

Evaluation von Präventionsprogrammen

Waldemar Mittag und Matthias Jerusalem

Die Nachfrage nach psychologischer Prävention und Gesundheitsförderung hat in den letzten Jahren rapide zugenommen. Inzwischen gibt es unzählige Programme und Maßnahmen zur Gesundheitsförderung, die mit hohem Sach- und Personalaufwand und hohen Kosten verbunden sind, aber im Hinblick auf ihre theoretische Fundierung, ihre praktische Umsetzung und ihre tatsächliche Wirksamkeit kaum wissenschaftlich evaluiert werden. Angesichts der immer knapper werdenden finanziellen Mittel und der stetig steigenden Kosten im Gesundheitswesen wird in den letzten Jahren zunehmend die Evaluation und Qualitätssicherung von Präventionsprogrammen gefordert. Das vorliegende Kapitel führt in das Forschungsgebiet der Evaluation ein und liefert einen Überblick über zentrale Themen der Evaluationsforschung. Zunächst werden die Begriffe „Evaluation“ und „Evaluationsforschung“ definiert, und es werden wichtige Grundlagen und Konzepte der Evaluationsforschung vorgestellt. Anschließend wird eine Unterscheidung der Evaluationsforschung in drei Hauptarten vorgeschlagen, die sich aus den einzelnen Stadien ergibt, die ein Programm von der Entwicklung über die Umsetzung bis hin zu seinen möglichen Wirkungen durchläuft. Diese Unterscheidung dient schließlich als heuristischer Rahmen, um in den nachfolgenden Abschnitten mit Bezug auf konkrete und fiktive Präventionsbeispiele zu verdeutlichen, welche spezifischen Aufgaben, Fragen und Probleme jeweils bei der Evaluation der Programmkonzeption, der Programmdurchführung und der Programmwirkung zu beachten sind.

Begriffe, Konzepte und Grundlagen der Evaluation

Zum Begriff „Evaluation“ findet man in der einschlägigen Literatur sehr unterschiedliche Vorstellungen, Beschreibungen und Definitionsversuche. Ein Grund dafür mag darin liegen, daß die Evaluationsforschung eine vergleichsweise junge wissenschaftliche Teildisziplin darstellt, in der bis heute divergierende Auffassungen im Hinblick auf bedeutsame Evaluationsfragen, auf Standards und Begriffe nebeneinander existieren. Eine kurze, relativ weitreichende Begriffsbestimmung von **Evaluation** stammt von Scriven (1967), der Evaluation als jegliche Art der Festsetzung des Wertes einer Sache versteht. Diese Beschreibung läßt allerdings offen, was den Gegenstandsbereich von Evaluationen ausmacht und mit welchen Methoden bzw. Techniken evaluiert werden soll. Folgt man beispielsweise Cook und Matt (1990), so läßt sich hinsichtlich des Gegenstandsbereiches von Evaluationen feststellen, daß prinzipiell alles evaluiert werden kann, einschließlich der Evaluation selbst. Ganz allgemein kann Evaluation als ziel- und zweckorientierter Bewertungs-

prozeß verstanden werden. Im Gesundheitsbereich geht es dabei vor allem um die Bewertung gesundheitsbezogener Maßnahmen oder Programme. Die Verwendung des Begriffes Evaluation impliziert jedoch nicht unbedingt, daß im Rahmen des Bewertungsprozesses wissenschaftliche Methoden oder Techniken verwendet werden (vgl. Wottawa & Thierau, 1990).

Im Gegensatz zu diesem sehr weitreichenden Verständnis von Evaluation hat der Begriff „**Evaluationsforschung**“ (evaluation research) eine engere Bedeutung. Obwohl beide Begriffe in der Literatur häufig synonym verwendet werden (z. B. Rossi & Freeman, 1993), umfaßt Evaluationsforschung nur solche Bewertungsprozesse, in denen systematisch wissenschaftliche Forschungsmethoden verwendet werden (vgl. Lösel & Nowack, 1987; Nöldner, 1990; Suchman, 1967). Folgt man dieser Begriffsbestimmung, so kann Evaluationsforschung in der Gesundheitspsychologie als „systematische Anwendung sozialwissenschaftlicher Forschungsmethoden zur Bewertung der Konzeption, Ausgestaltung, Umsetzung und des Nutzens“ von präventiven Maßnahmen und Programmen zur Erhaltung und Förderung der Gesundheit definiert werden (vgl. Rossi & Freeman, 1993, S. 5).

Die Definition macht deutlich, daß sich Evaluationsforschung auf verschiedene Entwicklungsphasen oder Stadien eines Programmes beziehen und von daher unterschiedliche Aspekte und Ziele zum Gegenstand haben kann. Deshalb ist es sinnvoll, zwischen verschiedenen **Arten der Evaluationsforschung** zu unterscheiden. Eine erste, bis heute sehr verbreitete Unterscheidung zwischen formativer und summativer Evaluation wurde bereits von Scriven (1967) eingeführt.

- Eine **formative Evaluation** richtet sich vornehmlich auf die Entwicklung, Konzeption, Ausgestaltung und Umsetzung eines Programmes. Dabei geht es im wesentlichen darum, verbesserungsbedürftige Programmelemente zu identifizieren und so zur Optimierung des Programmes beizutragen. Formative Evaluationen sind insbesondere bei der Planung, Ausgestaltung und Optimierung innovativer Programme in der Test- oder Vorlaufphase nützlich. Aber auch bei der Feinabstimmung bereits etablierter, laufender Programme können formative Evaluationen hilfreiche Hinweise zur Verbesserung liefern.
- Im Mittelpunkt einer **summativen Evaluation** steht dagegen die Bewertung der erzielten Wirkungen und des Nutzens eines Programmes. Das wesentliche Ziel liegt in der Feststellung aller Programmwirkungen, d. h. neben den geplanten Wirkungen sind auch unerwartete und negative Effekte zu berücksichtigen. Sinnvoll und nützlich sind summative Evaluationen vor allem bei bereits laufenden, etablierten Programmen, wenn es z. B. um die Entscheidung zwischen verschiedenen Programmalternativen und/oder die Weiterführung oder Einstellung einzelner Programme geht.

