieren (in welchem Ausmaß – hängt von der Stärke der Präferenz für ein ‚naturalistisches Design‘ ab). Weil sie *Wirkungsmodelle* prüfen, können sie aber auch Einsichten in das Prozessverständnis eröffnen, die in reinen Beobachtungsstudien wegen der Interdependenz von einzelnen Faktoren i. d. R. verschlossen bleiben.
4. Obwohl oft die Frage gestellt wird, ob noch einmal in Deutschland bewiesen werden muss, wofür ‚überwältigende internationale Evidenz‘ doch schon bestehe: dass einzelne Komponenten einer strukturierter Versorgung chronischer Erkrankungen erhebliche Verbesserungen der Outcomes bewirken können (vgl. dazu die anderen Beiträge dieses Heftes), wird in der gesundheitspolitischen Debatte hierzulande immer noch der grundsätzliche Wirkungsbeweis von DMP eingefordert [20, 21].
5. Für die Begründung von krankheitsspezifischen Interventionen (dem Betablocker für KHK-Patienten) oder Versorgungskomponenten (Case Management bei Herzinsuffizienz) wird Evidenz auf höchstem internationalen Niveau gefordert. Prinzipiell ähnliche methodische Maßstäbe müssen auch bei der Be-

wertung eines der weltweit umfassendsten Behandlungsprogramme: der deutschen DMP, gelten.

6. Das Prozessverständnis – über das ‚Innenleben‘ der DMP-Verträge – ist bereits kurz- bis mittelfristig erforderlich, um die DMP zu optimieren. Hierfür ist eine *kausale Interpretation* der Einflüsse von Teilaspekten notwendig. Zwar ist über eine kausale Interpretation des Ergebnisses einer randomisierten-kontrollierten Studie jeweils zu streiten (je nach dem Ausmaß, in dem Randbedingungen effektiv kontrolliert worden sind, bzw. in dem die Studie über eine bloß verfeinerte Beobachtung hinausgeht und den experimentellen Charakter der Prüfung der Wirkungen einer Intervention besitzt), zumindest sollte sich die *Richtung* einer Wirkungsbeziehung durch inferenzstatistische Verfahren darlegen lassen.
7. Schließlich wird es zukünftig auch darum gehen, die Effektivität und Effizienz verschiedener DMP und verschiedener Modelle strukturierter Versorgung im Vergleich bewerten zu können. Nur über *kontrollierte Studien* hinweg lassen sich ggf. normierbare Effektstärken vergleichen. So ist die Qualität der Diabetesbehandlung in Deutschland durch eine Vielzahl von Modellprojekten bereits vor dem Start der DMP gefördert worden (mit allerdings im internationalen Vergleich immer noch unklaren Ergebnissen), während die Behandlungsqualität der (Risikovorstufen einer) KHK wohl unumstritten noch unbefriedigend ist. Das Studienprotokoll wurde u. a. auch deswegen für zwei unterschiedliche Indikationsbereiche mit weitgehend gleicher Methodik ausgearbeitet, um eine solche Vergleichsmöglichkeit zu erproben.

Mit der Vorlage und Diskussion des Studienprotokolls für eine randomisierte kontrollierte, summative Evaluationsstudie der DMP soll das Thema der Evaluationsmethodik erneut auf die Tagesordnung der evaluativen Versorgungsforschung in Deutschland gesetzt werden und gezeigt werden, dass derartige Studien in Deutschland sowohl nötig

¹ Vgl. u. a. die Debatte um die sog. PROactive-Studie (zu Pioglitazon) – arznei-telegramm 11/2005 S. 95f.

wie möglich sind. Insbesondere angesichts der international herausragenden Reichweite (berechtigt sind allein für die beiden angesprochenen DMP geschätzt rund fünf Mio. Patienten) dürfen DMP nicht als politischer Großversuch mit beliebig interpretierbarem Ausgang aufgefasst werden können. Zwar wird es zumindest eine Studie geben [26], die weitergehende Aufschlüsse über das DMP Diabetes Typ II erbringen wird, zu fordern sind allerdings weitere Studien mit einer fortentwickelten Methodik evaluativer Versorgungsforschung, welche die Outcomes (auch anderer Zielerkrankungen) noch besser in den Blick nehmen können. Hierbei ist auch die Beteiligung aller Vertragspartner an den DMP unabhängig von kurzfristigen Interessen zu fordern, um die erfolgreiche Umsetzung strukturierter Versorgung und verbesserte Ergebnisse in der Versorgung chronisch Kranker zu fördern.

