

Nachhaltigkeit und Sicherung von Qualität

Evaluation im Gesundheitswesen

Max-Erik Niehoff

Nachhaltigkeit und Sicherung von Qualität

Evaluation im Gesundheitswesen

Max-Erik Niehoff

Impressum

Verfasser

Rechtsanwalt Max-Erik Niehoff, MBA HCM

geschäftsführender Gesellschafter der SalusCon GmbH – HealthCare Management

Geschäftsführer der SalusCon Akademie für Fort- und Weiterbildung im Gesundheitswesen

gemeinnützige UG

geschäftsführender Vorstand im bbb, Biotechnologieverbund Berlin-Brandenburg e.V.

Inhaltsverzeichnis

Abkürzungsverzeichnis	4
Einleitung	5
1 Einführung in die Evaluation	6
1.1 Basiswissen Evaluation	6
1.1.1 Evaluationskonzepte	8
1.1.2 Nutzen von Evaluationen	10
1.1.3 Ziele der Evaluation	11
1.2 Evaluationsansätze	12
1.2.1 Der CEval-Evaluationsansatz	14
1.2.1.1 Organisationsmodell	14
1.2.1.2 Innovations-/Diffusionsmodell	16
1.2.2 Evaluation von Nachhaltigkeit	17
1.3 Evaluationsdurchführung	18
1.3.1 Intern oder extern?	18
1.3.2 Ziel- oder wirkungsorientiert?	21
2 Konkurrierende und komplementäre Ansätze zur Evaluation	24
2.1 Controlling	24
2.2 Balanced Scorecard	25
2.3 Benchmarking	25
2.4 Audit	26
2.5 Fazit	27
3 Evaluation im Rahmen von Qualitätsmanagement und -sicherung	28
4 Evaluation im Gesundheitswesen	31
4.1 Evaluation von Gesundheitspolitik	31
4.2 Evaluation auf Seiten der Kostenträger	32
4.3 Evaluation auf Seiten der Leistungserbringer	33
5 Sozialer Kontext der Evaluation	35
6 Fallstudie	37
Zusammenfassung	38
Literaturverzeichnis	39

Abkürzungsverzeichnis

AQUA	Institut für angewandte Qualitätsförderung und Forschung im Gesundheitswesen GmbH
AV	abhängige Variable
BMG	Bundesministerium für Gesundheit
BSC	Balanced Scorecard
DeGEval	Deutsche Gesellschaft für Evaluation
DIN	Deutsches Institut für Normung
DKI	Deutsches Krankenhausinstitut
DRG	Diagnosis Related Groups
DZVhÄ	Deutscher Zentralverein homöopathischer Ärzte
EN	Europäische Norm
GKV	Gesetzliche Krankenversicherung
InEK	Institut für das Entgeltsystem im Krankenhaus
ISO	International Standard Organization
KHG	Krankenhausfinanzierungsgesetz
PKV	Private Krankenversicherung
SGB	Sozialgesetzbuch
UV	unabhängige Variable
WAMP	Wandel von Medizin und Pflege im DRG-System
ZEEZ	Zentrum für die Evaluation der Entwicklungszusammenarbeit

Einleitung

Die Versorgung der Bevölkerung mit Gesundheitsdienstleistungen als hoheitliche Aufgabe der Daseinsvorsorge wird in Deutschland durch eine Vielzahl von Normen geregelt. Zentrales Normenwerk des öffentlichen Gesundheitswesens ist hierbei das fünfte Sozialgesetzbuch (SGB V). Es regelt weitgehend die Beziehungen der Akteure im Gesundheitswesen (Patienten, Versicherte, Versicherungen und Leistungserbringer) untereinander.

„Die Krankenversicherung als Solidargemeinschaft hat die Aufgabe, die Gesundheit der Versicherten zu erhalten, wiederherzustellen oder ihren Gesundheitszustand zu bessern“ (§ 1 Abs. 1 S. 1 SGB V). „Die [hiernach erbrachten] Leistungen müssen ausreichend, zweckmäßig und wirtschaftlich sein; sie dürfen das Maß des Notwendigen nicht überschreiten. Leistungen, die nicht notwendig oder unwirtschaftlich sind, können Versicherte nicht beanspruchen, dürfen die Leistungserbringer nicht bewirken und die Krankenkassen nicht bewilligen.“ (§ 12 Abs. 1 SGB V)

Die Erfüllung dieser Zielsetzungen des Gesetzgebers bei bestehender Begrenztheit der dem Gesundheitswesen zur Verfügung stehenden Mittel (Geld, Personal etc.) bedarf der fortwährenden Überprüfung und Bewertung mit dem Ziel der zweckentsprechenden Ressourcenallokation.

Dabei konkurrieren sowohl einzelne Gesundheitsprogramme als auch die verschiedenen öffentlichen Haushalte um die zur Verfügung stehenden öffentlichen Mittel. In Zeiten knapper Kassen hat dabei die Evaluationsforschung, eine in Deutschland eher junge Disziplin, in den letzten Jahren stark an Bedeutung gewonnen.

Das vorliegende Skript ist eine systematische Einführung in den Bereich der Evaluation im Gesundheitswesen und verfolgt das Ziel, die Leser zu befähigen

- ⇒ Nutzen und Wirkung von berufsfeldbezogenen Maßnahmen oder Programmen selbständig zu analysieren, zu beschreiben und zu bewerten,
- ⇒ zentrale Ergebnisse von Evaluationsstudien zu beurteilen und auf den eigenen beruflichen Kontext zu übertragen sowie
- ⇒ Implementierungsstrategien für empirisch erprobte Konzepte für den eigenen beruflichen Kontext zu modifizieren.

1 Einführung in die Evaluation

Im vorliegenden Kapitel soll zunächst grundlegendes Wissen über Evaluationen vermittelt werden (Abschnitt 1.1). Nach einer Begriffserläuterung werden zentrale Evaluationsfragen identifiziert und systematisch erläutert. Zudem wird auf das Spannungsverhältnis zwischen Wissenschaftlichkeit und Nützlichkeit von Evaluation eingegangen.

Evaluationen dienen im Groben vier Zielen: Dem Erkenntnisgewinn, der Kontrolle, der Entwicklung sowie der Legitimation. Auch wenn diese Ziele eng miteinander verbunden sind, erfordern sie je nach Zielstellung unterschiedliche Herangehensweisen und Konzepte. Da Evaluationen unterschiedlichsten Anforderungen gerecht werden müssen, existiert eine Vielzahl von Evaluationsansätzen und -modellen, die in Abschnitt 1.2 näher beleuchtet werden. Außerdem wird dort ein am Centrum für Evaluation (CEval) entwickelter wirkungsorientierter Evaluationsansatz vorgestellt, der grundsätzlich in allen Phasen des Programmprozesses einsetzbar ist.

Anschließend wird in Abschnitt 1.3 näher auf die Evaluationsdurchführung eingegangen.

1.1 Basiswissen Evaluation

Der Begriff erfreut sich wachsender Beliebtheit. Dabei scheint nicht immer klar zu sein, was er eigentlich bedeutet, taucht der Begriff doch in den verschiedensten Zusammenhängen auf und wird zur Bezeichnung der unterschiedlichsten Verfahren verwendet. Auffällig ist, dass Evaluation nicht nur für einen Prozess (der Durchführung einer Evaluation) steht, sondern auch für das Ergebnis desselben.

Evaluation in seiner allgemeinsten Form stellt eine Bewertung oder Beurteilung eines Sachverhalts oder Objektes auf der Basis von Informationen dar (Stockmann 2007: 25). Hinzu kommt, dass Evaluationen zielgerichtet erfolgen, nämlich zu dem Zweck, eine Entscheidung zu treffen. Demnach erfolgt der Ablauf einer Evaluation unabhängig davon, welchem Zweck sie im Einzelnen dient, einem generellen Muster. Drei Schritte lassen sich identifizieren: 1. Informationen werden gesammelt, 2. bewertet und 3. auf dieser Basis Entscheidungen getroffen. Allerdings müssen die Personen, die eine Evaluation durchführen und diejenigen, die die Ergebnisse nutzen und Entscheidungen treffen, nicht identisch sein. Im Gegenteil, später wird gezeigt werden, dass diese beiden Funktionen in der Regel getrennt sind (vgl. Stockmann 2007: 25).

Gegenstand von Evaluationen können Gesetze, Produkte, Dienstleistungen, Organisationen, Personen, Prozesse sowie soziale Tatbestände jedweder Art oder gar Evaluationen selbst sein. Häufig sind die Untersuchungs- und Bewertungsgegenstände von Evaluationen allerdings Reformmaßnahmen, Projekte und Programme (vgl. Stockmann 2007: 26).

Wenn Sachverhalte oder Gegenstände von einer oder mehreren Personen bewertet werden, muss festgelegt werden, anhand welcher Kriterien dies erfolgen soll. Da die hierfür ausgewählten Bewertungskriterien natürlich sehr verschieden sein können, ist schon deshalb mit sehr unterschiedlichen Evaluationsergebnissen zu rechnen.

Neben den Aspekten was, von wem, anhand welcher Kriterien evaluiert wird, hat die Frage, wie bewertet wird, einen entscheidenden Einfluss auf das Evaluationsergebnis.

Diese Aspekte von Evaluation finden sich auch in den gängigen Definitionsversuchen wieder. Nach Mertens (1998: 219) ist **Evaluation**

„the systematic investigation of the merit or worth of an object (program) for the purpose of reducing uncertainty in decision making.“

Definitionen

Rossi et al. (1988: 3) definieren die **Evaluationsforschung**

„als systematische Anwendung sozialwissenschaftlicher Forschungsmethoden zur Beurteilung der Konzeption, Ausgestaltung, Umsetzung und des Nutzens sozialer Interventionsprogramme. Evaluationsforschung bezeichnet den gezielten Einsatz sozialwissenschaftlicher Forschungsmethoden zur Verbesserung der Planung und laufenden Überwachung sowie zur Bestimmung der Effektivität und Effizienz von (...) sozialen Interventionsmaßnahmen“.

Bei einer Evaluation kommt es demnach sehr darauf an, wer was anhand welcher Kriterien wie evaluiert. Besonders in öffentlichen Diskussionen wird dies aber häufig vergessen. Evaluationsergebnisse werden hier für die verschiedensten Begründungszusammenhänge herangezogen und damit ganz unterschiedliche Ziele verfolgt. Es ist daher wichtig, dass Evaluationen professionell, nach wissenschaftlichen Kriterien und unter Berücksichtigung fachlicher Standards von dafür qualifizierten Personen durchgeführt werden (vgl. Kromrey 2001: 106), andernfalls besteht die erhebliche Gefahr, dass „z. B. Sachverhalte einseitig oder gar ‚falsch‘ dargestellt, bestimmte Stakeholderinteressen über- oder unterbewertet oder keine einheitlichen Kriterien verwendet werden“ (Stockmann 2007: 27) sind. Entscheidend ist auch, dass die ‚richtigen‘ Fragen gestellt werden und die eingesetzten Designs oder Erhebungsmethoden zu den Fragestellungen passen.

„Wissenschaftlich oder professionell durchgeführte Evaluationen zeichnen sich dadurch aus, dass

- sie auf einen klar definierten Gegenstand (z.B. politische Interventionsmaßnahmen, Projekte, Programme, Organisationen, Personen, Prozesse, Produkte etc.) bezogen sind,
- sie von dafür besonders befähigten Personen (Experten) durchgeführt werden,
- die Bewertung anhand explizit auf den zu evaluierenden Sachverhalt und anhand präzise festgelegter und offengelegter Kriterien vorgenommen wird und
- für die Informationsgenerierung objektivierende empirische Datenerhebungsmethoden sowie
- für die Informationsbewertung systematisch vergleichende Verfahren herangezogen werden.“ (Stockmann 2007: 27)

Merkmale wissenschaftlicher Evaluationen

Nur professionelle Evaluationen sollten als Basis für Entscheidungen im öffentlichen Bereich herangezogen werden, wenn es z.B. um Förderentscheidungen, Programmverbesserungen oder gar die Selektion und Priorisierung von Maßnahmen geht.

Evaluation ist sonach ein Teil der angewandten Sozialforschung. Sie will zur Lösung praktischer, gesellschaftspolitischer Probleme beitragen. In diesem Sinnen versucht sie, gezielt und systematisch Grundlagen bereitzustellen, die außerwissenschaftlichen Entscheidungsprozessen dienen (vgl. Clemens, Strübing 2000: 215). Sie bedient sich hierzu der gesamten Bandbreite sozialwissenschaftlicher Theorien, Konzepte und Forschungsmethoden unter Anwendung der in der Wissenschaft geltenden Regeln für das Sammeln valider und reliabler Daten (vgl. Rossi et al. 1988: 1ff.; Kromrey 1995: 314f.; Wottawa, Thierau 1998: 9f.; Bortz, Döring 2002: 3).

Anders als die Grundlagenforschung, hat Evaluationsforschung zumeist einen Auftraggeber, der bestimmte Ziele verfolgt. Evaluation muss sich deshalb an extern gesetzten Zielbestimmungen orientieren. Evaluationen richten sich daher an außerwissenschaftlichen Erkenntnisinteressen und Verwertungskontexten aus (vgl. Kromrey 2002: 96f.).

Unterschiede zu Grundlagenforschung

Ein weiterer wesentlicher Unterschied zur Grundlagenforschung besteht darin, dass Evaluationen immer mit Bewertungen verbunden sind. Die Bewertungskriterien werden meist aus dem zu

evaluierenden Programm abgeleitet. Dabei werden „empirisch die Zustände in den Zielbereichen des Programms vor und nach den ergriffenen Maßnahmen festgestellt und untersucht, welche Veränderungen auf welche Elemente des Programms zurückzuführen sind. Durch Vergleich der empirischen Fakten (Ist-Zustand) mit den im Programm formulierten Zielniveaus (Soll-Zustand) können anschließend deduktiv Aussagen darüber abgeleitet werden, ob und in welchen Teilen das durchgeführte Programm erfolgreich bzw. nicht erfolgreich war. Um zu einer abschließenden Bewertung zu kommen, sind allerdings auch nicht-intendierte Wirkungen zu berücksichtigen. Darüber hinaus können zudem andere oder weitere Bewertungskriterien ergänzend herangezogen werden, die z.B. von Auftraggebern oder den betroffenen Zielgruppen formuliert werden“ (Stockmann 2007: 28).

Hinsichtlich der Auswahl des Untersuchungsgegenstandes sowie der Verwendung von Datenerhebungs- und Analysemethoden zur Identifizierung von Wirkungen und Kausalitäten (Ursache-Wirkungszusammenhänge) sind hingegen keine prinzipiellen Unterschiede zwischen Evaluations- und Grundlagenforschung zu erkennen. (Stockmann 2007: 29)

Die Evaluationsforschung bewegt sich hiernach in einem Spannungsverhältnis zwischen Wissenschaftlichkeit und Nützlichkeit.

Die Orientierung auf die Bewertung und Verwertung ist eine essentielle Voraussetzung für die Nützlichkeit von Evaluationen. Sie sind daher bei aller Wissenschaftlichkeit auch fast immer an den Interessen und Informationsbedürfnissen ihrer Initiatoren, Auftraggeber oder sonst Betroffenen ausgerichtet (vgl. Kromrey 2001; Vedung 2000; Patton 1997; Shadish et al. 1991). Umstritten ist, in welchem Maße einer möglichst objektiven Erkenntnisgewinnung dienende wissenschaftlich-methodische Ansprüche zu Gunsten der Erzeugung praktischen Handlungswissens reduziert werden dürfen. Klar ist jedoch, dass in der Praxis sehr häufig auf Lösungen zurückgegriffen wird, die aus sozialwissenschaftlicher Sicht fragwürdig sind.

Als Teil der empirischen Sozialforschung einerseits und Teil des politischen Prozesses andererseits zeichnet sich die Evaluationsforschung durch eine besondere Dualität aus. Aufgrund dieser Tatsache haben sich unterschiedliche theoretisch-methodische Ansätze herausgebildet, die sich entweder mehr an wissenschaftlichen Standards oder stärker an den Anforderungen der Auftraggeber oder den Bedürfnissen der Zielgruppen orientieren (siehe Abschnitt **Error! Reference source not found.**).

1.1.1 Evaluationskonzepte

Die Vielzahl ökonomischer, sozialer und ökologischer Problemstellungen die zukünftig zu bewältigen sein werden, stellt national wie international eine große Herausforderung dar. Die zu ihrer Lösung entwickelten (politischen) Strategien sind unterschiedlich konkret ausformuliert. Gemein ist ihnen aber, dass sie als wünschenswert definierte Ziele postulieren (vgl. Bussmann et al. 1997: 66f., 83), zu deren Verwirklichung es detaillierter Umsetzungspläne in Form von aufeinander abgestimmten Programmen, Projekten und einzelnen Maßnahmen bedarf.