Betrachtet man die Unterscheidung zwischen formativer und summativer Evaluation im Hinblick auf die einzelnen Stadien, die ein Programm von der Entwicklung über die Umsetzung bis hin zu seinen möglichen Wirkungen durchläuft, so bezieht sich die formative Evaluation auf die Beurteilung der Programmkonzeption und Programmdurchführung, während die summative Evaluation vornehmlich die Bewertung der Programmwirkungen zum Ziel hat. Angesichts der verschiedenen Aufgaben und Ziele, die mit der Evaluation dieser Stadien jeweils verbunden sind, kann man

auch eine Unterscheidung zwischen „**drei Hauptarten der Evaluationsforschung**“ vornehmen, wie sie beispielsweise von Rossi und Freeman (1993; Rossi, Freeman & Hofmann, 1988) vorgeschlagen wurde und in Abbildung 1 dargestellt ist.

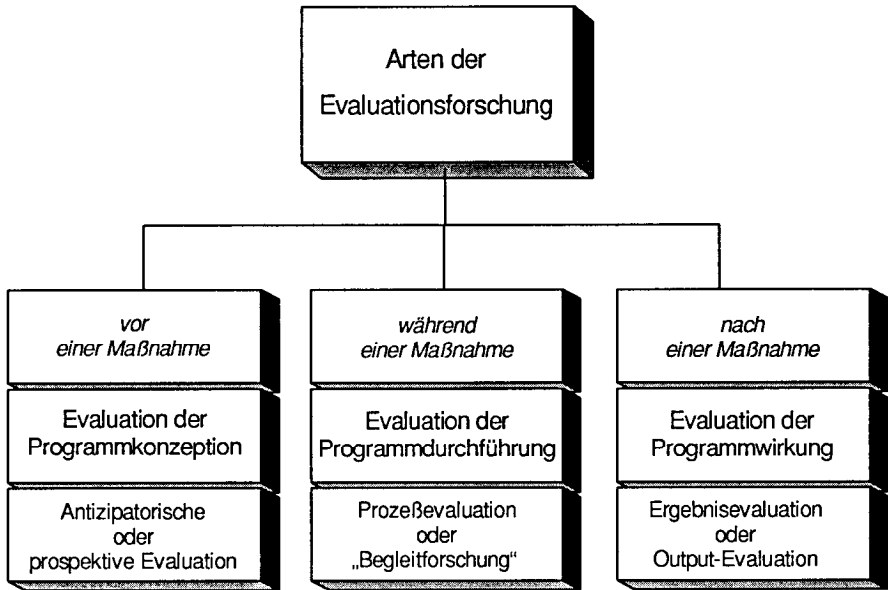


Abbildung 1. **Hauptarten der Evaluationsforschung.**

1. Bei der **Evaluation der Programmkonzeption** geht es darum, die Konzeption, Ausarbeitung und Ausgestaltung von geplanten Maßnahmen oder Programmen bereits vor deren Durchführung zu bewerten. Dies wird auch **als antizipatorische oder prospektive Evaluation** bezeichnet, da zu diesem Zeitpunkt in der Regel nur sehr wenige oder gar keine empirischen Hinweise zur Wirksamkeit des Programmes vorliegen. Gleichwohl ist es wichtig, bereits in diesem Stadium grundlegende Fragen über die Problem- und Zielbestimmung, die Angemessenheit und Umsetzbarkeit der geplanten Maßnahmen, die Kohärenz zwischen Programmzielen und Programmaßnahmen sowie zu Kosten-Nutzen-Gesichtspunkten anzugehen.
2. Eine **Evaluation der Programmdurchführung** hat die kontinuierliche Überprüfung der Umsetzung und Ausführung der geplanten Programmaßnahmen zum Ziel. Diese Form der **Prozeßevaluation** oder „**Begleitforschung**“ (vgl. Dlugosch & Wottawa, 1994; Rossi & Freeman, 1993) soll insbesondere Fehlentwicklungen oder unerwünschte Nebenwirkungen aufdecken und beseitigen helfen, durch die unter Umständen die Konzeption und Wirkung eines Programmes in Frage gestellt werden könnte.

3. Die **Evaluation der Programmwirkungen** beinhaltet die Bewertung der Ergebnisse eines Programmes nach dessen Durchführung. Dieser Schritt wird mitunter auch **als Ergebnisevaluation** oder **Outputevaluation** bezeichnet. Ergebnisevaluationen sollten nach Möglichkeit nicht allein auf die Feststellung der Programmwirkungen beschränkt bleiben, sondern darüber hinaus auch Kosten-Effizienz-Analysen mit einschließen (siehe unten).

Die Unterscheidung von Evaluationsarten in Abhängigkeit vom Entwicklungsstadium eines Programmes bzw. vom Zeitpunkt der Evaluation (vgl. Dlugosch & Wottawa, 1994) ist insoweit nützlich, als die jeweils spezifischen Ziel- und Aufgabenstellungen eines Evaluationsvorhabens transparenter werden. Unabhängig davon sollte die Evaluation eines Programmes nach Möglichkeit alle drei Evaluationsarten umfassen. Ein solcher übergreifender Evaluationsprozeß, der von der Konzeptentwicklung über die Ablaufplanung und Implementation bis hin zur Wirkungs- und Nutzenanalyse eines Programmes reicht, wird in der Literatur **als umfassende Evaluation** (comprehensive evaluation; Rossi & Freeman, 1993) oder auch **Programmevaluation** bezeichnet (vgl. Bengel & Koch, 1988; Nöldner, 1990; Wittmann, 1985).