Conflict of interest: Sämtliche Autoren haben – wie im Text beschrieben – an der Entwicklung eines Studienprotokolls zur Evaluation der DMP Diabetes Typ II und KHK mitgewirkt, die vom AOK-Bundesverband finanziell gefördert wurde. J. S. und F. M. G. haben darüber hinaus in Expertengremien des ehemaligen Koordinierungsausschusses (jetzt Gemeinsamer Bundesausschuss) an der Formulierung der Empfehlungen für die DMP KHK bzw. Asthma/COPD teilgenommen.

Literatur

- [1] § 137 SGB V; 10.12.2001; BGBl. I S. 3443
- [2] Bundesministerium für Gesundheit. Pressemitteilung. Stand: Dezember 2005. www.bmg.bund.de, acc. 21.02.06.
- [3] Sachverständigenrat für die Konzertierte Aktion im Gesundheitswesen. Bedarfsgerechtigkeit und Wirtschaftlichkeit. Band III.1 Grundlagen, Übersicht, Versorgung chronisch Kranker. Gutachten 2000/2001. Baden-Baden: Nomos Verlagsgesellschaft; 2001.
- [4] Busse R. Disease management programs in Germany's statutory health insurance system. A Gordian solution to the adverse selection of chronically ill people in competitive markets? *Health Affairs* 2004;23(1):56–67.
- [5] Erler A. Die Einführung von Disease-Management-Programmen in Deutschland im Spiegel unterschiedlicher Interessenlagen von Krankenkassen und Kassenärztlichen Vereinigungen. *Das Gesundheitswesen* 2002;64:572–7.
- [6] Bundesversicherungsamt. Pressemitteilung vom 07.11.2005. www.bundesversicherungsamt.de, acc. 21.03.06.
- [7] Weingarten SR, Henning JM, Badamgarav E, Knight K, Hasselblad V, Gano Jr A et al. Interventions used in disease management programmes for patients with chronic illness-which ones work? meta-analysis of published reports. *BMJ* 2002;325:1–8.
- [8] Wagner EH, Austin BT, Davis C, Hindmarsh M, Schaefer J, Bonomi A. Improving chronic illness care: Translating evidence into action. *Health Affairs* 2001;20(6):64–78.
- [9] Ofman JJ, Badamgarav E, Henning JM, Knight K, Gano AD Jr, Levan RK et al. Does disease management improve clinical and economic outcomes in patients with chronic diseases? A systematic review. *Am J Med* 2004;117: 182–92.
- [10] Wensing M, Wollersheim H, Grol R. Organizational interventions to implement improvements in patient care: a structured review of reviews. *BMC Implementation Science* 2006;1:2.
- [11] Gerlach FM, Beyer M, Muth C, Saal K, Gensichen J. Neue Perspektiven in der allgemeinmedizinischen Versorgung chronisch Kranker – Wider die Dominanz des Dringlichen. Teil I/Teil II. *Z ärztl Fortbild Qual* 2006;100:335–53.
- [12] Evaluationskriterien zu Diabetes mellitus Typ 2, i. d. F. v. 01.06.2005. <http://www.bundesversicherungsamt.de/Fachinformationen/Dmp/Evaluation-Diabetes-2.pdf>, acc. 21.03.06.
- [13] Nordrheinische Gemeinsame Einrichtung Disease-Management-Programme, editor. Qualitätssicherungsbericht 2004. Disease-Management-Programme in Nordrhein. Dezember 2004. http://www.kvno.de/importiert/qualbe_dmp04.pdf (acc. 02.02.2006). Gemeinsame Einrichtung DMP in Hessen: Qualitätssicherungsbericht 2005 Disease-Management-Programm Diabetes Mellitus Typ II. http://www.kv-hessen.de/documents_getdoc.cfm?pdfdoc=QS-Bericht_2005_DMP_Diabetes.pdf acc. 21.03.06.
- [14] veröffentlicht z. B. AOK-BV <http://www.aok-gesundheitspartner.de>