Programme folgen dabei einem Phasenmuster, das dem Lebensverlauf von Individuen gleicht. Deshalb können die Konzepte der Lebensverlaufsforchung verwendet werden, um die Programmentwicklung deutlich zu machen (vertiefend hierzu Stockmann 2006: 107 ff.).

Programmentwicklung

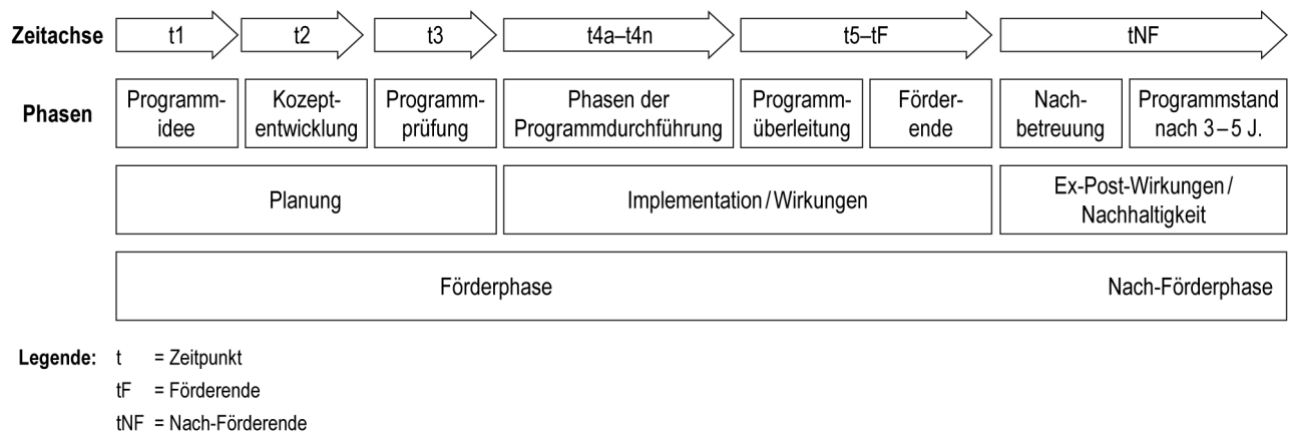


Abb. 1.1: Lebensverlaufsmodell (Stockmann 2007: 32)

Wichtig dabei ist zu verstehen, dass Programme nicht unabhängig von anderen, existierenden oder geplanten Programmen existieren. Sie sind immer auch im Kontext vorhandener gesellschaftlicher, institutioneller und organisatorischer Rahmenbedingungen zu betrachten. Soziale und regionale Kontexte sind ebenso zu berücksichtigen wie die teilweise ganz unterschiedlichen Interessen der verschiedenen Akteure. Programme müssen sich daher ökonomischen, sozialen, politischen, rechtlichen und kulturellen Veränderungen anpassen.

Insgesamt lässt sich die Entwicklung von Programmen aufgrund der Komplexität der sie umgebenden Umwelt nicht immer vorhersagen. Dennoch werden sie geplant und so gut wie möglich gesteuert, um die Programmziele möglichst in den vorgegebenen Fristen zu erreichen.

Der Lebensverlauf eines Programms lässt sich grob in drei Hauptphasen zerlegen (siehe **Error! Reference source not found..1**): In die **Planungsphase**, die **Implementationsphase** sowie die **Nachförder- oder auch Nachhaltigkeitsphase**. Am Beginn eines Programms steht die Programmidee (t1), sodann folgen die Ausarbeitung von Konzepten (t2) sowie die verschiedenen Lebensphasen eines Programms wie etwa die Programmprüfung (t3), die einzelnen Phasen der Durchführung (t4a-t4n) gefolgt von der Vorbereitung des Förderendes (t5-tF) sowie der Zeit nach der Programmförderung (tNF). Die einzelnen Phasen lassen sich durch eine Vielzahl prozessproduzierter Daten voneinander abgrenzen und sind daher gut analysierbar.

Evaluationen können sich mit verschiedenen Konzepten auf die einzelnen Phasen eines Programms beziehen, unterschiedliche Analyseperspektiven einnehmen und diversen Erkenntnisinteressen dienen (siehe **Tabelle 1.1**).

Tabelle 1.1: Dimensionen der Evaluationsforschung (Stockmann 2007: 34)

Phasen des Programmprozesses	Analyseperspektive	Erkenntnisinteresse	Evaluationskonzepte
Programmformulierung/Planungsphase	ex-ante	„analysis for policy“ „Science for action“	preformativ/formativ: aktiv gestaltend, prozessorientiert, konstruktiv
Implementationsphase	on-going	beides möglich	formativ/summativ
Wirkungsphase	ex-post	„analysis of policy“ „science for knowledge“	summativ: zusammenfassend, bilanzierend, ergebnisorientiert

Ex-ante-Evaluationen, auch „input-“ oder „preformative evaluations“ (Scriven 1991: 169) genannt, sind auf die Phase der Programmentwicklung, einschließlich ihrer Konzeptualisierung und

Planung gerichtet. Ihr Fokus liegt auf der Aufgabe, „die materiellen, personellen, institutionellen, finanziellen, theoretischen Rahmenbedingungen eines Programms“ (Brandstädter 1990: 217) zu untersuchen, um einen Beitrag zur Erstellung des Programmdesigns zu leisten (vgl. Brandstädter 1990: 217). Negative Effekte eines Programms sollen möglichst frühzeitig erkannt und Nachhaltigkeitschancen abgeschätzt werden.

Formative Evaluationen, auch „on-going-evaluations“ (Scriven 1991: 169) oder „Begleitforschung“ (Rossi et al. 1988: 11) unterstützen in der Implementationsphase vor allem das Programmmanagement bei der Steuerung. Indem Informationen über den Programmverlauf und die Programmergebnisse gesammelt, systematisiert und bewertet werden, sollen Entscheidungshilfen für die Durchführung gegeben und dadurch schon frühzeitige Korrekturen im Programmdesign ermöglicht werden (vgl. Rossi et al. 1988: 12, 31, 63; Wottawa, Thierau 1990: 54).

Summative Evaluationen, auch Ex-post evaluations genannt, werden nach Abschluss der Implementation eines Programms durchgeführt. Ihnen kommt die Aufgabe zu, den Umfang der Wirkungen eines Programms zu erfassen und zu bewerten sowie Zusammenhänge aufzudecken und die Frage nach dem Ursache-Wirkungs-Zusammenhang (Kausalität) genau zu untersuchen (vgl. Scriven 1991: 340). Darüber hinaus dienen sie dazu, die Nachhaltigkeit von Projekten und Programmen zu untersuchen.

1.1.2 Nutzen von Evaluationen

Der Nutzen von Projekt- und Programmevaluation lässt sich folgendermaßen kategorisieren (vgl. Stockmann 2007: 35f.):

- 1) Evaluationen können präformativ die Voraussetzungen für eine Programmdurchführung überprüfen und anschließend (formativ) die Ablaufprozesse beobachten. Es geht um die Identifikation von Problemen bei der Implementation eines Programms sowie um die Frage nach der Einhaltung geplanter Zeitabläufe. In diesem Zusammenhang geht es immer auch darum zu überprüfen, ob die Maßnahmen die Akzeptanz der Stakeholder finden, ob Interessenkonflikte auftreten, ob ausreichend qualifiziertes Personal für die Durchführung von Maßnahmen zur Verfügung steht, ob die Kommunikation und Koordination der ausführenden Stellen untereinander und mit den Zielgruppen des Programms funktioniert, ob die technische und finanzielle Ausstattung ausreichend ist, ob die mit dem Programm eingeführten Innovationen zielführend sind etc.
- 2) Eine wichtige Aufgabe von Evaluationen ist die Überprüfung der Zielerreichung. Dies geschieht in der Regel anhand eines Soll-Ist-Vergleichs, wobei die Sollwerte im Rahmen der Planung festgelegt wurden. Insoweit orientiert sich Evaluation immer strikt an den angestrebten Zielen. Dies kann zu einer Reihe von Problemen führen, auf die in Abschnitt **Error! Reference source not found.** näher eingegangen wird.
- 3) In der Regel erschöpfen sich Evaluationen nicht in einem Soll-Ist-Vergleich. Von Interesse ist darüber hinaus auch die Erfassung möglichst vieler (idealerweise aller) Wirkungen, die durch ein Programm bzw. eine Interventionsmaßnahme ausgelöst wurden. Dabei geht es vor allem darum, auch die nicht-intendierten Effekte zu erfassen, die die Zielerreichung unterstützen oder ihr zuwider laufen können, denn nur so kann erkannt werden, ob die positiven oder negativen Effekte eines Programms überwiegen.
- 4) Neben der reinen Prozessbetrachtung stellen Evaluationen auch die Programm- oder Maßnahmenziele selbst in Frage. Es gilt die Frage zu beantworten, ob ein Programm überhaupt dazu geeignet ist, relevante Entwicklungs- oder Innovationsleistungen zu erbringen oder ob Alternativen bestehen.

- 5) Schlussendlich ist von zentraler Bedeutung, ob die Wirkungen – intendiert oder nicht - überhaupt dem Programm zuzuordnen sind oder ob sie auf externen Faktoren beruhen. Die Lösung des Kausalitätsproblems ist eine der schwierigsten Aufgaben von Evaluation.

1.1.3 Ziele der Evaluation

Die Evaluationsziele sind eng mit den Aufgabenstellungen verknüpft. Gleichwohl macht eine Unterscheidung Sinn, da je nach Zielstellung unterschiedliche Herangehensweisen und Konzepte eingesetzt werden.

Evaluation hat vier wesentliche Ziele:

1) Gewinnung von Erkenntnissen

Ziel ist die Gewinnung von Erkenntnissen, um diese anhand vereinbarter oder im Programm bereits vorgegebener Bewertungskriterien zu beurteilen und daraus Steuerungsentscheidungen abzuleiten (vgl. Kromrey 2001: 114).

2) Ausübung von Kontrolle

Nicht immer steht bei der Verwertung der Erkenntnisse die Entscheidungsfindung im Vordergrund. Die Ausübung von Kontrolle ist ein weiteres wichtiges Ziel von Evaluation. Hier geht es vor allem um die Feststellung der Erreichung der in der Planung festgelegten Ziele. Kriterien wie Effektivität, Effizienz, Akzeptanz oder Nachhaltigkeit können hier herangezogen werden. Kontrollevvaluationen stellen eine Form administrativen Handelns dar (vgl. Kromrey 2001: 115).

3) Schaffung von Transparenz und Dialogmöglichkeiten

Durch die Offenlegung von Evaluationsergebnissen entsteht Transparenz hinsichtlich des Programmerfolgs und die Möglichkeit eines Dialogs zwischen verschiedenen Stakeholdern (Mittelgeber, Durchführungsorganisationen, Zielgruppen, sonstige Beteiligte und Betroffene). Im Vordergrund dieser Evaluationsfunktion stehen Lernprozesse, die für die Weiterentwicklung von Programmen genutzt werden sollen. Sie spielt bei formativen (programmgestaltenden) Evaluationen eine zentrale Rolle.

4) Legitimation der durchgeführten Maßnahmen

Zu guter Letzt dient Evaluation auch der Legitimation durchgeführter Programme und Maßnahmen. Mithilfe der durch Evaluation gewonnenen Daten kann nachprüfbar belegt werden, welcher Input zu welchem Output führte und welche Wirkungen über die Zeit hinweg erzielt werden konnten. Mit Ex-post-Evaluationen lassen sich Effizienz und Nachhaltigkeit von Programmen belegen.

1.2 Evaluationsansätze

Die Evaluationsforschung stellt eine Vielfalt von Ansätzen und Modellen bereit. Es handelt sich aber dabei nicht um Theorien im wissenschaftlichen Sinne. Es handelt sich mehr oder weniger um ausformulierte Vorschläge zur praktischen Strukturierung von Evaluationsvorhaben.

Diese Vielfalt wird teilweise beklagt und eine Synthese bestehender Ansätze gefordert (vgl. Gephardt 1978), andere hingegen glauben, dass dieses Bestreben sich kontraproduktiv auswirken würde, weil es die notwendige Entwicklung von Ansätzen in der Evaluationsforschung behindere (vgl. Raizen, Rossi 1981; Fitzpatrick et al. 2004).

Anwender von Evaluationen stehen somit vor dem Problem der Auswahl des geeigneten Ansatzes. Nicht selten endet dieses Dilemma mit der Entwicklung eines eigenen Ansatzes. Die Funktion von Ansätzen und Modellen wird dadurch weitgehend konterkariert.

Gründe für die **Vielfalt der Evaluationsansätze** gibt es viele. Zwei scheinen aber von besonderer Relevanz zu sein: (1) Der aus Weltanschauungen und philosophischen Grundeinstellungen abgeleitete wissenschaftstheoretische Standpunkt eines Autors sowie dessen aus persönlichen Erfahrungen und Einschätzungen gewonnenen Überzeugungen. (2) Die immer wieder neu formulierten Evaluationsfragen, die mit ihnen verfolgten Ziele und die jeweils unterschiedlichen Evaluationsgegenstände.

Die zunehmende Komplexität von Programmen, die Vielfalt der mit ihnen verbundenen Zielerwartungen, die Forderung nach stärkerer Einbeziehung der Stakeholder etc. stellt die Evaluation vor neue Herausforderungen. Ein Ordnungsversuch bestehender Ansätze auf zweierlei Weise erscheint sinnvoll: Auf der Basis

- (1) der wissenschaftstheoretisch-methodologischen Ausrichtung der Ansätze und
- (2) der grundlegenden Aufgabenstellung von Evaluationen.

Beide Aspekte sind häufig auch miteinander verknüpft.

Bei einer wissenschaftstheoretisch-methodologischen Systematisierung (siehe **Tabelle 1.2**) ist zu beobachten, dass sich die Evaluationsforschung nicht losgelöst von der allgemeinen wissenschaftlichen Forschung entwickelt hat, so dass sich in ihr auch die dort geführten Debatten mit nicht geringerer Heftigkeit wiederfinden.

Kriterien zur Unterscheidung von Evaluationsansätzen

Tabelle 1.2: Wesentliche Merkmale evaluationstheoretischer Ansätze (Lee 2006: 153; Übersetzung und Darstellung durch Niehoff)

(Post-)Positivistischer Ansatz	Interpretativer/ konstruktivistischer Ansatz	Transformativer/emanzipativer Ansatz
Es gibt eine einzige Realität, die objektiv erfassbar ist, indem alternative Erklärungen sukzessive ausgeschlossen werden.	Realität ist sozial aus verschiedenen Perspektiven konstruiert, die miteinander in Konflikt stehen können.	Es gibt mehrere Realitäten, die jeweils unterschiedliche Einflüsse auf den sozialen, politischen und kulturellen Kontext von Programmen und ihren Evaluationen haben können. Unterschiede bei der Programmzielerreichung und den -wirkungen werden insb. mit Faktoren wie ökonomischer Status, Ethnizität, Geschlecht, Behinderung etc. verknüpft.
Objektivität und Neutralität stellen die einzigen adäquaten Standards für Forschende dar	Es gibt keine „objektive“ Realität. Der Beobachter beeinflusst diejenigen, die er untersucht und wird umgekehrt von diesen beeinflusst. Forschende müssen sich dieser Tatsache und der daraus resultierenden Probleme bewusst sein.	Forschende sollten mit den Beteiligten und Betroffenen in einer Art und Weise interagieren, die die Benachteiligten berücksichtigt („empowers the observed“), ohne die Standards der Evaluation zu beeinträchtigen.
Die Verwendung wissenschaftlicher Methoden dient vor allem dem Zweck Ursache-Wirkungs-zusammenhänge aufzudecken, um die Generalisierung von Ergebnissen zu ermöglichen.	Die verwendeten Methoden sind qualitativ und zielen auf das Erklären/Verstehen der jeweils untersuchten Gegenstände im jeweiligen konkreten Kontext ab.	Der Schwerpunkt des Methodeneinsatzes liegt auf der Verwendung verschiedener, die unterschiedlichen Perspektiven verdeutlichender Methoden und mehr auf der Nebeneinanderstellung von Positionen denn einer Verallgemeinerung von Ergebnissen.

Trotz unterschiedlicher Sichtweisen herrscht weitgehend Konsens darüber, dass Evaluationen die **Perspektiven** und Bedürfnisse **der Stakeholder berücksichtigen** müssen und dass **quantitative und qualitative Methoden** oft gut miteinander **kombiniert werden können** (Multimethodenansatz; Stockmann 2007: 46).

Eine heute weit verbreitete Auffassung ist, „daß eine Untersuchung gleichzeitig strengen wissenschaftlichen Anforderungen genügen und für den Auftraggeber und andere Interessengruppen von maximalem Nutzen sein kann“ (Rossi et al. 1988: 10). Chelimsky (1995: 6) beschreibt die neuere Entwicklung so: „We think less today about the absolute merits of one method versus another, and more about whether and how using them in concert could result in more conclusive findings. [...] We have learned that the choice of methods (and measures and instruments and data) depends much more on the type of question being asked than on the qualities of any particular method“.