Betrachtet man die in den letzten Jahren publizierten Studien zur Evaluation von Gesundheitsprogrammen genauer, so ist festzustellen, daß diese häufig auf die Beurteilung der Programmwirkungen beschränkt bleiben, während Hinweise auf eine Evaluation der Programmkonzeption oder Programmdurchführung vergleichsweise seltener oder nur spärlich zu finden sind (vgl. Dlugosch & Wottawa, 1994). Angesichts der unterschiedlichen theoretischen Modellvorstellungen zu den personalen, sozialen und umweltbezogenen Bedingungen von Gesundheit (siehe z. B. Schwarzer, 1996; Schwenkmezger & Schmidt, 1994) ist es sinnvoll und notwendig, daß Maßnahmen der Programmentwicklung, Programmdurchführung und Wirkungskontrolle aufeinander abgestimmt werden und als verzahnte Aktivitäten eines umfassenden Evaluationsprozesses durchgeführt werden. Je besser die einzelnen Evaluationsschritte aufeinander abgestimmt werden, je fundierter theoretisch und methodisch geplant, durchgeführt und ausgewertet wird, um so eher können theoretisch interpretierbare Ergebnisse von gesundheitspsychologischen Präventionsprogrammen erwartet werden, aus denen sich schließlich sinnvolle Implikationen für Weiterentwicklungen in Theorie und Praxis ableiten lassen (vgl. Jerusalem & Mittag, 1994).

Natürlich wird sich diese Idealvorstellung im praktischen Alltag nicht immer realisieren lassen. Welche Aufgaben und Probleme in einem konkreten Evaluationsvorhaben mit welchen Methoden und Techniken zu lösen sind, hängt von einer Reihe von Faktoren ab. Neben dem jeweiligen Entwicklungsstadium eines Programmes bzw. dem Zeitpunkt der Evaluation spielen dabei auch die Zielrichtung, das Objekt, die möglichen Bearbeitungsformen einer Evaluation sowie die aktuell vorliegenden Rahmenbedingungen eine wichtige Rolle (vgl. Dlugosch & Wottawa, 1994). So erfordert beispielsweise die Evaluation innovativer Programme andere Methoden und Techniken als die Evaluation etablierter Programme oder die Modifikation und Feinabstimmung bereits laufender Projekte. Von daher benötigt jedes Programm eine speziell darauf zugeschnittene Evaluationsstrategie. Dennoch lassen sich einige

grundsätzliche Evaluationsaufgaben hervorheben, die nach Möglichkeit jede Evaluation von Gesundheitsprogrammen mit einschließen sollte: Problembestimmung und Feststellung des Programmbedarfs, Ziel- und Zielgruppenbestimmung, Angemessenheit der geplanten oder durchgeführten Maßnahmen, Überwachung der Programmausführung und Feststellung der Programmreichweite, Bewertung der intendierten Wirkungen einschließlich möglicher Nebenwirkungen sowie die Bestimmung der Nützlichkeit der Programmaßnahmen (vgl. Bengel & Koch, 1988; Nöldner, 1990). Diese zentralen Evaluationsaufgaben werden im folgenden genauer betrachtet.

Evaluation der Programmkonzeption

Die wichtigsten Aufgaben bei der Entwicklung und Konzipierung neuer Maßnahmen und Programme, aber auch bei der Modikation bereits laufender Programme sind die Problembestimmung, die Feststellung des Programmbedarfs, die Zielbestimmung, die Ableitung und Konzipierung geeigneter und umsetzbarer Programmaßnahmen sowie die prospektive Analyse der zu erwartenden Wirkungen und Nebenwirkungen.

Durch **die Problembestimmung** und **die Feststellung des Programmbedarfs** soll eine möglichst genaue Beschreibung bzw. ein Nachweis über Art, Ausmaß und Verteilung des Problems sowie über den zu erwartenden Bedarf an neuen oder zusätzlichen Maßnahmen erbracht werden (vgl. Nöldner, 1990; Rossi et al., 1988). Dies ist bedeutsam, weil die Vorstellungen über das Ausmaß und die Verteilung eines Problems bei den an der Programmplanung beteiligten und interessierten Personengruppen (Geld- und Auftraggeber, Projektleiter und Mitarbeiter, Evaluationsforscher etc.) weit auseinander liegen können. Sind solche Vorstellungen darüber hinaus ungenau, kann es - je nach Ausrichtung - zu einer Unter- oder Überschätzung des tatsächlichen Programmbedarfs kommen. Unangemessene und falsche Bedarfs-schätzungen wiederum können zu Problemen bei der Programmausführung, zu Verwischungen der Programmeffekte, aber auch zu unerwünschten und negativen Nebenwirkungen führen. Eine genaue Bedarfsprüfung kann solche Fehlentwicklungen in einem späteren Programmstadium verhindern.

Die **Zielbestimmung** ist eng mit der Problembestimmung und der Feststellung des Programmbedarfs verknüpft. Bei der Problembestimmung wird ein problematischer Ist-Zustand oder ein erwarteter, als negativ eingeschätzter Wird-Zustand festgelegt. Daran anknüpfend wird bei der Zielbestimmung ein positiv bewerteter Soll-Zustand formuliert. Als problematische und negative Ist-Zustände können im Rahmen gesundheitspsychologischer Interventionsprogramme zum Beispiel gesundheitliche Risikoverhaltensweisen bestimmt werden, wie etwa das Rauchverhalten, ein unangemessen hoher Alkohol- oder Medikamentenkonsum, unzureichende körperliche Aktivität, Fehlernährung und Übergewicht oder die mangelnde Nachfrage und Nutzung von Vorsorgeuntersuchungen. Zusätzlich sind im Zuge von Bedarfsprüfungen das Ausmaß und die Verteilung dieser gesundheitsriskanten Verhaltensweisen abzuschätzen. Auf der Grundlage von Problembestimmung und Bedarfsprüfung ist schließlich ein allgemeiner Zielzustand zu definieren, der z. B. darin liegen kann, in der Gesamt-

bevölkerung oder einer untersuchten Teilpopulation die Raucherquote zu senken oder die Nutzung von gesundheitlichen Vorsorgeuntersuchungen zu verbessern. Zur Zielbestimmung können außer Ist-Soll-Diskrepanzen auch Wird-Soll-Diskrepanzen herangezogen werden. Dies wird immer dann der Fall sein, wenn mit einem problematischen Zustand zu einem zukünftigen Zeitpunkt zu rechnen ist, falls nicht präventiv eingegriffen wird. So haben beispielsweise schulische Gesundheitsprogramme häufig zum Ziel, dem Aufbau sowie der Verfestigung gesundheitsriskanter Verhaltensweisen im Entwicklungsverlauf von Kindern und Jugendlichen präventiv entgegenzuwirken (vgl. Jerusalem, in diesem Band).