- de/bundesverband/dmp/meldungen/in dex_06558.html, acc. 21.03.06.
- [15] Altenhofen L, Haß W, Oliveira J, Brenner G. Modernes Diabetesmanagement in der ambulanten Versorgung. Köln: Deutscher Ärzte-Verlag, 2002.
- [16] Gerlach FM, Beyer M, Szecsenyi J, Raspe H. Evaluation von Disease-Management-Programmen – Aktuelle Defizite, Anforderungen, Methoden. Z ärztl Fortbild Qual 2003; 97(7):495–501.
- [17] Morfeld M, Wirtz M. Methodische Ansätze in der Versorgungsforschung. Das Beispiel Evaluation der Disease-Management-Programme. Bundesgesundheitsblatt-Gesundheitsforschung-Gesundheitsschutz 2006;49(2):107–10.
- [18] Øvretveit J. Evaluation gesundheitsbezogener Interventionen. Bern: Huber; 2002.
- [19] Campbell M, Fitzpatrick R, Haines A, Kinmonth AL, Sandercock P, Spiegelhalter D et al. Framework for design and evaluation of complex interventions to improve health. BMJ 2000; 321:694–6.
- [20] Bullmann C, Straub C. DMP zwischen Anspruch und Wirklichkeit. Eigentlich sollte alles besser werden. Z ärztl Fortbild Qual 2006;100:32–5.
- [21] Häussler B, Berger U. Bedingungen für effektive Disease-Management-Programme (2004) http://www.iges.de/sites/iges.de/myzms/content/e1788/e1818/e1820/Endfassunginkl.TitelundUmschlagseite_ger.pdf, acc. 21.03.06.
- [22] Raspe H, Sawicki P, Schmacke N. Sind vorliegende Disease-Management-Programme für Diabetes wirklich unnötig? GGW 2004;23–31.
- [23] Rychetnik L, Frommer M, Hawe P, Shiell A. Criteria for evaluating evidence on public health interventions. Journal of Epidemiology and Community Health 2002;56:119–27.
- [24] Deutsches Institut für evidenzbasierte Medizin, Köln, und Institut für Allgemeinmedizin, Universitätsklinikum Schleswig-Holstein, Campus Kiel, Sektion Allgemeinmedizin und Versorgungsforschung, Universitätsklinikum Heidelberg. Studienprotokoll zur kontrollierten Evaluation der Effekte von Disease-Management-Programmen (Diabetes Typ 2, KHK). Autoren: Sawicki, Kraus, Heise, Kaiser (DieM), Gensichen, Gerlach (Kiel), Wensing, Szecsenyi (Heidelberg). Das Studienprotokoll lag in der endgültigen Fassung mit Datum vom 27.05.2004 vor. Eine vollständige Fassung ist im Internet unter www.allgemeinmedizin.uni-frankfurt.de/dmp_evaluation.html zugänglich.
- [25] Campbell MK, Elbourne DR, Altman DG. CONSORT statement: extension to cluster randomised trials. BMJ 2004; 328:702–8.
- [26] Joos S, Rosemann T, Heiderhoff M, Wensing M, Ludt S, Gensichen J et al. ELSID-Diabetes study-evaluation of a large scale implementation of disease management programmes for patients with type 2 diabetes. Rationale, design and conduct – a study protocol. BMC Public Health 2005;5(99).
- [27] Mühlhauser I, Berger M. Surrogat-Marker: Trugschlüsse. Dt Ärzteblatt 1996; 93(49):A-3280–3.
- [28] Montori VM, Permyer-Miralda G et al. Validity of composite end points in clinical trials. BMJ 2005;330:594–6.

Korrespondenzadresse:

Dipl. Soz. Martin Beyer
 Institut für Allgemeinmedizin
 Johann Wolfgang Goethe-Universität
 Theodor-Stern-Kai 7
 60590 Frankfurt
 Tel.: ++49-(0)69-6301-5687
 Fax: ++49-(0)69-6301-6428
 E-Mail:
beyer@allgemeinmedizin.uni-frankfurt.de
 Homepage:
www.allgemeinmedizin.uni-frankfurt.de