Einigkeit besteht auch darin, dass Evaluationsergebnisse vor allem im politischen Prozess genutzt werden sollten (vgl. Chelimsky 1995: 8). Damit rückt das Kriterium der Nützlichkeit von Evaluationsergebnissen in den Vordergrund, welches seit Mitte der 80er Jahre ein Schwerpunkt in der Evaluationsdiskussion ist (vgl. Patton 1997, Alkin 1990).

Die Nützlichkeit von Evaluationsergebnissen, also die Frage, welche grundlegende Aufgabenstellung Evaluationen verfolgen, kann ebenfalls als Unterscheidungsdimension für die

Minimalkonsens evaluationstheoretischer Ansätze

Strukturierung von Evaluationsansätzen genutzt werden. Praktische Erwägungen im Hinblick auf die jeweils unterschiedlichen Anforderungen an Evaluationen stehen dabei im Vordergrund.

Fitzpatrick et al. (2004: 68f.) klassifizieren die evaluationstheoretischen Ansätze in fünf Kategorien:

Klassifizierung

- 1) **Zielorientierte Ansätze**, fokussieren auf die Identifizierung von Programmzielen sowie die Beurteilung der Zielerreichung
- 2) **Managementorientierte Ansätze**, sind auf die Identifizierung und Erfüllung der Informationsbedürfnisse von Entscheidungsträgern gerichtet
- 3) **Konsumentenorientierte Ansätze**, sollen produktbezogene Informationen und Bewertungen z.B. anhand von Produkt-Checklisten zur Verfügung stellen
- 4) **Expertenorientierte Ansätze**, bei denen eine Evaluation durch ausgesuchte Experten eines Praxisfeldes erfolgt
- 5) **Partizipative Ansätze**, sind vor allem durch die starke Einbeziehung der verschiedenen an einer Evaluation beteiligten oder von ihr betroffenen Interessengruppen (Stakeholder) bei der Evaluationsplanung und -durchführung charakterisiert.

Auch wenn die Vielfalt einzelner Ansätze keinen Werkzeugkoffer darstellt, aus dem der passende Ansatz für ein bestimmtes Evaluationsproblem ausgewählt werden kann, so bietet sie doch ein umfassendes Einsatzspektrum. Dieses Potenzial lässt sich nutzbar machen, um für spezifische Evaluationsaufgaben die passenden Kombinationen auszuwählen.

1.2.1 Der CEval-Evaluationsansatz

Der CEval-Evaluationsansatz, der am ehesten zu den managementorientierten Ansätzen gezählt werden kann, kann für zahlreiche Zwecke eingesetzt werden und lässt sich zur Aufklärung der Öffentlichkeit ebenso verwenden wie für die Befriedigung der Steuerungsbedürfnisse des Managements. Der CEval-Ansatz ist durch seinen wirkungsorientierten Fokus geprägt und grenzt sich deutlich von rein zielorientierten Ansätzen ab.

Das CEval-Konzept vereint Elemente aus verschiedenen Evaluationsansätzen zu einem wirkungsorientierten, theoriebasierten Ansatz, der multifunktionell einsetzbar ist. Das zentrale Anliegen ist die möglichst vollständige Erfassung der von Programminterventionen oder Maßnahmen ausgehenden Wirkungen und deren Ursachen. Hierzu ist es erforderlich, die Maßnahmen- bzw. Programmeffekte von anderen Einflüssen zu trennen. Die theoretisch-konzeptionellen Überlegungen des von Stockmann (1992, 1996, 2006) entwickelten CEval-Ansatzes gehen davon aus, dass Projekte und Programme regelhaft von Organisationen durchgeführt werden. Organisationstheoretische Ansätze für die Analyse von Zusammenhängen und die Entwicklung von untersuchungsleitenden Fragestellungen bieten sich daher besonders an. Des Weiteren lassen sich auch innovations- und diffusionstheoretische Ansätze verwenden, da Programme häufig Instrumente zur Einführung von Innovationen darstellen (vgl. Stockmann 2006).

1.2.1.1 Organisationsmodell

Programminnovationen können darauf abzielen, sowohl intern Veränderungen bei den programmdurchführenden Organisationen auszulösen, als auch extern in anderen sozialen Systemen. Je umfangreicher die durch ein Programm eingeführten Innovationen von anderen übernommen werden, desto erfolgreicher ist das Programm.

Bei Wirkungsevaluationen kommt den Beziehungen von Organisationen, die Programme durchführen, zu anderen Organisationen oder gesellschaftlichen Subsystemen eine besondere Bedeutung zu (siehe **Abb. 1.2**). Im Zentrum des Modells steht das in eine Trägerorganisation eingebettete Programm. Die Wirkungsmöglichkeiten des Programms werden durch die Trägerorganisation sowie durch die sie umgebenden Systeme beeinflusst. Dabei können die externen Umweltbereiche die Zielsetzungen des Programms sowohl unterstützen, als auch behindern.

Organisationen als offene soziale Systeme

Der CEval-Ansatz rekurriert auf das organisationstheoretische Erklärungsmodell, wonach Organisationen als offene soziale Systeme begriffen werden, die der Intention nach rational gestaltet sind, um spezifische Ziele zu erreichen (vgl. Thompson 1967: 66 ff.; Kieser, Kubicek 1992: 4 ff.; Kieser 2002: 169 ff.; Kieser, Walgenbach 2003: 6 ff.; Müller-Jentsch 2003: 20 ff.; Scott 2003: 33 ff., 82 ff., 141 ff.).

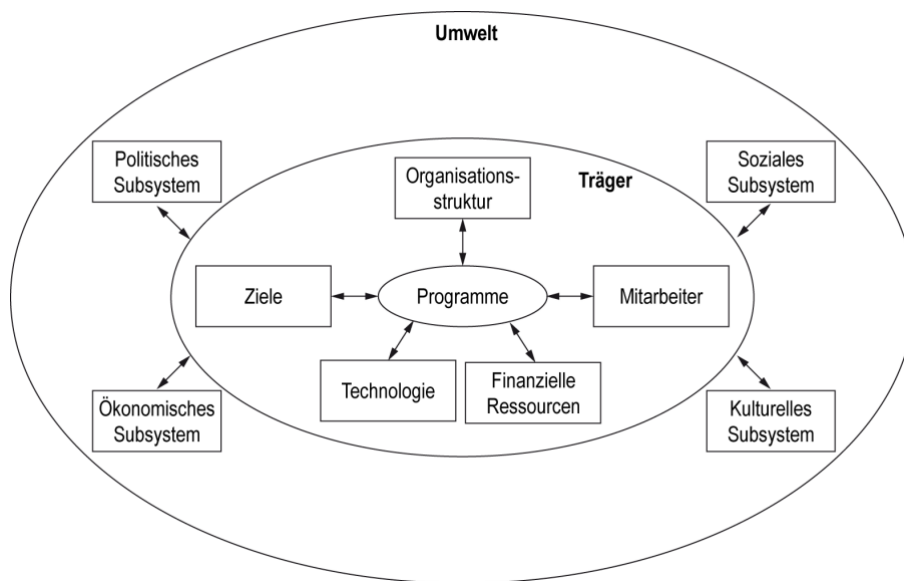


Abb. 1.2: Wirkungsmodell (Stockmann 2007: 52)

Anders als Strukturen und Prozesse oder Technologien einer Organisation werden deren finanzielle Ressourcen in den gängigen soziologischen Organisationskonzepten nicht berücksichtigt. Dieser für die Aufgabenerfüllung zentrale Aspekt wird beim CEval-Ansatz in die Analyse einbezogen. Die Elemente, die eine Organisation konstituieren, lassen sich damit wie folgt ableiten: Die **Ziele**, die **Mitglieder** (Beteiligte), die **formale Struktur** (Organisationsstruktur), die **Technologie** und die **finanziellen Ressourcen** einer Organisation.

Elemente einer Organisation

Das Wirkungsmodell lässt verschiedene kausale Betrachtungsweisen zu. Zwei Analyseperspektiven können nacheinander eingenommen werden:

In einem ersten Schritt werden die Programminterventionen als unabhängige Variablen (UV) und die Organisationselemente als abhängige Variablen (AV) betrachtet. Ziel ist es zu prüfen, ob die Interventionen (Inputs) Veränderungen bewirkt haben. Wenn die einzelnen Organisationselemente durch die Programminterventionen effektiv gestaltet werden konnten, stellt dieses Ergebnis gleichzeitig einen internen, auf die Durchführungsorganisation bezogenen Programmoutput dar (vgl. Stockmann 2007: 53).

Sodann werden in einem zweiten Schritt die internen Programmoutputs zu unabhängigen Variablen, die auf Veränderungen in der Trägerumwelt zielen. Diese externen Bereiche (z.B. das Beschäftigungs- oder Bildungssystem, das ökologische System, das Rechtssystem) werden nun zu abhängigen Variablen. Mithilfe von Indikatoren können nun die Diffusionswirkungen der

Durchführungsorganisation in den externen Bereichen, als Maßstab für die Programmeffektivität, gemessen werden.

1.2.1.2 Innovations-/Diffusionsmodell

Die Diffusionsforschung beschäftigt sich mit der Frage, unter welchen Bedingungen eine Verbreitung von Innovationen stattfindet. Rogers (1995: 5) definiert Diffusion als „the process by which an innovation is communicated through certain channels over time among the members of a social system“. Der Diffusionsprozess kann durch verschieden Faktoren dabei positiv wie negativ beeinflusst werden. Mohr (1977: 19ff.) unterscheidet hier vier Gruppen von Variablen:

1. Spezifischen Eigenschaften der jeweiligen Innovation selbst;
2. Umweltvariablen;
3. Personen, die eine Innovationsidee aufgreifen, über ihre Einführung entscheiden und gegebenenfalls ihre Realisierung durchsetzen;
4. Elemente der formalen Organisationsstruktur, die Innovationen einführt.

Variablen, die den Diffusionsprozess beeinflussen

Nach der von CEval verwendeten Organisationskonzeption werden Personen, die an einer Innovation mitwirken, als Mitglieder von Organisationen behandelt, so dass hier nur drei Variablengruppen unterschieden werden (siehe *Error! Reference source not found. 1.3*):

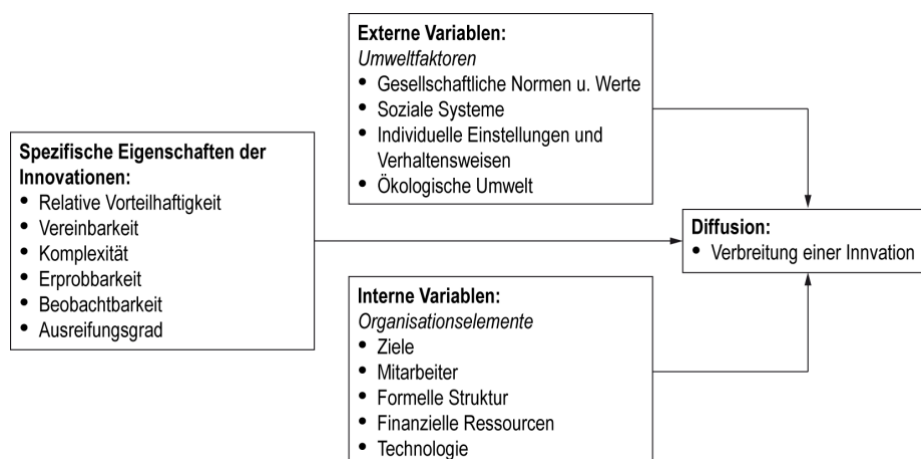


Abb. 1.3: Diffusionsmodell (in Anlehnung an Mohr 1977: 43)

Die organisations- und innovations-/diffusionstheoretischen Überlegungen ermöglichen die Entwicklung eines Evaluationsleitfadens (siehe **Tabelle 1.3**) für die Evaluation von Programmen. Dabei definieren die verwendeten Variablen die einzelnen Untersuchungsfelder. Die **internen Wirkungsfelder** (der Veränderungen in der Trägerorganisation des Programms) werden mithilfe der identifizierten Organisationsparameter (vgl. *Error! Reference source not found. 1.2*) untersucht. Für die **externen Wirkungsfelder** werden die Veränderungen in den Zielsystemen (Politikfeldern) und bei den Zielgruppen (z.B. Menschen oder auch andere Organisationen) untersucht, bei denen Veränderungen bewirkt werden sollten. Aus der Lebensverlaufsperspektive folgt zudem, dass der Planungs- und Implementationsprozess nur bewertet werden kann, wenn die einzelnen Phasen eines Programms betrachtet werden. Dem analytischen Teil ist ein Abschnitt vorangestellt, in dem das zu evaluierende Programm sowie dessen Kontextbedingungen beschrieben werden (vgl. Stockmann 2006: 126ff.).

Tabelle 1.3: Leitfaden für die Programmevaluation (Stockmann 2007: 56)

1. Programm und Umwelt

1.1 Programmbeschreibung (u.a. Programmdaten und -konzeption, Innovationskonzeption, Ressourcen)
1.2 Umwelt-/Kontextbedingungen (u.a. Beschreibung des Praxis-/Politikfelds bzw. des gesellschaftlichen Subsystems, Zielgruppen)
2. Programmverlauf
2.1 Planung
2.2 Steuerung
2.3 Förderende
2.4 Nachbetreuung
3. Interne Wirkungsfelder
3.1 Ziele/Zielakzeptanz bei der Durchführungsorganisation und/oder politisch übergeordneten Organisationen (z.B. Geldgeber)
3.2 Personal (insb. Qualifikation)
3.3 Organisationsstruktur (insb. Funktionalität und Funktionsfähigkeit)
3.4 Finanzielle Ressourcenverfügbarkeit
3.5 Technologie: Technische Infrastruktur (insb. Ausstattung)
3.6 Technologie: Organisationsprogramm/-konzeption
4. Externe Wirkungsfelder
4.1 Zielakzeptanz bei den Zielgruppen
4.2 Zielgruppenerreichung
4.3 Nutzen für die Zielgruppen
4.4 Zielgruppenübergreifende Wirkungen
4.5 Wirkungen im Politikfeld des Programms
4.6 Politikfeldübergreifende Wirkungen

1.2.2 Evaluation von Nachhaltigkeit

Der CEval-Evaluationsansatz wird ergänzt durch ein Nachhaltigkeitsmodell, das zwischen der Nachhaltigkeit auf Makroebene und der auf Programmebene unterscheidet.

Nachhaltigkeitsmodell

Auf der Makroebene knüpft der Evaluationsansatz an das Konzept der „**nachhaltigen Entwicklung**“ (sustainable development) an. Dieses geht davon aus, dass ökonomische, soziale und ökologische Entwicklungsprozesse untrennbar miteinander verbunden sind und so ausbalanciert werden müssen, dass die natürlichen Ressourcen nachfolgenden Generationen erhalten bleiben. Es existieren unterschiedlichste Konzepte, mit denen diese drei Zielgrößen auf sehr verschiedene Weise operationalisiert werden. **Tabelle 1.4** stellt eine vereinfachte und leicht zu handhabende Operationalisierung dar, die sich vor allem für die Evaluation von Programmen eignet (vgl. Stockmann 2006: 134ff., 157ff.).

Tabelle 1.4: Dimensionen der Nachhaltigkeit auf der Makroebene (Stockmann 2007: 57)

Zielgröße	Operationalisierung	Kriterium erfüllt wenn, ...
ökonomisch	Effizienz	mit einem möglichst geringen Mitteleinsatz (Input) ein Optimum an Leistungen (Output) und möglichst alle intendierten Wirkungen (Outcome) erzielt werden.
sozial	Gesellschaftspolitische Relevanz	die durch die erbrachten Leistungen entstandenen (intendierten wie nicht-intendierten) Wirkungen insgesamt als gesellschaftspolitisch relevant und nützlich eingestuft werden.

ökologisch	Umweltverträglichkeit	mit den Ressourcen zur Leistungserstellung umweltschonend umgegangen wird und wenn die erbrachten Leistungen und die daraus entstandenen (intendierten wie nicht-intendierten) Wirkungen umweltverträglich sind.
------------	-----------------------	--

Die bewusst abstrakt gehaltenen Kriterien bedürfen im Einzelfall einer weitergehenden Konkretisierung, um die Sammlung messbarer (quantitativer und/oder qualitativer) Daten zu ermöglichen.