Bei der Zielbestimmung unterscheidet man ferner zwischen allgemeinen Zielen und Operationszielen (vgl. Rossi & Freeman, 1993). **Allgemeine Ziele** (goals) sind zunächst die von den beteiligten Personengruppen gewünschten Ergebnisse. Diese sind bei vielen Programmen anfangs sehr unspezifisch und vage. Darüber hinaus können die Zielvorstellungen der beteiligten Personengruppen stark divergieren. **Operationsziele** (objectives) stellen dagegen Operationalisierungen der allgemeinen Programmziele dar. Die Formulierung von Operationszielen dient dazu, die zunächst sehr allgemeinen Zielvorstellungen zu präzisieren und konkrete Maßnahmen abzuleiten. Bei gesundheitsbezogenen Präventionsprogrammen waren beispielsweise im Rahmen der Zielbestimmung die jeweilige theoretische Grundposition zu klären und aus dem zugrunde gelegten, gesundheitspsychologischen Modell spezifische Hypothesen bzw. entsprechende Zwischenziele abzuleiten. Als ein nützliches Hilfsmittel zur **Ableitung und Konzipierung von Programmaßnahmen** hat sich insbesondere im pädagogisch-psychologischen sowie im klinisch-psychologischen Bereich die Aufstellung von Ziel-Mittel-Bäumen erwiesen. Auf der Grundlage solcher Ziel-Mittel-Bäume lassen sich konkrete Maßnahmen ableiten, wobei zur Entscheidung für oder gegen eine Maßnahme bestimmte Kriterien wie z. B. Effektivitäts- und Ökonomiegesichtspunkte, die Manipulierbarkeit des in Frage stehenden Verhaltens und die Vereinbarkeit mit den jeweiligen Rahmenbedingungen sowie die Abwesenheit negativer Effekte zu berücksichtigen sind.

Schließlich ist im Rahmen der Zielbestimmung auch zu klären, ob die Programmmaßnahmen unmittelbar oder mittelbar auf die intendierte Zielgruppe gerichtet sein sollen (**direkte versus indirekte Zielgruppe**), ob **generalpräventiv** oder eher auf bestimmte **Risikogruppen** bezogen vorgegangen werden soll, und wie genau sich **die Ziel- bzw. Untersuchungseinheiten** eines Programmes bestimmen und abgrenzen lassen. Gesundheitspsychologische Maßnahmen und Programme sind vornehmlich auf Individuen gerichtet. Als Ziel- bzw. Untersuchungseinheiten können neben Individuen aber auch bestimmte Gruppen von Individuen sowie physische, betriebliche, geographische oder politische Einheiten in Frage kommen. So zielt beispielsweise betriebliche Gesundheitsförderung häufig auf bestimmte Arbeitsgruppen oder Abteilungen innerhalb eines Betriebes oder einer Verwaltung, und in groß angelegten Programmen wie der Deutschen Herz-Kreislauf-Präventionsstudie werden ausgewählte Segmente der Gesamtbevölkerung innerhalb bestimmter geographischer Regionen als Ziel- und Untersuchungseinheit ausgewählt (siehe z. B. v. Troschke et al., 1990). Diese Beispiele verdeutlichen, daß die Auswahl einer geeigneten Ziel- und Untersuchungseinheit vom jeweiligen Programm und Evaluationsprojekt abhängt. Die

Festlegung auf eine konkrete Analyseeinheit sollte jedoch möglichst frühzeitig erfolgen, da dies sowohl im Rahmen diagnostischer Voruntersuchungen zur Bestimmung des Programmbedarfs und der Ziele als auch für die nachfolgende Messung und Analyse der Programmwirkungen von zentraler Bedeutung ist.

Für die Festlegung der Ziel- und Untersuchungseinheiten ist auch die Unterscheidung zwischen direkten und indirekten Zielgruppen bedeutsam. Während **direkte Zielgruppen** unmittelbar mit den Programmaßnahmen behandelt werden (z. B. die medizinische Behandlung bestimmter Personengruppen), erhalten **indirekte Zielgruppen** die Programmaßnahmen mittelbar, d. h. auf indirektem Wege (vgl. Rossi et al., 1988). Bei indirekten Zielgruppen werden zunächst ausgewählte Vermittlungspersonen in Seminaren und Kursen mit den Programmaßnahmen vertraut gemacht und geschult, um diese anschließend an die intendierte Zielgruppe weiterzugeben. So können zum Beispiel im Rahmen der betrieblichen Gesundheitsförderung bestimmte Maßnahmen von ausgewählten Angestellten oder Abteilungsleitern an die übrige Belegschaft weitergegeben werden, und bei schulischen Präventions- und Gesundheitsprogrammen können Lehrer, Eltern, Gleichaltrige oder externe Personen als Programmvermittler in Frage kommen. Ist die Vermittlung und Durchführung der Programmaßnahmen durch solche Personengruppen geplant, sind zwei Aspekte zu beachten. Zum einen ist zu überlegen, ob und inwieweit bei der Entwicklung der Programmaßnahmen und der Evaluation der Programmkonzeption auf die Fachkompetenz der betreffenden Personen zurückgegriffen werden kann. Lehrer kennen den Schulalltag und ihre Klassen am besten und können mitunter hilfreiche Hinweise und Verbesserungsvorschläge liefern, wie Programmbausteine unter den gegebenen schulischen Rahmenbedingungen zu realisieren und im Hinblick auf die intendierten Programmwirkungen zu optimieren sind. Zum anderen ist zu berücksichtigen, daß die Wirksamkeit des Programmes im wesentlichen davon abhängt, in welcher Beziehung die trainierte Personengruppe zur indirekten Zielgruppe steht. Als nützlich hat sich in dieser Hinsicht der **Bezugsgruppen-Ansatz** erwiesen (vgl. Fuchs & Schwarzer, 1994; Rossi et al., 1988). Besonders im Rahmen schulischer Gesundheitsprogramme scheint der aktuellen Forschungslage zufolge ein Einsatz von Gleichaltrigen oder etwas älteren Kindern und Jugendlichen als Programmvermittler oder Multiplikatoren erfolgversprechend (**peer education-, peer leader- oder opinion leader-Konzept**). Verschiedene Evaluationsstudien konnten deutliche Programmwirkungen vor allem dann nachweisen, wenn die Maßnahmen von Gleichaltrigen durchgeführt wurden (vgl. Schaps, DiBartolo, Moskowitz, Palley & Churgin, 1981; Telch, Miller, Killen & Cooke, 1990), während in von Lehrern durchgeführten Interventionsklassen zum Teil keine oder sogar negative Verhaltenseffekte zu beobachten waren (Botvin, Baker, Filazzola & Botvin, 1990). Wie wichtig im Zusammenhang mit der Interpretation solcher Ergebnisse die Festlegung geeigneter Ziel- und Analyseeinheiten sein kann, wird beispielsweise in der Evaluationsstudie von Botvin et al. (1990) deutlich. Durch detaillierte Analysen konnten auch in von Lehrern angeleiteten Interventionsgruppen positive Verhaltensänderungen nachgewiesen werden, allerdings nur dann, wenn die Maßnahmen im Hinblick auf die Programmvorgaben und intendierten Programmziele besonders gut umgesetzt und ausgeführt wurden. Der Nachweis solcher differentieller Effekte setzt allerdings voraus, daß geeignete Ziel- und Analyseeinheiten bei der