Nachhaltigkeitsmodelle bedürfen politischer Strategien und Programme. Dabei stellt sich die Frage, ob Maßnahmen über die Phase der finanziellen Förderung hinaus Wirkung entfalten, ob also dauerhaft Strukturen geschaffen und Verhaltensänderungen herbeigeführt werden können. Nur in diesem Fall ist Nachhaltigkeit gegeben. Die Nachhaltigkeit von Programmen wird anhand von vier Dimensionen bestimmt (**Tabelle 1.5**):

Tabelle 1.5: Dimensionen der Nachhaltigkeit auf Programmebene (Stockmann 2007: 58)

Dimension	Typ	Merkmal
I	projekt-/programmorientiert	Die Zielgruppe und/oder Trägerorganisation führt die Innovationen im eigenen Interesse und zum eigenen Nutzen fort.
II	output-/leistungsorientiert	Andere Gruppen/Organisationen haben die Innovationen in ihrem eigenen Interesse und zum eigenen Nutzen dauerhaft übernommen.
III	systemorientiert	Die Innovationen führen über Diffusionsprozesse zu einer Leistungssteigerung des gesamten Systems (z.B. des Gesundheits- oder Bildungssystems).
IV	innovationsorientiert	Die Zielgruppe/die Trägerorganisation verfügt über ein Innovationspotential, um auf veränderte Umweltbedingungen flexibel und angemessen zu reagieren.

1.3 Evaluationsdurchführung

Da der organisatorische Ablauf einer Evaluation hier aus Platzgründen nicht näher dargestellt werden kann, beschränken sich die Ausführungen auf einige grundlegende Begriffe und Sachverhalte, die für das Verständnis wichtig sind.

1.3.1 Intern oder extern?

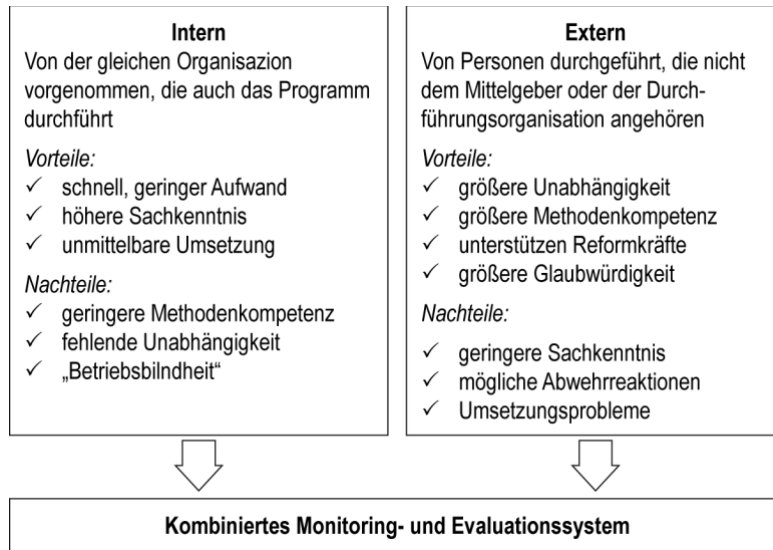
Evaluationen können auf zweierlei Weise durchgeführt werden, als interne oder externe Evaluation. Sie ist intern, wenn sie durch die programmdurchführende Organisation selbst vorgenommen wird. Von einer Selbstevaluation spricht man, wenn diese interne Evaluation von der Abteilung durchgeführt wird, die gleichzeitig mit der operativen Durchführung des Programms betraut ist. Zu den Vor- und Nachteilen interner Evaluationen siehe **Tabelle 1.6**.

Externe Evaluationen hingegen werden von Personen durchgeführt, die nicht dem Fördermittelgeber oder der Durchführungsorganisation angehören (Stockmann 2007: 62). Neben ihrer größeren Unabhängigkeit verfügen externe Evaluatoren meist über große Methodenkompetenz und breites Evaluationswissen und kennen das Fachgebiet, dem das Programm bzw. das Projekt zuzuordnen ist. Externe Evaluationen können Reformkräften

innerhalb einer Organisation zudem zusätzliche Legitimität verleihen, die benötigt wird, um Veränderungsprozesse auszulösen (vgl. Pollitt 2000: 72).

Externe Evaluationen können aber umgekehrt auch Angstgefühle bei den Evaluierten auslösen und zu Abwehrreaktionen führen, was sich auf die spätere Umsetzung von Evaluationsergebnissen negativ auswirken kann.

Tabelle 1.6: Interne und externe Evaluation (Stockmann 2007: 61)



Im Hinblick auf die Evaluationsziele ist zu beobachten, dass Evaluationen mit Fokus auf die Entwicklungsfunktion (Lernen) häufig intern durchgeführt werden, wohingegen Erkenntnis- und kontrollorientierte Evaluationen sowohl intern als auch extern durchgeführt werden. Evaluationen, die primär der Legitimation dienen, werden fast ausschließlich extern in Auftrag gegeben.

Eng mit dem Instrument der Evaluation verbunden ist das Monitoring. Das Monitoring kann auch als eine **Form der internen Evaluation** betrachtet werden, wobei allerdings nicht jede interne Evaluation eine Form des Monitoring darstellt. Monitoring kann auf der Ebene des **Gesamtsystems**, eines **Politikfeldes**, eines **Programms**, eines **Projekts** oder einzelner **Interventionsmaßnahmen** ansetzen. Es lassen sich **Input-, output- und Wirkungsdaten** erfassen. Ein bekanntes Beispiel für ein Monitoring-System auf Politikfeldebene ist das Umweltmonitoring, das Messdaten über den Zustand der Umwelt liefert. Auf der gesamtgesellschaftlichen Ebene informiert z.B. ein Sozialindikatorensystem über die Entwicklung der Lebenslagen in Deutschland und ergänzt dadurch die amtliche Statistik (vgl. Stockmann 2007: 63).

Auf Projekt- und Programmebene hat ein Monitoring-System die Aufgabe, das Management kontinuierlich mit Daten über den Verlauf und die Zielerreichung zu versorgen. Rossi et al. (2004: 231) definieren deshalb: „Program monitoring is a form of evaluation designed to describe how a program is operating and assess how well it performs its intended functions“. Anders als bei Evaluationen, die singular zu einem bestimmten Zeitpunkt durchgeführt werden, ist **Monitoring eine fortlaufende, routinemäßige Tätigkeit** mit dem Ziel, zu überwachen, ob die Planungsvorgaben und angestrebten Ziele möglichst effizient und unter Einhaltung der verfügbaren Ressourcen und der vorgegebenen Zeit erreicht werden (Stockmann 2007: 63).

Monitoring kontrolliert demnach den planmäßigen Vollzug. Dabei werden der Programm- bzw. der Projektplan und die ihm zu Grunde liegenden Entwicklungshypothesen nicht in Frage gestellt. Dies sowie die Analyse von Wirkungszusammenhängen ist Gegenstand und Aufgabe von Evaluationen.

Beim Monitoring spielt die kausale Zuordnung beobachteter Veränderungen eine untergeordnete Rolle. Monitoring ist eine weitgehend **deskriptive Tätigkeit**, mit der möglichst zuverlässig **Daten in periodischen Abständen gesammelt werden** sollen, so dass kontinuierlich Zeitreihen entstehen, die Entwicklungsverläufe erkennen lassen. Dies kann im Rahmen von Einzelevaluationen oft nur schwer oder gar nicht geleistet werden (Stockmann 2007: 63).

Monitoring

Ziel: Entwicklungsverläufe darstellen

1.3.2 Ziel- oder wirkungsorientiert?

Fast jeder Evaluationsauftrag dient auch der Überprüfung, ob die in einem Programm angestrebten Ziele erreichbar sind (präformativ/formativ) oder erreicht wurden (summativ). Dies geschieht in Form von Soll-Ist-Vergleichen. Zielorientierte Evaluationen weisen einige Probleme auf. Hervorgehoben werden kann, dass Programmziele häufig nicht klar formuliert werden oder sogar ganz fehlen. Neben offiziell deklarierten Zielen bestehen manchmal andere (informelle) Zielvorgaben, die sich als die eigentlich angestrebten herausstellen können. Zudem unterliegen Ziele Veränderungen über die Zeit hinweg und nicht alle Akteure verfolgen immer die gleichen Ziele.

Die Nichteinhaltung grundlegender Regeln bei der Formulierung der Programmziele führt deshalb zu erheblichen Evaluationsproblemen.

„Eine vollständige Zielformulierung muss folgende fünf Elemente beinhalten:

- **Zielobjekt:** Worauf bezieht sich das Ziel? (z.B. eine bestimmte Zielgruppe)
- **Zielinhalt:** Was soll erreicht werden? (z.B. die Umsetzung empfohlener Maßnahmen)
- **Zielwert:** Wie viel soll erreicht werden? (z.B. die Zahl der Umsetzungen)
- **Ortsbezug:** Wo soll das Ziel wirksam werden? (z.B. in einer bestimmten Region)
- **Zeitbezug:** Wann soll das Ziel erreicht werden? (z.B. der Zeitpunkt des Projektendes).“
(Stockmann 2007: 64)

Elemente einer Zielformulierung

Der Zielformulierungsprozess kann auch durch die Evaluation selbst beeinflusst werden. Wird in einem Programm davon ausgegangen, dass der Erfolg desselben vor allem an der Zielerreichung zu messen ist, dann besteht die große Versuchung, Ziele entweder schwammig zu formulieren oder die Zielerreichungsmargen sehr niedrig zu hängen. Bei Evaluationen, die auf reinen Soll-Ist-Vergleichen beruhen, besteht daher die Gefahr, dass sie allein der Vollzugsverkündung dienen.

Unabhängig davon besteht bei zielorientierten Evaluationen auch die Gefahr, dass nicht-intendierte Effekte, die im Hinblick auf die Ergebnisbewertung von besonderem Interesse sein können, systematisch ausgeblendet werden (Tunnelblick).

Wirkungsorientierte Evaluationen bieten sich hier als Alternative an. Diese gehen nicht primär von den Zielen eines Programms aus, sondern versuchen, hypothesengeleitet potentiellen Wirkungen auf die Spur zu kommen. Hierfür eignet sich z.B. der von CEval entwickelte und in Kapitel 1.2.1 vorgestellte Evaluationsansatz. Mit dem dargestellten Evaluationsleitfaden lässt sich die Suche nach intendierten und nicht-intendierten Wirkungen steuern und strukturieren.

Obwohl der Begriff **Wirkung** in der Evaluation von zentraler Bedeutung ist, ist nicht immer klar, worum es sich hierbei überhaupt handelt. Nicht verwechselt werden darf er mit dem Begriff „Output“. Outputs sind die Produkte oder Leistungen, die von einer Organisation erbracht werden, wie z.B. die Anzahl verteilter Essenspakete, die Anzahl behandelter Kranker, die Anzahl der durchgeführten Beratungen etc. Wirkungen stellen die Veränderungen dar, die Folgen dieser Leistungen sind, also z.B. gesättigte Menschen, verbesserter Gesundheitszustand der Behandelten oder Personen, die nach der Beratung wissen, was zu tun ist (vgl. Stockmann 2007: 65).

Da die Qualität eines Programms nur in seiner gesamten Komplexität bewertet werden kann, sind die nicht-intendierten Folgen bei der Bilanzierung eines Programms immer mit zu berücksichtigen. Während intendierte Wirkungen in der Regel im Hinblick auf die Zielerreichung positiv bewertet werden, können nicht-intendierte Wirkungen sowohl positiv als auch negativ ins Gewicht fallen.

Positive intendierte Wirkungen wären z.B. dann gegeben, wenn eine Finanzmittelerhöhung im Schulwesen, nicht nur dazu geführt hat, dass mehr Lehrer eingestellt wurden, sondern dass deshalb auch die Klassen kleiner wurden und die Schüler jetzt mehr lernen. Ein nicht erwünschter Effekt könnte darin bestehen, dass weniger qualifizierte Lehrkräfte eingestellt wurden, weil nicht genügend qualifizierte verfügbar sind und dadurch die Ausbildungsqualität in den Schulen eher sinkt und nicht wie intendiert zunimmt.

Fallbeispiel

Wirkungen können sich in der Veränderung von Strukturen, Prozessen oder individuellen Verhaltensweisen zeigen. Sie lassen sich analytisch auf drei Dimensionen bestimmen (Stockmann 2007: 66):

1. Dimension Struktur – Prozess – Verhalten

Wirkungen können Strukturen (z.B. von Organisationen oder gesellschaftlichen Teilsystemen), Prozesse und/oder individuelle Verhaltensweisen betreffen.

2. Dimension: Geplant – Ungeplant

Wirkungen können wie geplant (intendiert) oder ungeplant (nicht-intendiert) auftreten.

3. Dimension: Positiv – Negativ

Die geplant oder ungeplant auftretenden Wirkungen können die Programm- oder Leistungsziele unterstützen (+) oder ihnen zuwiderlaufen (–) (siehe **Tabelle 1.7**).

Wirkungsdimensionen

Tabelle 1.7: Wirkungsdimensionen (Stockmann 2007: 66)

Wirkungsdimension	Geplant	Ungeplant
Struktur	+/-	+/-
Prozess	+/-	+/-
Verhalten	+/-	+/-

Ziel von Wirkungsevaluationen ist es, mit größtmöglicher Zuverlässigkeit festzustellen, ob eine Intervention die intendierten Wirkungen auslöst. Daher sind Einflüsse anderer Faktoren, die ebenfalls für die gemessenen Veränderungen ursächlich sein könnten, auszuschließen. Dies gehört zu den größten Herausforderungen von Evaluationen. Dies liegt zum einen an der Komplexität der sozialen Umwelt, zum anderen daran, dass Interventionen in der Regel nur einen geringen Eingriffsspielraum und ein niedriges Veränderungspotenzial haben. Oft sind die Programm- oder Leistungswirkungen so schwach ausgeprägt, dass die Gefahr besteht, dass sie gar nicht wahrgenommen werden. Die von der Sozialwissenschaft entwickelten Erklärungsmodelle und Messmethoden für soziale Phänomene sind im Rahmen von Evaluationen oft nur unzureichend anwendbar (vgl. Meyer 2007: 195ff.). Standards

In vielen Berufs- und Arbeitsfeldern gibt es berufliche Standards. Sie definieren nicht nur grundlegende Qualitätsansprüche und ethische Richtlinien, sondern zielen auch darauf ab, Kunden und die Öffentlichkeit vor schädlichen Praktiken und unsachgemäßem Vorgehen zu schützen. Standards bieten darüber hinaus eine Kontroll- und Beurteilungsbasis bei potenziellen Streitfragen zwischen Kunden und Anbietern und fördern eine Orientierung an den jeweils anerkannten *Best Practices* in einem Tätigkeitsfeld (vgl. DeGEval 2002; Owen, Rogers 1999; Stufflebeam 2000; Rossi et al. 2004).

Standards für die Evaluationsforschung

Erste Standards für die Evaluationsforschung wurden in den 80er Jahren in den USA entwickelt. 1981 publizierte das *Joint Committee on Standards for Educational Evaluation* Standards für den erziehungswissenschaftlichen Bereich. Im Laufe der Jahre fanden diese Standards in immer mehr Themenfeldern Anwendung (vgl. Widmer 2006). 1994 präsentierte das *Joint Committee* eine überarbeitete Fassung mit dem Titel „The Program Evaluation Standards“ (Beywl, Widmer 2000: 250). Die Standards blieben im deutschen Sprachraum zunächst weitgehend unbeachtet.

Dies änderte sich erst mit der Gründung der Deutschen Gesellschaft für Evaluation (DeGEval) im Jahr 1997. 2001 veröffentlichte sie eigene Standards.

Diese Standards beanspruchen „Gültigkeit für verschiedene Ansätze der Evaluation, für unterschiedliche Evaluationszwecke sowie eine Vielzahl von Evaluationsfeldern“ (DeGEval 2002: 6). Sie richten sich sowohl an „Evaluatoren als auch an Personen und Einrichtungen, die Evaluationen in Auftrag geben, sowie an Beteiligte und Betroffene im Bereich des Evaluationsgegenstandes“ (DeGEval 2002: 12).

Die DeGEval (2002) benennt folgende **Funktionen der Standards**:

- Hilfe bei der Sicherung und Entwicklung der Qualität von Evaluationen,
- Dialoginstrument und fachlicher Bezugspunkt für einen Austausch über die Qualität von professionellen Evaluationen,
- Orientierungshilfe bei der Planung und Durchführung von Evaluationen,
- Anknüpfungspunkt für die Aus- und Weiterbildung in Evaluationen,
- Bezugsrahmen für die Evaluation von Evaluationen (Meta-Evaluation) sowie
- Schaffung von Transparenz über Evaluationen als professionelle Praxis.

Gute Evaluationen sollen nach Auffassung der DeGEval vier grundlegende Eigenschaften aufweisen: **Nützlichkeit, Durchführbarkeit, Fairness und Genauigkeit** (Wissenschaftlichkeit) (vgl. DeGEval 2002):

Eigenschaften guter Evaluationen

- Die Nützlichkeitsstandards sollen sicherstellen, dass die Evaluation sich an den erklärten Evaluationszwecken sowie am Informationsbedarf der vorgesehenen Nutzerinnen und Nutzer ausrichtet,
- die Durchführbarkeitsstandards sollen sicherstellen, dass eine Evaluation realistisch, gut durchdacht, diplomatisch und kostenbewusst geplant und ausgeführt wird.
- die Fairnessstandards sollen sicherstellen, dass in einer Evaluation respektvoll und fair mit den betroffenen Personen und Gruppen umgegangen wird.
- die Genauigkeitsstandards sollen sicherstellen, dass eine Evaluation gültige Informationen und Ergebnisse zu dem jeweiligen Evaluationsgegenstand und den Evaluationsfragestellungen hervorbringt und vermittelt.

Die DeGEval-Standards entsprechen weitgehend den Standards des Joint Committee (vgl. DeGEval 2002: 42). Die Standards sind nicht verbindlich. Sie stellen vielmehr einen grundlegenden Orientierungsrahmen zur Bestimmung von Qualitätsaspekten bei der Planung und Durchführung von Evaluationen dar. Zudem stehen einige Standards auch zueinander in Konkurrenz, so dass im Einzelfall zu entscheiden ist, welchem Standard der Vorrang einzuräumen ist (vgl. Widmer 2006).

Zwischen den vier grundlegenden Eigenschaften *guter* Evaluationen besteht keine unterschiedliche Gewichtung. Die Praxis hat jedoch genau dies zu leisten. Häufig steht der wissenschaftliche Anspruch, der in den Genauigkeitsstandards zum Ausdruck kommt, mit der Forderung nützliche Ergebnisse (Nützlichkeitsstandards) unter zumeist erheblichen Zeit- und Finanzrestriktionen zu produzieren (Durchführbarkeitsstandards), in Konflikt.

2 Konkurrierende und komplementäre Ansätze zur Evaluation

Außerhalb der Wissenschaft wird Evaluation vor allem im öffentlich-rechtlichen und non-profit Bereich eingesetzt. Beide Bereiche sind durch wenig direkten Wettbewerb und fehlende Gewinnorientierung gekennzeichnet. Öffentliche Einrichtungen, Stiftungen, Verbände und Nonprofit-Organisationen verwenden Evaluation zur Optimierung ihrer Planung und Steuerung sowie zur Überprüfung von Zielerreichung, Wirksamkeit und Nachhaltigkeit der von ihnen gewährten Leistungen oder durchgeführten Maßnahmen, Projekte und Programme.

In der Privatwirtschaft hingegen wird Evaluation als Steuerungs- und Qualitätssicherungsinstrument kaum angewendet. Hier wird vor allem auf aus der Betriebswirtschaft stammende Konzepte und Instrumente zurückgegriffen, von denen einige ähnlichen Zwecken dienen wie die Evaluation. Auch Nonprofit-Organisationen machen sie zunehmend für ihre Zwecke nutzbar. Einige dieser Instrumente wie Controlling und Balanced Scorecard, Benchmarking und Audit werden – teilweise komplementär zur Evaluation – für die Unternehmenssteuerung bzw. für die rationale Steuerung öffentlicher Programme und Einrichtungen eingesetzt. Deshalb werden sie hier kurz vorgestellt.

2.1 Controlling

Controlling als Teil des Managements ist ein System, das Informationen und relevante Daten über bereits eingetretene oder mögliche künftige Umweltveränderungen bereitstellt, um damit die Voraussetzungen für Anpassungshandlungen zu schaffen (Stockmann 2007: 75).

Das Controlling liefert kontinuierlich organisationsinterne Daten über das Verhältnis der geplanten und tatsächlichen Entwicklungen (Soll-Ist-Vergleich) an die Entscheidungsträger, um diesen zielgerichtete Korrekturen (Steuerungsentscheidungen) zu ermöglichen.

Zudem erfüllt Controlling Koordinationsaufgaben für das Führungssystem, indem es die Voraussetzungen zur Abstimmung des Handelns der einzelnen Abteilungen einer Organisation schafft. Hierfür werden u.a. Planungs-, Kontroll- und Informationssysteme aufgebaut.

Werden die Aufgabenstellungen des Controllings mit denen des Monitorings verglichen, dann finden sich einige Übereinstimmungen, aber auch gravierende Unterschiede (siehe **Tabelle 2.1**):

**Vergleich zwischen
Controlling und Monitoring**

Tabelle 2.1: Controlling und Monitoring im Vergleich (Stockmann 2007: 77)

Kriterien	Controlling	Monitoring
Historischer Ursprung	Buchhaltung	Sozialwissenschaften
Aufgabe	nicht bewertende Informations-sammlung und -koordination	nicht bewertende Informationssammlung
Adressat	Management	Management, aber auch Gesellschaft insgesamt
Berichtsfokus	vor allem Kostenaspekte, betriebsinterne Prozesse, keine Wirkungsdaten, input-/output-zentriert	breit, interne und externe Strukturen und Prozesse, auch Daten zu intendierten und nicht-intendierten Wirkungen
Kriterien	Controlling	Monitoring
Häufigkeit	kontinuierlich	kontinuierlich
Organisation	top down	top down oder bottom up
Kriterien	weitgehend kodifiziert	weitgehend offen

Methoden	insb. Kostenrechnung	sozialwissenschaftliches Spektrum
Datenquellen	ausschließlich quantitative	quantitative und qualitative

2.2 Balanced Scorecard

Die Balanced Scorecard (BSC) ist ein aus dem Controlling stammendes Konzept, das die Unternehmensstrategie in ein ausgewogenes Kennzahlensystem übersetzt.

Als Instrument des strategischen Managements nimmt die Scorecard neben finanziellen Kennzahlen auch Ursachenfaktoren (Treibergrößen) in die Betrachtung mit auf, um so ein ausgewogenes Bild über die Vision und Strategie eines Unternehmens aufzuzeigen. Die Finanzen, die Kunden, interne Geschäftsprozesse sowie Lernen und Entwicklung bilden die vier Perspektiven der Balanced Scorecard.

Die Zielerreichung wird anhand vorgegebener ausgewählter Kennzahlen gemessen. Neben quantitativen werden auch qualitative Tatbestände erfasst.

Auch die BSC stellt **kein Evaluationsverfahren** dar. Zwar weist sie mit ihrer Ausrichtung an Zielen, der Formulierung von Ursache-Wirkungs- oder Wertschöpfungsketten, der Messung von unabhängigen (Leistungstreibern) und abhängigen Variablen (Ergebniskennzahlen) sowie der Erfassung von quantitativen und qualitativen Tatbeständen durchaus evaluative Elemente auf, doch das übergeordnete Ziel, welches sich in finanziellen Kennzahlen ausdrückt, wird genauso wenig hinterfragt wie die Fixierung auf die drei weiteren Perspektiven. Die Berücksichtigung verschiedener Interessengruppen und deren Werte wird nicht, zumindest nicht explizit, verlangt. Neben Kunden und Mitarbeitern werden keine anderen Stakeholdergruppen mit einbezogen. Anders als Evaluation scheint BSC zudem vornehmlich top-down eingesetzt zu werden. Die Fixierung auf die Umsetzung der Unternehmensstrategie und die Frage, wie diese zur Erfüllung der finanziellen Ziele beiträgt, lässt keine breitere evaluative Sichtweise, die z. B. nicht-intendierte Wirkungen mit einschließt, zu (vgl. Stockmann 2007: 77ff.).

Am ehesten lässt sich auch die BSC mit dem Monitoring vergleichen. Auf der Basis theoretischer Annahmen über programm- und unternehmensrelevante Zusammenhänge werden kontinuierlich Daten gesammelt und ausgewertet. Dabei ist das Datengerüst beim BSC-Ansatz durch die vier vorgegebenen Perspektiven bereits stark vorstrukturiert und weit weniger entscheidungsoffen bei der Auswahl der Indikatoren als ein Monitoring-System, sodass es sein könnte, dass wichtige Einflussgrößen und Wirkungen gar nicht erfasst werden (vgl. Stockmann 2007: 82).

Welcher Methodensets (außer Dokumentenanalyse und einzelner Interviews) sich der BSC-Ansatz zur Datengewinnung bedient, bleibt weitgehend unklar. Dies könnte unter anderem daran liegen, dass nur leicht zu messende Indikatoren verwendet werden, die keine komplexen methodischen Erhebungsverfahren notwendig machen. Deshalb dürften z. B. auch eher Input- und Output- als Outcome- oder Impactindikatoren erfasst werden (vgl. Stockmann 2007: 82).

Vergleich der BSC mit dem Monitoring

2.3 Benchmarking

Benchmarking ist eine Managementmethode zur Leistungsoptimierung mithilfe zielgerichteter Vergleiche unter mehreren Unternehmen. Benchmarking beruht auf der Orientierung an den Besten innerhalb einer Vergleichsgruppe. Diese Vorgehensweise bezeichnet man als *Best Practice*. Best Practice existiert auf verschiedenen Betrachtungsebenen. Auf der obersten Betrachtungsebene werden Konzepte in Frage gestellt („die richtigen Dinge tun“). Auf der untersten Betrachtungsebene werden Prozesse in Frage gestellt („Dinge richtig tun“), um die Prozesseffizienz zu verbessern.

Benchmarking weist einige Nähe zur Evaluation auf. Die Etablierung von Benchmarks, die Messung von Leistungsfaktoren im eigenen und in der/den Vergleichsorganisation(en), der bewertende Vergleich sowie die Suche nach Erklärungen für die Unterschiede, als Voraussetzung für die Verbesserung der eigenen Prozesse lassen Ähnlichkeiten zu ex-ante oder formativen Evaluationen erkennen. Eine Verbindung zum Monitoring wird in den weiteren Arbeitsschritten deutlich, bei denen es explizit um die Überwachung der durch die implementierten Maßnahmen hervorgerufenen Wirkungen (z.B. Leistungssteigerung, Rationalisierungseffekte etc.) geht (vgl. Stockmann 2007: 83ff.; Owen, Rogers 1999: 180).

Vergleich zwischen Benchmarking und Evaluation

Anders als bei Evaluationen werden aber beim Benchmarking keine theoretischen Überlegungen angestellt und daraus Bewertungskriterien abgeleitet oder gar Stakeholder beteiligt. Die Vergleichskriterien sind weitgehend gesetzt und orientieren sich an den Best practices der Vergleichsorganisation(en). Demgegenüber ist die Bandbreite evaluativer Fragestellungen weitaus umfangreicher. Gleichwohl kann Benchmarking in das Portfolio evaluativer Instrumente aufgenommen werden. Wie schon eingangs erwähnt können die mit verschiedenen Instrumenten gewonnenen Daten je nach Zielsetzung für ein Benchmarking genutzt werden (vgl. Stockmann 2007: 83ff.).

2.4 Audit

Das Audit entwickelte sich wie das Controlling aus dem Rechnungswesen. Neben dem klassischen Finanzaudit werden u.a. das Operational- und Management-Audit oder im öffentlichen Bereich das Performance-Audit unterschieden.

Aufgabenstellung und organisatorischer Ablauf eines Performance-Audits weisen große Ähnlichkeiten mit Formen der Evaluation auf. Audit und Evaluation sind Instrumente einer modernen an Rationalitätskriterien orientierten Politik. Beide Ansätze leisten einen Beitrag zur Steuerungsfähigkeit und öffentlichen Kontrolle.

Auditing hat sich in der Aufgabenstellung und der Anwendung der Verfahren in den letzten Jahren der Evaluation angenähert. Gravierende Differenzen resultieren aus dem institutionellen Kontext von Evaluation und Audit. Diese lassen sich letztlich nicht auflösen.

Vergleich zwischen Auditing und Evaluation

Tabelle 2.2 zeigt die Unterschiede zwischen Evaluation und Auditing auf:

Tabelle 2.2: Auditing und Evaluation im Vergleich (Stockmann 2007: 95)

Kriterien	Auditing	Evaluation
Fokus	Überwachung (regularity and compliance) Administrative Orientierung	Theorie und Erklärung Wissenschaftliche Orientierung
Mission	Verbesserung der Transparenz und Kontrollierbarkeit öffentlicher Institutionen	Beitrag zur Lösung sozialer Probleme
Phasen des polit. Prozesses	Implementation	Planung, Implementation, Wirkung
Zentrale Aufgaben	Analyse des Implementationsprozesses und der unmittelbaren Leistungen (output) und Wirkungen (outcome)	Analyse des Implementationsprozesses sowie output, outcome, Nachhaltigkeit, Relevanz, Nützlichkeit und weitere Aspekte
Untersuchungsperspektive	Retrospektiv	Prospektiv, gegenwärtig, retrospektiv
Untersuchungslogik	Normativ, Soll-Ist-Vergleich	Wirkungsorientiert, Ursache-Wirkungs-Logik
Bewertungskriterien	Vorab genau definiert anhand übergreifender Standards	Weniger klar definiert. Auswahl und Klärung teilweise während des Evaluationsprozesses.
Adressaten	Legislative	Auftraggeber, Peergroups, Stakeholder, Öffentlichkeit
Untersuchungsdesigns	Vornehmlich Querschnittsanalysen	(Feld-)Experimente, Längsschnitt-, Querschnittsanalysen, u. a.
Methoden	Dokumentenanalyse, Interviews	Gesamtes Sozialwissenschaftliches Spektrum
Berücksichtigung von Stakeholderinteressen	Nicht wichtig	Sehr wichtig
Berichtsergebnisse	Verpflichtend	Nicht bindend

2.5 Fazit

Ganz allgemein kann man feststellen, dass Evaluation generell offener, wissenschaftlicher, umfassender, teilhabeorientierter, weniger normativ, methodenreicher (quantitative und qualitative Aspekte umfassend), vom Untersuchungsfeld und den verwendeten Kriterien her breiter angelegt und an einen vielschichtigen Adressatenkreis gerichtet ist. Dafür ist Evaluation jedoch auch weniger fokussiert und hat für die Steuerungsverantwortlichen weniger Verbindlichkeit als betriebswirtschaftliche Instrumente und das Audit (vgl. Stockmann 2007: 100).

3 Evaluation im Rahmen von Qualitätsmanagement und -sicherung

Da Qualität sich vor allem an der Bewertung des Nutzens durch die Kunden bemisst, machen die meisten Qualitätsmanagementsysteme die Qualität zum wichtigsten Erfolgskriterium eines Unternehmens. Alle Prozesse und Strukturen sind daher an der Befriedigung der Kundenbedürfnisse orientiert.

Verschiedene Modelle stehen zur Gestaltung des Qualitätsmanagements zur Verfügung. Sie können auf der DIN EN ISO 9000-9004 oder auf dem Leitgedanken eines umfassenden Qualitätsmanagements (Total Quality Management = TQM) aufbauen.

Allen Qualitätsmanagementsysteme fokussieren auf die Organisationsstruktur, die Prozesse und die Sicherung der Qualität des zu produzierenden Produkts bzw. der zu erbringenden Dienstleistung. Standards und Normen sollen helfen, eine einheitliche Informationsbasis zu schaffen, mit dessen Hilfe Unternehmen bewertet und miteinander verglichen werden können. Entscheidendes Bewertungskriterium ist dabei die Qualität (bzw. die Beschaffenheit) des Produkts bzw. der Dienstleistung (vgl. Raidl 2001: 53).

Die DIN EN ISO definiert Mindestanforderungen an ein Qualitätsmanagementsystem und zielt auf die vollständige Dokumentation der Arbeitsabläufe. Dem liegt der Gedanke zugrunde, bestmögliche Qualität sei das Ergebnis logisch geplanter, vereinheitlichter und für alle Beteiligten transparenter Leistungserstellung. Die Normen der DIN EN ISO regeln alle, im Rahmen eines Qualitätsmanagementsystems zu definierenden und zu organisierenden Bereiche.

Der Total Quality Management (TQM) Ansatz stellt ein alternatives Qualitätskonzept dar. TQM baut nicht auf fixierte Normen auf, Stattdessen werden alle Unternehmensprozesse am Vorrang der Qualität ausgerichtet. TQM wird vom Management eingeführt und es steuert die kontinuierliche Verbesserung aller Unternehmensprozesse. Die Qualitätssicherung und -entwicklung liegt jedoch in der Verantwortung aller Mitarbeiter. Die europäische Version des TQM ist die European Foundation for Quality Management (EFQM). EFQM unterscheidet fünf Befähiger- und vier Ergebniskriterien, die jeweils in weitere (insgesamt 32) Unterkriterien aufgeschlüsselt werden. Anhand der Kriterien für die Befähiger (Führung, Politik und Strategie, Mitarbeiter, Ressourcen, Prozesse) können Handlungsweisen, Tätigkeiten und Prozesse in einem Unternehmen sowie deren Anwendungsgrad untersucht und bewertet werden. Letztlich ist das Ziel aller Qualitätsbemühungen die Verbesserung der Geschäftsergebnisse. Daher werden als Schwerpunkte auf der Ergebnisseite die (Geschäfts-)Ergebnisse, die Mitarbeiter- und Kundenzufriedenheit sowie das Image außerhalb des Unternehmens systematisch bewertet (Stockmann 2007: 102).