Entwicklung der Programmkonzeption berücksichtigt und im Verlauf der Wirkungsmessung miterfaßt werden. Darüber hinaus ist dazu eine Überwachung und Evaluation der Programmdurchführung erforderlich, die im nächsten Abschnitt genauer behandelt wird.

Die bisherigen Ausführungen und Beispiele durften deutlich gemacht haben, welche grundlegenden Aufgaben bei der Evaluation der Programmkonzeption zu bewältigen sind und daß im Rahmen dieses Prozesses in Abhängigkeit von verschiedenen Faktoren zahlreiche Fragen und Probleme zu lösen sind. Welche Verfahren und Methoden dabei zur Anwendung kommen, hängt ebenfalls von verschiedenen Faktoren ab. Es wurde bereits darauf hingewiesen, daß zur Evaluation innovativer Programme andere Methoden und Techniken erforderlich sind als zur Evaluation etablierter Programme oder zur Modifikation und Feinabstimmung bereits laufender Projekte. Grundsätzlich lassen sich zur Problem- und Zielbestimmung alle systematischen sozialwissenschaftlichen Methoden und Techniken heranziehen. Dies können z. B. Experteninterviews, öffentliche Anhörungen und sekundär-statistische Auswertungen fremder Daten, aber auch eigene diagnostische Voruntersuchungen, Umfragen oder kleinere Modellversuche sein (vgl. Rossi et al., 1988). Lassen sich keine entsprechenden Vergleichszahlen oder brauchbaren Datenquellen heranziehen und eigene Voruntersuchungen aus verschiedenen Gründen nicht durchführen, so sollte in jedem Fall angestrebt werden, die unterschiedlichen Interessen der an einem Programm beteiligten Personen oder Personengruppen zu bündeln und zu einem Konsens im Hinblick auf die Problem- und Zielbestimmung des Programmvorhabens zu führen, damit eine sinnvolle Evaluation überhaupt möglich wird.

Evaluation der Programmdurchführung

Bereits im vorauslaufenden Abschnitt wurde anhand einiger Anwendungsbeispiele und Probleme auf die Bedeutung der Programmdurchführung und ihrer Überprüfung hingewiesen. Mit einer Evaluation der Programmdurchführung sollen mögliche Fehlentwicklungen bei der Ausführung eines Programmes aufgedeckt werden, so daß gegebenenfalls korrigierend eingegriffen werden kann. Dieser Evaluationsschritt stellt nicht nur eine wichtige Ergänzung, sondern auch eine notwendige Voraussetzung für eine valide Evaluation der Programmwirkungen dar. So bedeutet ein fehlender Nachweis von Programmwirkungen nicht unbedingt, daß ein Programm unwirksam oder ungeeignet ist. Es kann auch sein, daß die geplanten Maßnahmen nur unvollständig, falsch oder überhaupt nicht durchgeführt wurden, daß die intendierte Zielgruppe nicht erreicht wurde, oder daß eine nicht-standardisierte Durchführung zu einer Verwischung der Programmeffekte geführt hat.

Die Evaluation der Programmdurchführung hat im wesentlichen zwei Aufgaben (vgl. Abbildung 2). Durch **die Kontrolle der Programmausführung** wird überprüft, ob und inwieweit die Maßnahmen in Übereinstimmung mit der Konzeption und den Vorgaben des Programmes ausgeführt werden. Die **Prüfung der Programmreichweite** soll dagegen klären, ob und in welchem Umfang die intendierte Zielgruppe mit den

Programmaßnahmen erreicht wird. Die Ergebnisse dieser beiden Evaluationsaufgaben - die Kontrolle der Programmausführung und die Prüfung der Programmreichweite - ergeben zusammen die **Programmleistung** (siehe Abbildung 2; vgl. Rossi et al., 1988).

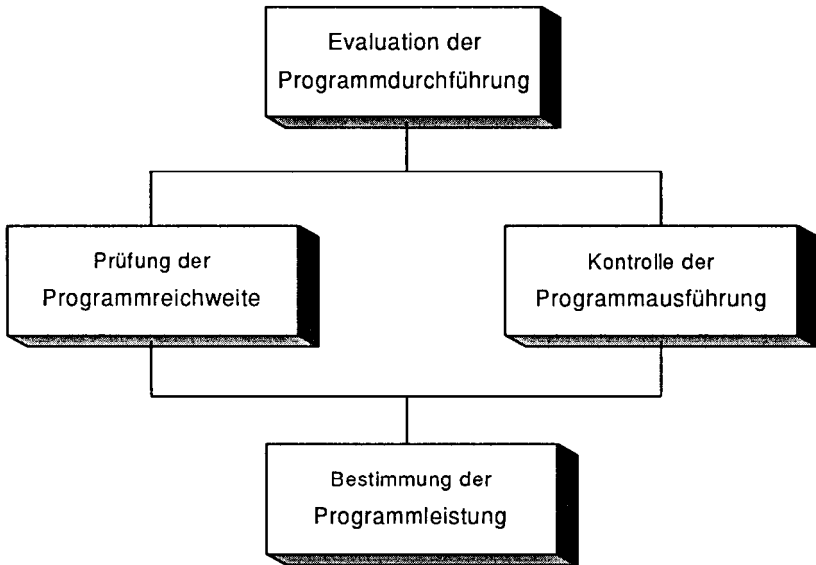


Abbildung 2. **Zentrale Aufgaben der Evaluation der Programmdurchführung zur Bestimmung der Programmleistung.**