Die in Kapitel 1 und 2 erläuterten Instrumente werden eigenständig zur Unternehmenssteuerung und Qualitätsentwicklung verwendet. Sie lassen sich aber auch in die darüber hinausgehenden, umfassenderen Qualitätsmanagementsysteme nach DIN EN ISO oder EFQM integrieren. Dabei dürfte kaum überraschen, dass **sozialwissenschaftliche Konzepte wie Monitoring und Evaluation kaum, betriebswirtschaftliche hingegen häufig eingesetzt werden**. Jeder (DIN EN ISO-) Zertifizierung bspw. geht ein Audit voraus, in dem belegt werden muss, dass die vorgeschriebenen Normen erreicht werden. Kennzahlensysteme, die durch das Controlling oder eine Balanced Scorecard vorgegeben sind, lassen sich anschließend zur kontinuierlichen Normüberprüfung einsetzen, wobei die zu kontrollierenden Größen durch die DIN EN ISO- oder EFQM-Kriterien vorgegeben sind. Die Balanced Scorecard weist dabei eine große Nähe zum EFQM-Modell auf. Die Leistungstreiber- und Ergebniskennzahlen die im Rahmen der BSC erhoben werden, lassen sich zu einem Großteil in das EFQM-Modell mit seinen Befähiger- und Ergebniskriterien überführen. Zudem wird die BSC zur Führung durch Zielvereinbarungen (Management by Objectives) im TQM eingesetzt.

Benchmarking spielt im Rahmen des Total Quality Management eine zentrale Rolle. Es liefert den Vergleich mit Unternehmen (oder Unternehmensteilen), die im Qualitätsbereich vorbildlich sind, und gibt Anregungen für Qualitätsverbesserungen im Unternehmen. Insbesondere bei der nach EFQM vorzunehmenden Selbstbewertung wird ausdrücklich der objektive Blick nach außen verlangt. (vgl. Siebert, Kempf 2002: 25).

Das Interesse an Qualitätsmanagement hat seit Mitte/Ende der 90er Jahre auch im Non-profit-Bereich stark zugenommen. In Ermangelung eigener Konzepte wird dabei häufig versucht, die für den Profit-Sektor entwickelten Modelle (wie DIN EN ISO und EFQM) zu übertragen. Allerdings nicht immer mit Erfolg (vgl. Klausegger, Scharitzer 1998: 371, 387; Langnickel 2003: 45; Beckmann 2004: 9f.). „A principal difficulty is that the skills and techniques of the private sector are not directly transferable to the public sector. In particular, private sector techniques, with their emphasis on profitability as a measure of effectiveness, are of little value in evaluating policy outcomes“ (Smith 1996: 169). Dies liegt vor allem daran, dass Profit- und Non-profit-Organisationen eine Reihe organisations- und situationsbezogener Unterschiede aufweisen (zur näheren Erläuterung vgl. Stockmann 2006: 44ff.).

Die Herstellung wettbewerbsähnlicher Strukturen, die Ausrichtung an den Kundenbedürfnissen sowie die Schaffung eines Qualitätsbewusstseins erzeugt die Notwendigkeit leistungs- und wirkungsbezogener Steuerung. Dies bringt methodische Schwierigkeiten mit sich, die mit den in der Verwaltung üblichen Kontroll- und Finanzinstrumenten nicht zu überwinden sind. Aus diesem Grund finden zunehmend umfassendere Kennzahlensysteme und Benchmarking Einsatz (vgl. Stockmann 2007: 104).

Ein auf betriebswirtschaftlichen Kennzahlen basierendes Controlling und Benchmarking mag dazu beitragen, die Effizienz öffentlich erbrachter Leistungen (Output) zu steigern. Eine Aussage über die damit erzielten Wirkungen, die den eigentlichen Zweck der Leistungserstellung ausmachen, ist damit jedoch nicht möglich (vgl. Brüggemeier 2004: 375). „An appropriate public sector measurement of outputs necessitates the evaluation of the extent to which the organization has met its primary goals. We have argued that the public sector exists to meet need, and therefore it is appropriate that the extent to which need is met should be the basis of efficacy measures“ (Smith 1996: 169).

Selbst in Großbritannien, wo die Leistungsmessung deutlich weiter entwickelt ist als in Deutschland, liegt der Fokus stärker auf betrieblichen Indikatoren und weniger auf der Effektivität politischer Maßnahmen oder gar ihrer Wirksamkeit (vgl. Stockmann 2007: 1005).

Die Einengung auf betriebswirtschaftliche Steuerungsgrößen wird teilweise als „ökonomischer Reduktionismus“ (Kuhlmann 2004a: 12) kritisiert. Die ausschließliche Nutzung betriebsbezogener Kennzahlen und Benchmarking reicht offenbar nicht für eine leistungs- und wirkungsorientierte Steuerung, wie sie z.B. im Rahmen des *New Public Management* notwendig ist, aus. Zwar gibt es Anzeichen, dass sich international die Steuerung zunehmend an Outcomes orientiert (vgl. Bühler 2002: 273). In Deutschland kann eine vergleichbare Entwicklung jedoch nicht beobachtet werden (vgl. Brüggemeier 2004: 375; Kuhlmann 2004b: 99ff.; Pedé 2000: 335ff.). Hierzulande dominieren weiterhin quantitativ und monetär ausgerichtete Mess- und Bewertungsverfahren. Daten zur Qualität von Prozessen werden nur selten erhoben (vgl. Bogumil 2004: 393), obwohl gerade dort die Potenziale des Benchmarking liegen. Auf der Basis reiner Zahlenvergleiche ist die Idee des Benchmarking im öffentlich-rechtlichen Sektor kaum zu verwirklichen (vgl. Kuhlmann 2004b: 103). Darüber hinaus verspricht Benchmarking dann seinen größten Nutzen, wenn es branchenunabhängig vorgenommen wird. Im öffentlichen Sektor werden jedoch nur gleiche Einheiten (z.B. Kommunen, Altenheime, Sozialhilfeeinrichtungen etc.) miteinander verglichen, so dass das Innovationspotenzial von vornherein stark eingeschränkt ist (Stockmann 2007: 106).

Zentrales Defizit der bisher praktizierten Kennzahlensysteme, Leistungsmessungen und Benchmarking ist die zu geringe Ausrichtung auf Wirkungen und der Mangel an Aussagen

**Stärke des
Evaluationsansatzes**

darüber, wie *Best practices* entstehen (vgl. Bogumil 2004: 394). Hier aber liegt die größte Stärke des Evaluationsansatzes, dessen Hauptaufgabe in der Entdeckung und Messung von Wirkungen sowie in der Ursache-Wirkungs-Analyse liegt, der in Deutschland bis dato allerdings nur wenig Anwendung findet (vgl. Wollmann 2003: 41).

Die Außerachtlassung von Ursache-Wirkungs-Zusammenhängen im Rahmen von Controlling und Benchmarking lässt wesentliche Innovations- und Lernpotenziale ungenutzt. Evaluation könnte hier Abhilfe schaffen. Mit Evaluationen lassen sich umfassende Output- und Outcome-Bewertungen vornehmen und große Lerneffekte erzielen. Es erscheint daher sinnvoll, sowohl betriebswirtschaftlich basierte Instrumente für die Unternehmens- oder Organisationssteuerung, als auch evaluative Ansätze einzusetzen.

Eine Kombination aus kennzahlenbezogenen Instrumenten (Controlling, Balanced Scorecard), Ansätzen zur kontinuierlichen Leistungserfassung von quantitativen und qualitativen Daten (Monitoring) und Evaluationen, die nicht nur die Zielerreichung prüfen, sondern die Ziele selbst auf den Prüfstand stellen, intendierte und nicht-intendierte Wirkungen einander gegenüberstellen sowie weitergehende Kriterien (wie z.B. Nachhaltigkeit und gesellschaftspolitische Relevanz) für die Bewertung einsetzen und die Ursachen für beobachtete Veränderungen erkunden, könnte der Schlüssel für eine erfolgreiche Organisationssteuerung, insbesondere im öffentlichen Sektor sein und stärker als bisher zu einer Ausschöpfung von Innovations- und Lernpotenzialen führen (vgl. Stockmann 2007: 107).

4 Evaluation im Gesundheitswesen

4.1 Evaluation von Gesundheitspolitik

Seit einigen Jahren lässt sich eine zunehmende Tendenz zur gesetzlich verpflichtenden Evaluation und Begleitforschung zu den Auswirkungen der Gesundheitspolitik beobachten. Genannt seien hier beispielhaft

- §17b Abs.8 KHG, der eine Pflicht zur Begleitforschung zu den Auswirkungen des DRG-Systems festschreibt,
- §119b Abs.3 SGB V – Evaluation des Versorgungsgeschehens im Bereich der vertragsärztlichen Versorgung im Zusammenhang mit der ambulanten Behandlung in stationären Pflegeeinrichtungen sowie
- §116b Abs.9 SGB V – Evaluation der Auswirkungen der ambulanten spezialfachärztlichen Versorgung auf Kostenträger, Leistungserbringer und Patientenversorgung.

**Gesetzlich verpflichtende
Evaluationen**

So begrüßenswert die Etablierung einer Politikfolgenforschung ist, so diametral stehen sich Anspruch und Wirklichkeit in der Praxis teilweise gegenüber. Verdeutlicht sei dies am Beispiel der Begleitforschung zu den Auswirkungen der DRG-Einführung:

Die Einführung der DRG als Form der Vergütung stationärer Leistungen im Jahr 2010 markiert den vorläufigen Endpunkt der schrittweisen Umstellung von einer retrospektiven Einzelleistungsvergütung hin zu einer prospektiven Vergütung auf der Grundlage von Fallpauschalen. In der Bundesrepublik laufen die Bemühungen um die Einführung von DRGs seit 1983. Die ab 1. Januar 1996 geltende Bundespflegesatzverordnung führte 73 Fallpauschalen und 147 Sonderentgelte ein, mit denen rund 20% aller Leistungen im Krankenhaus abgedeckt wurden. Sie stellt eine Art Vorläufer der DRGs dar. Die Einführung der DRGs als Vergütungskonzept stationärer Leistungen in der Bundesrepublik ist durch §17b KHG im Jahr 2000 geregelt worden. Mit der organisatorischen Vorbereitung und Einführung wurden die Deutsche Krankenhausgesellschaft, der Spitzenverband Bund der GKV sowie der Verband der PKV betraut. Für die technische Aufbereitung der deutschen DRGs gründeten die Akteure im Jahr 2001 das Institut für das Entgeltsystem im Krankenhaus (InEK).

**Begleitforschung zu
Auswirkungen des
DRG-Systems**

Die DRGs wurden ab 2003 zunächst zur schrittweisen Erlösermittlung für die Krankenhäuser eingesetzt. Nach einer mehrfach verlängerten sogenannten Konvergenzphase ist die Honorierung stationärer Leistungen nach diesem fallpauschalierten Entgeltsystem ab 2010 verbindlich geworden (vgl. Niehoff 2010: 69).

Trotz einer vom Gesetzgeber für die DRG-Einführung vorgeschriebenen mehrjährigen Lernphase gibt es bis heute weder systematische interdisziplinäre Untersuchungen zu den Aushandlungsprozessen der Krankenhausbudgets noch Analysen zu den Wirkungen der Budgets auf das soziale System Krankenhaus und seine Hauptakteure. Das Bundesministerium für Gesundheit (BMG) verzichtete nach der Einführung der Fallpauschalen trotz eines gegenteiligen Votums des Bundesrates vorsätzlich auf die Analyse der Auswirkungen auf die Versorgungsqualität (vgl. Braun et al. 2010: 9).

Der Bundesrat nämlich stimmte der Bundespflegesatzverordnung am 8. Juli 1994 nur unter der Maßgabe einer umfassenden Evaluation zu: „Neben den betriebswirtschaftlichen Fragen, die mit der Einführung von Fallpauschalen und Sonderentgelten verbunden sind, sollen insbesondere auch deren Auswirkungen auf die Leistungsqualität und Leistungsstruktur in die Untersuchung einbezogen werden“ (BR-Drs. 381/2/94: 14).

Das BMG setzte sich über dieses Votum hinweg und hielt seinerseits lediglich eine Begleitforschung zu den betriebswirtschaftlichen Auswirkungen. Die Risiken für die Versorgung

wurden nicht näher untersucht. Die mit der Evaluation beauftragten Institute hatten dem BMG in Kenntnis des Bundesratsbeschlusses zwei Untersuchungsmodule vorgelegt, eines zu den rein ökonomischen Auswirkungen und ein zweites zu den Auswirkungen auf die Patientenversorgung. Jedoch wurden lediglich Teile des ersten, rein betriebswirtschaftlichen Moduls vom BMG finanziert. Die von den Instituten geplante Expertenbefragung lehnte es ab und beendete das Forschungsprojekt zudem ein Jahr früher als geplant (vgl. DKI, #KGesundheitsforschung 1999: 15). Die Beschränkung des Untersuchungsauftrages auf rein ökonomische Fragestellungen führte nach Einschätzung der beauftragten Institute dazu, „dass mögliche Veränderungen auf der Patientenebene im Hinblick auf die Behandlung nicht bzw. nur unzureichend einbezogen werden können und somit ein zentraler Aspekt der Wirkungen auf die Leistungsqualität unberücksichtigt bleibt“ (DKI, I+G Gesundheitsforschung 1999: 14).

Trotz der ab 2003 wirksamen klaren gesetzlichen Verpflichtung zur Durchführung einer Begleitforschung zu den Auswirkungen der DRG in §17b Abs.8 KHG, startete diese erst 2009, also fast am Ende der Einführungs- und Konvergenzphase.

Die heute vorliegenden Erkenntnisse zu den Auswirkungen der DRG auf die Versorgungsqualität gehen fast vollständig auf Initiativen wissenschaftlicher Einrichtungen und Einzelpersonen zurück. Zuvorderst zu nennen sind hier das Wissenschaftszentrum Berlin und das Zentrum für Sozialpolitik der Universität Bremen, die seit 2003 mit Unterstützung der Hans Böckler Stiftung, der Gmünder Ersatzkasse, der LÄK Hessen und der Techniker Krankenkasse das Projekt „Wandel von Medizin und Pflege im DRG-System (WAMP)“ (Wissenschaftszentrum Berlin für Sozialforschung 2014) durchführt.

Interessant zu erwähnen ist auch, dass das 2001 zunächst als gemeinnützige GmbH gegründete InEK 2007 in eine werbende GmbH umgewandelt wurde. Es darf die Frage gestellt werden, ob ein Unternehmen mit Gewinnerzielungsinteresse, das die G-DRG heute erfolgreich in die ganze Welt verkauft, über die notwendige Unabhängigkeit verfügt, um als Institut nach §17b Abs.8 KHG nach wie vor die DRG-Begleitforschung durchzuführen und zwar finanziert durch den DRG-Systemzuschlag. Es gibt Anlass zu der Vermutung, dass das InEK aus unternehmerischem Eigeninteresse selbst die Evaluation behindert respektive in ihrem Interesse steuert.

4.2 Evaluation auf Seiten der Kostenträger

Auf Seiten der Kostenträger, namentlich der Gesetzlichen Krankenversicherungen, findet Evaluation im Wesentlichen im den ihnen gesetzlich auferlegten Umfang statt. Im Internet finden sich eine Reihe von Evaluationsberichten einzelner Krankenkassen. Beispielfhaft können hier genannt werden:

- 1) Evaluationen zu den Wirkungen von Disease Management Programms gem. §137f. Abs.2 Nr. 6 i.V.m Abs.4 SGB V, z.B. die Evaluation des Disease Management Programms Brustkrebs der Techniker Krankenkasse in der Region Hamburg (vgl. Techniker Krankenkasse 2013).

Die Evaluationen zum DMP Brustkrebs einzelner Krankenkassen unternehmen immerhin den *Versuch*, auch Auswirkungen auf die Ergebnisqualität zu evaluieren. Ansonsten bilden die gesetzlich/vertraglich vorgeschriebenen Datenerhebungen quasi ausschließlich Struktur- und Behandlungsqualität ab.

Beim DMP Brustkrebs heißt das z. B.:

- Relevante Endpunkte werden nicht erhoben (Tumor-/Rezidivfreiheit, Überleben),
- zeitliche Zusammenhänge sind schwer nachvollziehbar (Diagnosestellung, OP, adjuvante Therapie, Nachsorge).

- Teilweise erlauben die erhobenen Parameter nicht einmal die Bewertung expliziter Qualitätskriterien (Behandlung in einem am DMP teilnehmenden Krankenhaus).
- Eine Verknüpfung zum Mammographiescreening ist nicht gegeben.

Der MDK Berlin/Brandenburg hat im Rahmen des Qualitätsberichtes für das DMP Brustkrebs in Berlin z.B. die Häufigkeit von Lymphödemen als Merkmal der Lebensqualität, einem Indikator im Rahmen der Messung von Ergebnisqualität, gemessen (vgl. Reiner-Theisen et al. 2013).

Außerdem erhalten die teilnehmenden Ärzte vom MDK individuelle Feedbackberichte, die ihnen ihre Arbeit vergleichend widerzuspiegeln versuchen.

- 2) Evaluationen zu den Wirkungen von Wahlтарifen gem. § 53 Abs.9 SGB V. Es liegen hier noch keine Ergebnisse vor.
- 3) Evaluationen zu den Wirkungen der Hausarztzentrierten Versorgung gem. § 73b SGB V, z.B. die Evaluation von fünf Ersatzkassen-Hausarztverträgen auf Grundlage des § 73b SGB V (vgl. AQUA-Institut 2013).
- 4) Evaluationen zu den Wirkungen der Integrierten Versorgung, z.B. die Studie zur Evaluation des Integrierten Versorgungsvertrages Homöopathie (vgl. DZVhÄ 2013).
- 5) Evaluationen zu den Wirkungen von Präventionsprogrammen gem. § 20ff SGB V. Hierzu liegen derzeit keine Evaluationsberichte vor. Der GKV-Spitzenverband hat allerdings ein Anwenderhandbuch für die Durchführung von Evaluationen in diesem Bereich veröffentlicht (vgl. Arbeitsgemeinschaft der Spitzenverbände der Krankenkassen 2008a; 2008b).

Die Evaluationspflichten werden zum Teil auch durch den Spitzenverband der Gesetzlichen Krankenversicherung wahrgenommen.

Festzustellen ist, dass die Qualität der Arbeiten in erheblichem Maße divergiert. Teilweise handelt es sich bei den als Evaluation bezeichneten Arbeiten lediglich um Auswertungen der im Rahmen der beleuchteten Programme erhobenen Daten.

Probleme bereiten Evaluationen auf dieser Ebene insofern auch, als die einzelnen Krankenkassen für die Evaluation lediglich auf die Daten ihrer eigenen Versicherten zugreifen können. Aufgrund der teilweise erheblich unterschiedlichen Morbiditätsstruktur des Versichertenkreises sind Rückschlüsse auf die Auswirkungen für die GKV insgesamt schwierig.

4.3 Evaluation auf Seiten der Leistungserbringer

Auf der Ebene der Leistungserbringer orientiert sich die Leistungsbewertung vor allem an betriebswirtschaftlichen Kennzahlensystemen. Allerdings sind die Leistungserbringer nach SGB V heute gesetzlich auch dazu verpflichtet, einrichtungsintern ein Qualitätsmanagement einzuführen und weiterzuentwickeln (§ 135a Abs.2 Nr.1 SGB V). Wie bereits oben dargelegt, orientieren sich die meisten Qualitätsmanagementsysteme in erster Linie an der Bewertung des Nutzens einer Leistung durch den Kunden als oberstes Kriterium für Qualität. Im Rahmen der Datenerhebung für das Qualitätsmanagement durch die Leistungserbringer wird daher in hohem Maße auf die Kundenzufriedenheit abgestellt. Zumindest für den Leistungsbereich der Gesetzlichen Krankenversicherung und anderer Sozialgesetzbücher stellt sich damit die Frage, wer Kunde der Leistungserbringer ist. Im ökonomischen, wie auch juristischen Verständnis ist Kunde, wer die erbrachte Leistung bezahlt. Dies sind im Anwendungsbereich der Sozialgesetzbücher dann aber, anders als in der Praxis gehandhabt, nicht die Patienten, sondern die Kostenträger.

Ob aus Sicht der Kostenträger allerdings das Kriterium der Patientenzufriedenheit ein entscheidendes ist, hängt letztlich von der ganz grundsätzlichen Frage ab, welche Primärziele die

Gesetzliche Krankenversicherung und andere Kostenträger verfolgen. Das SGB V, um ein Beispiel zu nennen, definiert hier die Aufgabe, die Gesundheit der Versicherten zu erhalten, wiederherzustellen oder ihren Gesundheitszustand zu verbessern (§1) und zwar durch eine ausreichende, zweckmäßige und wirtschaftliche Versorgung (§12) mit präventiven, medizinischen, pflegerischen und rehabilitativen Leistungen. Es darf zumindest die Frage gestellt werden, ob, wo und wenn ja, in welchem Umfang die Patientenzufriedenheit als subjektives Kriterium hier Berücksichtigung zu finden hat.

Der Gesetzgeber selbst macht zwar das Kriterium der Qualität medizinischer Versorgung zu einem zentralen Aspekt der Leistungserbringung, legt aber selbst nicht fest, anhand welcher Kriterien die Qualität zu messen sei. Er hat vielmehr dem Gemeinsamen Bundesausschuss die Aufgabe übertragen, Richtlinien zur Qualitätssicherung zu erlassen (vgl. § 137 SGB V). Das Gesetz zumindest erwähnt die Patientenzufriedenheit nicht und auch der Gemeinsame Bundesausschuss ist beim Erlass von Richtlinien in erster Linie dem Primärziel einer ausreichenden, zweckmäßigen und wirtschaftlichen Leistungserbringung verpflichtet. Patientenzufriedenheit kann, so wünschenswert sie sein mag, nur insofern Berücksichtigung finden, als sich dieses Kriterium in übergeordnete Ziele einfügen lässt.

Für den Bereich der Selbstzahler (inkl. der Privaten Krankenversicherung) sieht dies selbstverständlich anders aus, da hier Personenidentität zwischen Patient und Kostenträger besteht.

5 Sozialer Kontext der Evaluation

Evaluation umfasst neben einer methodischen auch eine wesentliche soziale Komponente, die daraus resultiert, dass Evaluationsvorhaben in einen sozialen Kontext eingebunden sind und Nutzen stiften wollen. (vgl. Fitzpatrick et al. 2004; Owen, Rogers 1999; Patton 1997; Rossi et al. 2004).

„Evaluation will dazu beitragen, die Planung und Umsetzung sozialer Interventionen, sei es im Rahmen einzelner Maßnahmen kurzfristiger Projekte, lang andauernder Programme oder gar unbefristeter Dienstleistungsangebote wirkungsvoller zu gestalten, um letztlich die Qualität von Programmen und Dienstleistungen zu verbessern“ (Stockmann 2006: 261).

In diesem Sinne stellt Evaluation selbst eine soziale Intervention dar. Sie zielt darauf ab (vgl. Stockmann 2006: 17f.),

- a) Programmentwicklungen zu unterstützen und die Planung von Interventionen zu verbessern („ex-ante-Evaluation“),
- b) die effektive und effiziente Umsetzung sozialer Maßnahmen zu befördern, um den Wirkungsgrad zu erhöhen („on-going-Evaluation“) oder
- c) Wirkungen von Maßnahmen retrospektiv zu bewerten („ex- post-Evaluation“).

Evaluation berührt damit unmittelbar Fragen der Gestaltung, Kontrolle und Steuerung von Programmen und gerät dadurch, da immer auch interessensgeleitet, nahezu automatisch in soziale und politische Konfliktfelder (vgl. Stockmann 2004; 2006). Die im folgenden aufgeführten Interessengruppen (oder auch Einzelpersonen), sind regelhaft, ob mittelbar oder unmittelbar, an einer Evaluation beteiligt oder von ihr betroffen und können deren Planung und Durchführung mehr oder weniger stark fördern oder behindern (vgl. auch Fitzpatrick et al. 2004; Owen, Rogers 1999; Patton 1997; Rossi et al. 2004; Wottawa, Thierau 2003; DeGEval 2002):

- **Auftraggeber der Evaluation,**
- **Geldgeber der Evaluation** (müssen nicht mit den Auftraggebern der Evaluation identisch sein),
- **Geldgeber des Programms,**
- **Programmanager, -mitarbeiter** (die Personen, die für das Programm auf strategischer oder operativer Ebene unmittelbar verantwortlich sind),
- **Zielgruppen des Evaluationsgegenstandes,**
- **das Evaluationsteam** selbst sowie
- **sonstige Betroffene** (Personen oder Gruppen mit wenig oder ohne Einfluss auf die Evaluationsdurchführung, die jedoch von dieser oder durch den Evaluationsgegenstand selbst beeinflusst werden).

Interessengruppen

Eine weitere Aufgliederung (vgl. z.B. die Checkliste in Fitzpatrick et al. 2004: 202) wäre ohne weiteres möglich. Jedoch zeigt sich bereits an der vorliegenden Unterteilung, dass einzelne Gruppen oder Personen sehr unterschiedliche Interessen hinsichtlich eines Evaluationsgegenstandes und damit auch an einer Evaluation selbst haben oder entwickeln können. Der Erfolg eines Programmes wird entsprechend den Interessenlagen der unterschiedlichen Gruppen daher unter Umständen sehr heterogen bewertet werden. Die Geldgeber eines Programms etwa werden in aller Regel primär daran interessiert sein, zu erfahren, wie dieses im Hinblick auf die Kosteneffizienz abschneidet. Die Programmmitarbeiter hingegen werden aller Wahrscheinlichkeit nach vorrangig an einer Weiterfinanzierung ihres Arbeitsplatzes interessiert sein, während die Programmteilnehmer vor allem ihren individuellen Nutzen vor Augen haben werden.

Evaluatoren müssen in der Lage sein, sowohl mit unterschiedlichen Interessengruppen zu kommunizieren als auch gegebenenfalls zwischen diesen zu vermitteln und dabei berücksichtigen, dass sich unterschiedliche Interessen nie zur vollen Zufriedenheit Aller auflösen lassen. Evaluationsergebnisse, die bestimmten Interessen oder Positionen von Stakeholdern zuwiderlaufen, stoßen nicht selten auf Ablehnung und erbitterten Widerstand. Stakeholder, die ihre eigenen Positionen gefährdet sehen, versuchen oftmals, die Evaluation und die Evaluatoren zu diskreditieren. Es kann mithin davon ausgegangen werden, dass vor dem Hintergrund der Interessenvielfalt nahezu zwangsläufig, und zwar unabhängig von den jeweiligen Ergebnissen und der methodischen Qualität einer Evaluation, irgendjemand immer unzufrieden sein wird (vgl. Rossi et al. 2004: 374).

Um eine möglichst breite Akzeptanz des Evaluationsvorhabens und damit schließlich auch die Umsetzung und Nutzung der Evaluationsergebnisse zu fördern, besteht die Möglichkeit, die Interessen und Perspektiven relevanter Stakeholdergruppen bereits bei der Planung von Evaluationsvorhaben zu berücksichtigen und die verschiedenen Akteure in die folgenden Durchführungsschritte einzubinden bzw. diesen die jeweiligen Schritte zu kommunizieren (vgl. Owen, Rogers 1999: 121 f.; Weiss 1998; vgl. auch den Ansatz der Utilization-Focused Evaluation von Patton 1997). Damit lassen sich zwar nicht alle Konflikte vollkommen auflösen, es wird aber eine möglichst hohe Transparenz gegenüber allen beteiligten Akteuren geschaffen, was auch eine höhere Akzeptanz der Evaluation zur Folge haben dürfte (vgl. DeGEval 2002).

Evaluatoren sollten ebenfalls berücksichtigen, dass Evaluation letztlich immer nur eine von mehreren Entscheidungsgrundlagen darstellt, die bei politischen Entscheidungsprozessen Berücksichtigung finden (vgl. Rossi et al. 2004: 375; Weiss 1987: 48 ff.; Weiss 1998: 310 ff.). Neben rein sachbezogenen Argumenten sind immer auch andere Faktoren für die Entscheidungsfindung von Bedeutung, die meist außerhalb der Einflussosphäre von Evaluatoren liegen. Weiss (1987: 48 ff.) macht darauf aufmerksam, dass politische Entscheidungsträger unabhängig von ihrer Einstellung zu den Ergebnissen einer Evaluation immer auch Akteure innerhalb ihres d.h. des politischen Systems sind, welches ihre Entscheidungen beeinflusst:

„[...] policymakers [...] are members of a policymaking system that has its own values and its own rules. Their model of the system, its boundaries and pivotal components, goes far beyond concern with program effectiveness. Their decisions are rooted in all the complexities of the democratic decision-making process: the allocation of power and authority, the development of coalitions, and the trade-offs with interest groups, professional guilds, and salient publics. How well a program is doing may be less important than the position of the congressional committee chairman, the political clout of its supporters, or demands on the budget. A considerable amount of ineffectiveness may be tolerated if a program fits well with prevailing values, if it satisfies voters, or if it pays off political debts“ (Weiss 1987: 53 f.).

6 Fallstudie

Seit der Veröffentlichung von Avedis Donabedian „Evaluating the Quality of Medical Care“ orientiert man sich auch in der Medizin und in den Gesundheitswissenschaften bei der Evaluation unterschiedlichster Maßnahmen (Gesundheitsförderung, ärztliche Therapien, Pflege) an der von ihm konzipierten Unterscheidung zwischen Strukturqualität, Prozessqualität und Ergebnisqualität. Eine schwedische Studie (Kunkel et al. 2007) hat nun erstmals empirisch und an einer größeren Studie von 386 Kliniken und stationären Versorgungseinrichtungen überprüft, inwieweit diese drei unterschiedlichen Bewertungsebenen miteinander zusammenhängen. Bedeutsam ist diese Frage insofern, als vielen Studien aus finanziellen oder organisatorischen Gründen die Ergebnisqualität (also z.B. in der medizinischen Versorgung der direkte Therapieerfolg) nicht gemessen wird und daher Indikatoren der anderen beiden Bewertungsebenen hilfsweise herangezogen werden.

Von den in der Studie etwa 600 angeschriebenen Einrichtungen beteiligten sich 75 Prozent und beantworteten den Fragenbogen, der unterschiedliche Fragen zu den drei Bewertungsebenen enthielt. So wurde zur Strukturqualität etwa gefragt, ob Mitarbeiter und Leiter genug Zeit finden, um sich mit Systemen zur Qualitätsverbesserung zu beschäftigen, ob es eine Dokumentation zur Qualitätssicherung gibt usw. Im Abschnitt Prozessqualität zielten Fragen darauf, ob Mitarbeiter unerwünschte Zwischenfälle auch tatsächlich mitteilen, ob Mitarbeiter auch genügend Unterstützung von Vorgesetzten und Kollegen erhalten, wenn sie organisatorische Verbesserungen einführen möchten. Und im Fragenblock zur Ergebnisqualität sollte angegeben werden, ob die Klinik exakt formulierte Qualitätsziele hat, ob die Erreichung dieser Ziele regelmäßig geprüft wird, ob dies auch dokumentiert wird usw.

In der Analyse der Antworten aus den Kliniken wurde dann festgestellt, dass der Zusammenhang zwischen den drei Bewertungskriterien recht hoch ausfällt. Ausgedrückt in Korrelationen (Pearsons r) heißt dies (vgl. Kunkel et al. 2007: 5):

- Struktur und Prozess korrelierten mit 0.72,
- Struktur und Ergebnis korrelierten mit 0.60,
- Prozess und Ergebnis korrelierten eher schwach mit 0.20,
- der Zusammenhang von Prozess und Struktur (gleichzeitig) mit Ergebnis betrug 0.75.

Die Ergebnisse der Studie sind wahrscheinlich eine nachträgliche Legitimation für viele Evaluations-Projekte, die sich aufgrund zeitlicher Einschränkungen oder organischer Schwierigkeiten nur mit Struktur- oder Prozessdaten beschäftigen konnten. Gleichwohl sollten die Befunde nicht ohne Weiteres als methodische Richtschnur genommen werden. In zweierlei Hinsicht lässt die Befragungsstudie Wünsche offen. Zum einen wurden keine externen Bewertungen vorgenommen, sondern Aussagen der jeweiligen Mitarbeiter. Hier ist jedoch davon auszugehen, dass solche Urteile auch eine hohe Konsistenz aufweisen. Es ist wenig wahrscheinlich, dass ein Mitarbeiter die Ergebnisqualität überaus positiv, die Strukturqualität jedoch weniger gut einschätzt. Von daher sind die hohen Korrelationen zumindest teilweise ein Effekt der Methodik durch Befragung. Zum anderen wurden keine tatsächlichen „Outcome“-Indikatoren erfasst (Daten zur Versorgungsqualität: Mortalität, unerwünschte Vorkommnisse und Komplikationen, medizinische Messwerte, Wiedereinweisungs-Quoten usw.), sondern lediglich Angaben, ob hierzu bestimmte Voraussetzungen vorhanden sind (Qualitätsmanagement, Dokumentation) (vgl. Kunkel et al. 2007).

Zusammenfassung

Die Evaluationsforschung ist eine in Deutschland junge Wissenschaft. Ihre Methoden und Instrumente finden im öffentlichen Dienstleistungssektor erst seit einigen Jahren vermehrte Anwendung. Dabei sollte darauf geachtet werden, Evaluationen im sozialwissenschaftlichen Sinne nicht mit den betriebswirtschaftlichen Methoden der Leistungsmessung zu verwechseln. Auch wenn teilweise Gemeinsamkeiten bestehen, überwiegen doch die strukturellen Unterschiede und grundlegenden Zielsetzungen.