Darunter fallen alle Maßnahmen, die entsprechend den Programmvorgaben durchgeführt werden und auch tatsächlich die intendierte Zielgruppe erreichen. Das zentrale Ziel der Evaluation der Programmdurchführung besteht also darin, die Programmleistung zu bestimmen und damit die Voraussetzung für eine valide Bewertung der Programmwirkungen zu schaffen. Zusätzlich kann die Evaluation der Programmdurchführung aber auch dazu genutzt werden, die zur Programmdurchführung notwendigen und anfallenden Ressourcen und Kosten zu dokumentieren, damit im Rahmen der Evaluation der Programmwirkungen auch das Verhältnis von Kosten und Wirkungen bzw. Nutzen eines Programmes analysiert werden kann (siehe unten).

Viele Präventionsprogramme scheitern beispielweise daran, daß die Programmmaßnahmen unvollständig, uneinheitlich, falsch oder gar nicht ausgeführt werden. Durch **Kontrolle der Programmausführung** könnten solche Fehlentwicklungen aufgedeckt und verhindert werden. Besteht beispielsweise die Gefahr, daß ein Programm an einer uneinheitlichen Ausführung der Maßnahmen scheitern könnte, kann unter Umständen noch rechtzeitig genug durch geeignete Maßnahmen gegengesteuert werden. Wird dagegen eine Variation in der Programmausführung festgestellt, ohne daß der Programmernfolg dadurch grundsätzlich in Frage gestellt ist, liefert die Kontrolle der Programmausführung wichtige Informationen, um im Rahmen der Wirkungsanalysen zu prüfen, unter welchen Ausführungsvarianten die größten

Programmeffekte aufgetreten sind (siehe z. B. Botvin et al., 1990). Zur Kontrolle der Programmausführung bieten sich verschiedene Methoden an. An erster Stelle sind hier Verfahren der Verhaltensbeobachtung zu nennen, die von der nicht-standardisierten Form mit narrativer Protokollierung bis hin zur systematischen Form mit einem hoch strukturierten Beobachtungs- und Bewertungsschema reichen können. Eine andere Möglichkeit ist die Erfassung und Analyse von Programmunterlagen, Aufzeichnungen und Daten der Programmmitarbeiter oder -Vermittler. Auch diese Daten können in ihrer Genauigkeit und Komplexität variieren. Sie können in Form narrativer Berichte vorliegen oder als standardisierte Fragebogen oder Interviews, durch die konkret erfaßt wurde, ob und wie bestimmte Programmbausteine durchgeführt wurden, welche Probleme dabei auftraten, wie diese aus der Sicht der Mitarbeiter von den Programmteilnehmern aufgenommen wurden und welche Ergebnisse dabei beobachtet werden konnten. Neben diesen Verfahren kann aber auch eine direkte Befragung der Programmteilnehmer in Frage kommen. Dabei sollten jedoch nicht nur allgemeine Zufriedenheitsmaße, sondern auch konkrete Fakten und Daten zum Programmablauf erhoben werden. Für welches Verfahren man sich schließlich entscheidet, hängt außer von der Fragestellung und dem Gegenstand der Evaluation auch von den vorhandenen Ressourcen und Möglichkeiten des Evaluationsprojektes ab. Der Einsatz von Beobachtungsverfahren ist im allgemeinen aufwendiger und teurer als die Befragung der Programmmitarbeiter oder Programmteilnehmer. Sind jedoch hinreichende Ressourcen vorhanden, sollte nach Möglichkeit mehr als nur eine Datenquelle und ein Verfahren genutzt werden. Dies gilt insbesondere dann, wenn ein Programm umfassend und formativ evaluiert werden soll.

Aussagen über die Wirksamkeit und den Nutzen eines Programmes können in ihrer Gültigkeit aber auch dadurch eingeschränkt oder völlig in Frage gestellt werden, daß die intendierte Zielgruppe mit den Maßnahmen nicht oder nur unzureichend erreicht wurde. Diese Art der Validitätsgefährdung tritt insbesondere dann auf, wenn die Programmteilnahme freiwillig ist und von der Akzeptanz oder Attraktivität des Programmes abhängt. Von daher sollte eine Evaluation der Programmdurchführung auch **die Prüfung der Programmreichweite** umfassen. Dabei ist festzustellen, ob und in welchem Umfang die intendierte Zielgruppe tatsächlich an dem Programm teilnimmt und wie hoch das Ausmaß an Verzerrungen durch Selbstselektion und Ausfälle bei den Programmteilnehmern ist. Dies gilt insbesondere dann, wenn im Rahmen der Evaluation der Programmwirkungen ein Vergleich zwischen Interventions- und Kontrollgruppe vorgesehen ist und beide Gruppen in unterschiedlichem Ausmaß von Verzerrungen betroffen sind. Voraussetzung für die Prüfung der Programmreichweite ist natürlich, daß eine Zielgruppe bei der Entwicklung der Programmkonzeption explizit festgelegt und durch empirisch erfaßbare Merkmale operational definiert wurde. Im Idealfall könnte dann durch entsprechende Auswahlkriterien über Programmteilnahme oder Ausschluß entschieden werden. In der Realität wird dieses Vorgehen allerdings nur selten praktiziert. Bei vielen Programmen ist es mitunter sehr schwierig oder sogar unmöglich, klare und hinreichend trennscharfe Kriterien zur Abgrenzung der intendierten Zielgruppe zu finden und festzulegen. Auch wird häufig auf die Anwendung diagnostischer Auswahlverfahren verzichtet, weil dies zu aufwendig und in Relation zu den sonstigen Programmkosten

zu teuer ist. Schließlich ist insbesondere bei Programmen mit generalpräventiver Zielsetzung eine Eingrenzung auf eine bestimmte Zielgruppe oft gar nicht wünschenswert, weil ein möglichst großer Teil der Gesamtpopulation oder eines bestimmten Segmentes der Population erfaßt und behandelt werden soll.