Zu Beginn jeder Evaluation gilt es, die Frage zu beantworten, wessen Interessen die Evaluation dienen soll. Ein klar definierter Evaluationsgegenstand (z. B. politische Interventionsmaßnahmen, Projekte, Programme, Organisationen, Personen, Prozesse, Produkte etc.) ist dabei ebenso wichtig, wie die Bewertung anhand präzise auf den zu evaluierenden Sachverhalt ausgerichteter Kriterien.

Die Evaluationsforschung bewegt sich damit immer in einem Spannungsfeld zwischen Wissenschaftlichkeit und Nützlichkeit. Dabei stellt die zunehmende Komplexität von Programmen, die Vielfalt der mit ihnen verbundenen Zielerwartungen, die Forderung nach stärkerer Einbeziehung der Stakeholder etc. die Evaluationsforschung vor neue Herausforderungen. Der Versuch einer Kategorisierung der verschiedenen Evaluationsansätze ist, trotz vielfältiger Versuche, dabei bisher ohne Erfolg geblieben.

Gleichwohl leisten sozialwissenschaftliche Evaluationen einen unverzichtbaren und wesentlichen Beitrag zur effizienten und effektiven Allokation öffentlicher Mittel. Es bleibt zu hoffen, dass Sie auch in der Bundesrepublik deutlich stärker, als dies bisher – gerade auch im Bereich der Gesundheitspolitik – der Fall ist, tatsächlich zur Grundlage politischer Entscheidungsfindung gemacht wird.

Literaturverzeichnis

Alkin, M.C. (1990): Debates on Evaluation. Newbury Park u.a.: Sage.

AQUA-Institut für angewandte Qualitätsförderung und Forschung im Gesundheitswesen GmbH (2013): Evaluation von fünf Ersatzkassen-Hausarztverträgen auf Grundlage des §73b SGB V. URL: http://www.aqua-institut.de/aqua/upload/CONTENT/Projekte/2013-09-16_HzV-Evaluation_vdek.pdf [Stand: 10.2.2013].

Arbeitsgemeinschaft der Spitzenverbände der Krankenkassen (2008a): Gemeinsame und einheitliche Evaluationsverfahren zu §20 SGB V der Spitzenverbände der Krankenkassen. Anwenderhandbuch Evaluation Teil 1 : Evaluation des Individuellen Ansatzes (Bewegungs-, Ernährungs- und Stressreduktionskurse) in der Fassung vom Juni 2008. URL: http://www.gkv-spitzenverband.de/media/dokumente/krankenversicherung_1/praevention__selbsthilfe__beratung/praevention/praevention_evaluation/individualansatz_kurse/Praev_Handbuch_1_Individualansatz_2008-06.pdf [Stand: 10.2.1014].

Arbeitsgemeinschaft der Spitzenverbände der Krankenkassen (2008b): Gemeinsame und einheitliche Evaluationsverfahren zu §20 SGB V der Spitzenverbände der Krankenkassen. Anwenderhandbuch Evaluation Teil 2: Evaluation von betrieblicher Gesundheitsförderung Fassung vom Juni 2008. URL: http://www.gkv-spitzenverband.de/media/dokumente/krankenversicherung_1/praevention__selbsthilfe__beratung/praevention/praevention_evaluation/betriebl_gesundheitsfoerderung/Praev_Eva_Handbuch_2_BGF_2008-06.pdf [Stand: 10.2.1014].

Beckmann, C. (Hrsg.) (2004): Qualität in der sozialen Arbeit: zwischen Nutzerinteresse und Kostenkontrolle. Wiesbaden: Verlag für Sozialwissenschaft.

Beywl, W.; Widmer, T. (2000): Handbuch der Evaluationsstandards. Die Standards des „Joint Committee for Educational Evaluation“. Opladen: Leske + Budrich.

Bogumil, J. (2004): Probleme und Perspektiven der Leistungsmessung in Politik und Verwaltung. In: Kuhlmann, S.; Bogumil, J.; Wollmann, H. (Hrsg.) (2004): Leistungsmessung und -vergleich in Politik und Verwaltung. Konzepte und Praxis. Wiesbaden: Verlag für Sozialwissenschaft: 392–398.

Bortz, J.; Döring, N. (2002): Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler. Berlin u.a.: Springer.

Braun, B.; Buhr, P.; Klinke, S.; Müller, R.; Rosenbrock, R. (2010): Pauschalpatienten, Kurzlieger und Draufzähler – Auswirkungen der DRGs auf Versorgungsqualität und Arbeitsbedingungen im Krankenhaus. Bern: Huber.

Brandtstädter, J. (1990): Evaluationsforschung: Probleme der wissenschaftlichen Bewertung von Interventions- und Reformprojekten. In: Zeitschrift für Pädagogische Psychologie 4/4: 215–228.

BR-Drs. 381/2/94: Verordnung zur Neuordnung des Pflegesatzrechtes. URL: <http://dipbt.bundestag.de/doc/brd/1994/D381+94.pdf> [Stand: 23.12.2013].

Brüggemeier, M. (2004): Von der Kunst, erfolgreich zu scheitern – Wirkungsorientiertes Controlling in öffentlichen Verwaltungen. In: Kuhlmann, S.; Bogumil, J.; Wollmann, H. (Hrsg.): Leistungsmessung und Leistungsvergleich in Politik und Verwaltung. Wiesbaden: Verlag für Sozialwissenschaft: 374–391.

Bühler, B. M. (2002): Von Outputs zu Outcomes. Internationale Erfahrungen mit outcome-orientierter Steuerung. In: Verwaltung & Management 8/5: 273–278.

Bussmann, W. et al (Hrsg.) (1997): Einführung in die Politikevaluation. Basel: Helbing und Lichtenhahn.

Chelimsky, E. (1995): New dimensions in evaluation. In: World Bank Operations Evaluations Department (OED): Evaluation and Development: proceedings of the 1994 World Bank Conference. Washington D.C.: 3–11.

Clemens, W.; Strübing, J. (Hrsg.) (2000): Empirische Sozialforschung und gesellschaftliche Praxis. Bedingungen und Formen angewandter Forschung in den Sozialwissenschaften. Opladen: Leske + Budrich.

DeGEval – Deutsche Gesellschaft für Evaluation e.V. (2002): Standards für Evaluation. Köln. URL: <http://www.degeval.de/degeval-standards/standards> [Stand: 23.12.2013].

DZVhÄ – Deutscher Zentralverein homöopathischer Ärzte (2013): Studie zur Evaluation des Integrierten Versorgungsvertrages Homöopathie mit der Techniker Krankenkasse gestartet. URL: <http://dzvhae-homoeopathie-blog.de/?p=5898> [Stand: 10.2.2014].

DKI – Deutsches Krankenhausinstitut e.V.; I+G Gesundheitsforschung (vormals Infratest) (1999): Begleitforschung zur Bundespflegesatzverordnung 1995 – Abschlussbericht. Düsseldorf: Deutsches Krankenhausinstitut.

Fitzpatrick, J. L.; Sanders, J. R.; Worthen, B. R. (2004): Program Evaluation: Alternative Approaches and Practical Guidelines. Boston u.a.: Pearson.

Gephart, W. J. (1978): The facets of the evaluation process: A starter set. Unpublished manuscript, presented at the Annual Meeting of the American Educational Research Association in Toronto.

Kieser, A.; Kubicek, H. (1992): Organisation. Berlin u.a.: de Gruyter.

Kieser, A.; Walgenbach, P. (2003): Organisation. 4. überarbeitete und erweiterte Auflage. Stuttgart: Schäffer-Poeschel.

Kieser, A. (2002): Organisationstheorien. Stuttgart: Kohlhammer.

Klausegger, C.; Scharitzer, D. (1998): Instrumente für das Qualitätsmanagement in NPOs. In: Eschenbach, R. (Hrsg.): Führungsinstrumente für die Nonprofit-Organisation. Stuttgart: Schäffer-Poeschel.

Kromrey, H. (1995): Empirische Sozialforschung. Modelle und Methoden der Datenerhebung und Datenverarbeitung. Opladen: Leske und Budrich.

Kromrey, H. (2001): Evaluation – Ein vielschichtiges Konzept. Begriff und Methodik von Evaluierung und Evaluationsforschung. Empfehlungen für die Praxis. In: Sozialwissenschaften und Berufspraxis 24/2: 105–131.

Kromrey, H. (2002): Empirische Sozialforschung. Opladen: Leske u. Budrich.

Kuhlmann, S. (2004a): Einleitung: Leistungsmessung und Evaluation in Politik und Verwaltung. In: Kuhlmann, S.; Bogumil, J.; Wollmann, H. (Hrsg.): Leistungsmessung und -vergleich in Politik und Verwaltung: Konzepte und Praxis (Stadtforschung aktuell Bd. 96). Wiesbaden: Verlag für Sozialwissenschaften: 11–20.

Kuhlmann, S. (2004b): Interkommunaler Leistungsvergleich in Deutschland: Zwischen Transparenzgebot und Politikprozess. In: Kuhlmann, S.; Bogumil, J.; Wollmann, H. (Hrsg.): Leistungsmessung und -vergleich in Politik und Verwaltung: Konzepte und Praxis (Stadtforschung aktuell Bd. 96). Wiesbaden: Verlag für Sozialwissenschaften: 94–122.

Kuhlmann, S.; Bogumil, J.; Wollmann, H. (Hrsg.) (2004): Leistungsmessung und -vergleich in Politik und Verwaltung: Konzepte und Praxis (Stadtforschung aktuell Bd. 96). Wiesbaden: Verlag für Sozialwissenschaften.

- Kunkel, S.; Rosenquist, U.; Westerling, R. (2007): The structure of quality systems is important to the process and outcome, an empirical study of 386 hospital departments in Sweden. In: BMC Health Services Research 7: 104. URL: <https://www.biomedcentral.com/content/pdf/1472-6963-7-104.pdf>
[Stand: 10.2.2014].
- Langnickel, H. (2003): Das EFQM-Modell für Excellence – Der Europäische Qualitätspreis. In: Boeßenecker, K.-H.: Qualitätskonzepte in der sozialen Arbeit. Weinheim: Beltz: 38–47.
- Lee, B. (2006): Theories of Evaluation. In Stockmann, R. (Hrsg.): Evaluationsforschung. Grundlagen und ausgewählte Forschungsfelder. 3. Auflage. Münster: Waxmann: 137–176.
- Mertens, D. M. (1998): Research methods in education and psychology: Integrating diversity with quantitative and qualitative approaches. Thousand Oaks, CA: Sage.
- Meyer, W. (2007): Messen: Indikatoren – Skalen – Indizes – Interpretationen. In: Stockmann, R. (Hrsg.): Handbuch der Evaluation. Eine praktische Handlungsanleitung. Münster: Waxmann: 195–222.
- Mohr, H.-W (1977): Bestimmungsgründe für die Verbreitung von neuen Technologien. Berlin: Duncker & Humblot.
- Müller-Jentsch, W. (2003): Organisationssoziologie. Eine Einführung. Frankfurt a.M.: Campus.
- Niehoff, J.-U., Braun, B. (2010): Handwörterbuch Sozialmedizin und Public Health. Baden-Baden: Nomos.
- Owen, J. M.; Rogers, P. J. (1999): Program Evaluation. Forms and Approaches. London u. a.: Sage.
- Patton, M. Q. (1997): Utilization – Focused Evaluation: The New Century Text. 3. ed. Thousand Oaks, London, New Delhi: Sage.
- Pede, L. (2000): Wirkungsorientierte Prüfung der öffentlichen Verwaltung. Bern: Haupt.
- Pollitt, C. (2000): Public management reform: a comparative analysis. New York: Oxford Univ. Press.
- Raidl, M. (2001): Qualitätsmanagement in Theorie und Praxis – eine Verbindung von Instrumenten der empirischen Sozialforschung und der Einsatz und Nutzen für die Praxis. Eine empirische Studie in einer süddeutschen Privatklinik. München u. a.: Hampp.
- Raizen, S. A.; Rossi, P. H. (1981): Program evaluation in education: When? How? To what ends? Washington, D.C.: National Academic Press.
- Reiner-Theisen, I.; Braatz, F.; Winkler, H.-P. (2013): 8. Qualitätsbericht zum Disease Management Programm Brustkrebs Berlin. URL: http://www.aok-gesundheitspartner.de/imperia/md/gpp/bln/dmp/bk/bln_bk_qs_2012.pdf
[Stand: 10.2.2014].
- Rogers, E. M. (1995): Diffusion of innovations. 4. ed. New York.
- Rossi, P. H.; Freeman, H. E.; Hofmann, G. (1988): Programm Evaluation: Einführung in die Methoden angewandter Sozialforschung. Stuttgart: Enke.
- Rossi, P. H.; Lipsey, M. W.; Freeman, H. E. (2004): Evaluation: a systematic approach. 7. ed. Thousand Oaks, Calif.: Sage.
- Scott, R. (2003): Organizations: Rational, Natural, and Open Systems. 5. ed. New Jersey: Prentice Hall.
- Scriven, M. (1991): Evaluation Thesaurus. 1. ed. Newbury Park u. a.: Sage.

Shadish, W. R.; Cook, T. D.; Leviton, L. C. (1991): Foundations of Program Evaluation: Theory and Practice. London: Sage.

Siebert, G.; Kempf, S. (2002): Benchmarking. Leitfaden für die Praxis. Wien: Hanser.

Smith, P. (Hrsg.) (1996): Measuring Outcome in the Public Sector. London, Bristol: Taylor & Francis.

Stockmann, R. (1992): Die Nachhaltigkeit von Entwicklungsprojekten. Eine Methode zur Evaluierung am Beispiel von Berufsbildungsprojekten. Opladen: Westdeutscher Verlag.

Stockmann, R. (1996): Überlegungen zur Gründung eines Zentrums für die Evaluation der Entwicklungszusammenarbeit (ZEEZ). Auszug aus einem Vortrag anlässlich einer Konferenz des Deutschen Übersee-Instituts zur „Erfolgskontrolle in der entwicklungspolitischen Zusammenarbeit“ am 10. und 11. Juni 1996 in Hamburg.

Stockmann, R. (2006): Evaluation und Qualitätsentwicklung. Eine Grundlage für wirkungsorientiertes Qualitätsmanagement. Münster: Waxmann.

Stockmann, R. (2007): Handbuch zur Evaluation – Eine praktische Handlungsanleitung. Münster: Waxmann.

Stufflebeam, D. L. (2000): Foundational Models for 21st Century Program Evaluation. In: Stufflebeam, D. L.; Madaus, G. F.; Kellaghan, T. (Hrsg.): Evaluation Models. Viewpoints on Educational and Human Services Evaluation. Boston u.a.: Kluwer: 33–83.

Techniker Krankenkasse (2012): Disease Management Programm Brustkrebs der Techniker Krankenkasse in der Region Hamburg. Evaluationsbericht zum 30.9.2012. URL: <https://www.tk.de/centaurus/servlet/contentblob/34218/Datei/111421/Evaluationsbericht%20DMP%20Brustkrebs%20zum%2030.09.2012.pdf> [Stand: 10.2.2014].

Thompson, J. D. (1967): Organizations in Action. New York: McGraw-Hill.

Vedung, E. (2000): Evaluation Research and Fundamental Research. In: Stockmann, R. (Hrsg.): Evaluationsforschung. Opladen: Leske + Budrich: 103–127.

Weiss, C. H. (1987): Where Politics and Evaluation Meet. In: Palumbo, D. (Hrsg.): The Politics of Program Evaluation. Newbury Park (CA): Sage: 47–70.

Weiss, C. H. (1998): Evaluation. 2. ed. Upper Saddle River (NJ): Prentice Hall Inc.

Widmer, T. (2006): Qualität der Evaluation – Wenn Wissenschaft zur praktischen Kunst wird. In: Stockmann, R. (Hrsg.): Evaluationsforschung. Grundlagen und ausgewählte Forschungsfelder. Münster: Waxmann: 85–112.

Wissenschaftszentrum Berlin für Sozialforschung (2014): Wandel von Medizin und Pflege im DRG-System (WAMP). URL: <http://www.wzb.eu/de/forschung/beendete-forschungsprogramme/public-health/projekte/wandel-von-medizin-und-pflege-im-drg-system> [Stand: 10.2.2014].

Wollmann, H. (2002): Verwaltungspolitik und Evaluierung. Ansätze, Phasen und Beispiele im Ausland und in Deutschland. In: Zeitschrift für Evaluation. Bd. 1/2002: 75–100.

Wollmann, H. (Hrsg.) (2003): Evaluation in Public Sector: Reform Concepts and Practice in International Perspective. Cheltenham: Edward Elgar Publishing Limited.

Wottawa, H.; Thierau, H. (1990): Lehrbuch Evaluation. Bern: Huber.

Wottawa, H.; Thierau, H. (1998): Lehrbuch Evaluation. 2. Auflage. Bern: Huber.

Wottawa, H.; Thierau, H. (2003): Lehrbuch Evaluation. 3. korrigierte Auflage. Bern: Huber.