Das Problem der Programmreichweite läßt sich auch an den in jüngerer Zeit von den Krankenkassen angebotenen Kursen und Seminaren zur Prävention und Gesundheitsförderung verdeutlichen. Hier stellt sich angesichts der inzwischen sehr umfangreichen Angebotspalette die Frage nach der Inanspruchnahme durch die Versicherten. Wie hoch ist der Teilnahmegrad insgesamt bei den Versicherten? Wie hoch ist er bei Versicherten aus bestimmten Risikogruppen? Stimmt die Programmreichweite mit der Zielsetzung einzelner Kurse im besonderen und mit den Zielen des Gesamtangebotes im allgemeinen überein, oder ist die Programmreichweite bei den Versicherten insgesamt und in den artgezielten Risikogruppen eher gering, weil starke Verzerrungen durch die wiederholte Teilnahme einiger weniger, besonders aktiver Versicherter vorliegen? Als Verfahren zur Bestimmung der Programmreichweite bietet sich in diesem Beispiel unter anderem die Analyse der Akten und Daten der Versicherten an, soweit dies im Rahmen des Datenschutzes möglich ist. Darüber hinaus können aber auch Befragungen der Programmteilnehmer oder repräsentative Umfragen bei den Versicherten in Betracht kommen. Die Ergebnisse dieser Analysen und Umfragen sind für den Versicherungsträger in dreierlei Hinsicht von Nutzen. Erstens können sie als Rechenschaftsnachweis über den Versorgungsgrad der Versicherten durch Programme zur Prävention und Gesundheitsförderung verwendet werden (vgl. z. B. v. Troschke et al., 1990). Zweitens können diese Informationen - soweit notwendig - dazu herangezogen werden, um durch gezielte Werbekampagnen die Motivation zur Programmteilnahme in der Gruppe der eher passiven Versicherten zu erhöhen. Und drittens sind detaillierte Informationen über die Programmreichweite notwendig, um im Rahmen von Wirkungsanalysen zu validen Aussagen über die Wirksamkeit der Programmaßnahmen zu gelangen.

Evaluation der Programmwirkung und des Nutzens

Sichtet man die Literatur zur Evaluation gesundheitsbezogener Präventionsprogramme, so ist zunächst anzumerken, daß sich der überwiegende Teil der Studien auf die Evaluation der Programmwirkungen beschränkt, d. h. die Programmaßnahmen werden lediglich im Hinblick auf die erzielten bzw. intendierten Ergebnisse bewertet. Eine solche **Wirkungsanalyse** (impact assessment) erlaubt jedoch keine Schlußfolgerungen darüber, ob und inwieweit ein Programm oder bestimmte Maßnahmen auch unter dem Gesichtspunkt der Kosten positiv zu bewerten sind. Die Klärung dieser Frage ist Gegenstand von **Kosten-Effizienz-Analysen** (efficiency assessment), die allerdings im Rahmen gesundheitsbezogener Präventionsprogramme bislang eher selten durchgeführt worden sind. Kosten-Effizienz-Analysen werden in der Evaluationsliteratur auch als Kosten-Ergebnis-Analysen, Kosten-Ergebnis-Beurteilungen oder auch nur kurz als Kosten- bzw. Effizienz-Analysen bezeichnet (vgl. Bengel &

Koch, 1988; Bühringer & Hahlweg, 1988; Nöldner, 1990). Auf die verschiedenen Formen oder Ansätze der Kosten-Effizienz-Analyse wird später genauer eingegangen. **Als Nützlichkeitsanalyse** (Utility assessment) wird schließlich ein Bewertungsprozeß bezeichnet, der sowohl die Wirkungsanalyse als auch die Kosten-Effizienz-Analyse umfaßt (vgl. Abbildung 3).

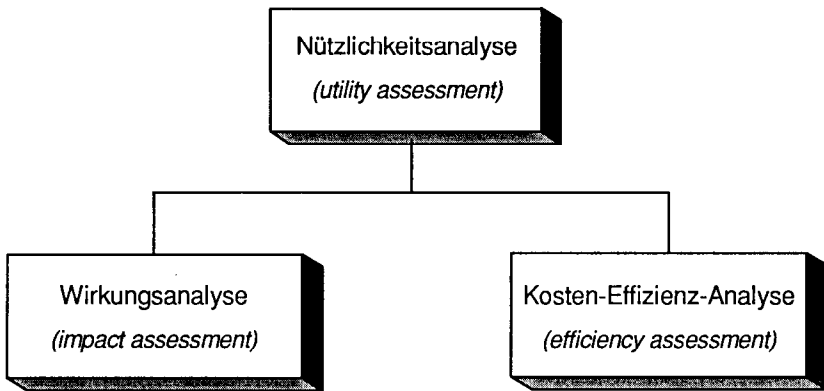


Abbildung 3. **Wirkungsanalyse und Kosten-Effizienz-Analyse als Unterformen der Nützlichkeitsanalyse.**

Wirkungsanalysen

Wirkungsanalysen prüfen, ob die durchgeführten Programmaßnahmen zu den beabsichtigten Ergebnissen geführt haben oder nicht. Die Beantwortung dieser Frage ist insoweit schwierig, als die soziale Realität sehr komplex ist und neben den Programmaßnahmen auch andere Faktoren und Störvariablen als Ursachen für die beobachteten Ergebnisse in Frage kommen können. Ferner ist zu berücksichtigen, daß die durchgeführten Maßnahmen auch zu unerwünschten Nebenwirkungen oder in Wechselwirkung mit anderen Faktoren zu sehr verschiedenartigen Ergebnissen führen können. Angesichts dieser Schwierigkeiten sind von Präventionsprogrammen in der Regel eher geringe Wirkungen und Veränderungen zu erwarten, die zudem möglicherweise nur sehr schwer nachzuweisen sind. Von daher ist es wichtig, daß Evaluationsprojekte so systematisch wie möglich geplant und durchgeführt werden, damit im Rahmen von Wirkungsanalysen der Einfluß anderer Faktoren und Fehlerquellen weitgehend kontrolliert und alternative Erklärungsmöglichkeiten für das Zustandekommen der beobachteten Ergebnisse ausgeschlossen werden können.

In der einschlägigen Literatur ist wiederholt auf verschiedene Fehlerquellen und Faktoren hingewiesen worden, die die Gültigkeit von Untersuchungsergebnissen gefährden. In diesem Zusammenhang sei an die von Cook und Campbell (1979) entwickelte Systematik der **Validitätsbedrohung** erinnert (vgl. Nöldner, 1990).

Bezogen auf die Wirksamkeit von Präventionsprogrammen kann prinzipiell zwischen programm-externen und programm-internen Störfaktoren unterschieden werden, die jeweils zusätzlich oder unabhängig von den Programmaßnahmen auf die Zielvariablen einwirken können (vgl. Rossi & Freeman, 1993; Rossi et al., 1988). **Als programm-externe Störfaktoren** gelten Umwelteinflüsse, wie z. B. endogene und exogene Veränderungen, historische Ereignisse und Reifungsprozesse sowie die unkontrollierte Auswahl der Programmteilnehmer durch Selbstselektion und Ausfälle. **Programm-interne Störfaktoren** sind dagegen Design-Effekte, die aus der Anlage und Durchführung der Evaluation resultieren. Dazu zählen u. a. stochastische Effekte, die Reliabilität und Validität der Meßinstrumente, die Auswahl der Ergebnisindikatoren und der Meßzeitpunkte, Fehlwerte und Stichprobeneffekte sowie mögliche Untersuchungseinflüsse durch „Hawthorne-“ bzw. „Placebo-Effekte“.

Da der Einfluß programm-externer und -interner Störfaktoren nie vollständig auszuschließen ist, muß davon ausgegangen werden, daß sich die insgesamt beobachteten Ergebnisse (**Bruttowirkung**) aus den tatsächlichen Wirkungen des Programmes, den Wirkungen programm-externer Störfaktoren und Design-Effekten zusammensetzen. Zentrales **Ziel der Wirkungsanalyse** ist die Bestimmung der **Nettowirkung** eines Programmes, d. h., denjenigen Anteil an den insgesamt beobachteten Ergebnissen zu ermitteln, der allein auf die durchgeführten Programmaßnahmen zurückzuführen ist. Dies setzt voraus, daß der Einfluß potentieller Störfaktoren bereits bei der Planung und Durchführung einer Evaluation in die Überlegungen einbezogen wird. Inwieweit dann im Rahmen von Wirkungsanalysen mögliche Störfaktoren berücksichtigt, kontrolliert und ausgeschaltet werden können, hängt von der **Qualität des Untersuchungsdesigns** ab. Nach Möglichkeit sollte ein Design mit mindestens einer Interventions- und einer Kontrollgruppe, einem Prä- und Posttest sowie einem zusätzlichen Follow-up gewählt werden, damit ein Vergleich von Programmteilnehmern und Nichtteilnehmern vor und nach der Intervention möglich ist und auch langfristige Programmwirkungen geprüft werden können. Darüber hinaus sollten mögliche Prätest-Unterschiede zwischen Interventions- und Kontrollgruppe durch Verwendung geeigneter Versuchspläne vermieden bzw. kontrolliert werden. Ob dazu **„echte“ Experimente** mit randomisierten Kontrollen oder **Quasi-Experimente** mit konstruierten und/oder statistischen Kontrollen herangezogen werden können, hängt von den jeweiligen Rahmenbedingungen ab. Bei innovativen Programmen in der Test- und Vorlaufphase sind beispielsweise eher Experimente mit randomisierten Kontrollen möglich, während bei bereits laufenden und etablierten Programmen allenfalls quasi-experimentelle Versuchspläne mit konstruierten oder statistischen Kontrollen in Frage kommen. Die Kontrolle potentieller Störfaktoren durch experimentelle oder quasi-experimentelle Versuchspläne ist allerdings nur bei Programmen mit partieller Erfassung möglich, d. h., wenn nicht alle Personen oder Einheiten der Zielpopulation gleichzeitig durch das Programm erfaßt bzw. behandelt werden. Ist dagegen eine vollständige Erfassung der Zielpopulation vorgesehen, wie dies bei groß angelegten Programmen wie z. B. der Deutschen Herz-Kreislauf-Präventionsstudie der Fall ist (vgl. z. B. v. Troschke et al., 1990), wird man sich in der Regel auf reflexive Kontrollen durch **Prätest-Posttest-Untersuchungen, Panel-Untersuchungen** oder **Zeitreihenanalysen** ohne Kontrollgruppe beschränken müssen. Unter bestimmten

Bedingungen ist bei Programmen mit vollständiger Erfassung allenfalls noch die Einführung von Wartekontrollgruppen möglich. Bei den zuletzt genannten Designalternativen ohne echte Kontrollgruppe kann der Einfluß bestimmter Störfaktoren (z. B. Reifungsprozesse, exogene Veränderungen und historische Ereignisse) auf die Programmresultate nicht mehr kontrolliert werden, so daß alternative Erklärungen für das Zustandekommen der Ergebnisse grundsätzlich nicht auszuschließen sind.

Die bislang erwähnten Versuchspläne eignen sich für Wirkungsanalysen, die üblicherweise zur Evaluation einzelner Programme durchgeführt werden. Angesichts der inzwischen - vor allem in den USA - zahlreich publizierten Studien zur Evaluation einzelner Gesundheitsprogramme besteht natürlich ein theorie- und praxisbezogenes Interesse, die Ergebnisse dieser Programme zusammenfassend zu evaluieren und miteinander zu vergleichen. Neben der umfassenden Frage nach der generellen Wirksamkeit gesundheitsbezogener Präventionsprogramme wird man vielleicht wissen wollen, welche spezifischen Maßnahmen sich auch über verschiedene Programme hinweg als wirksam erwiesen haben, inwieweit die Programmwirkungen möglicherweise von der Dauer und Intensität der Implementation abhängen oder auch davon, ob zur Vermittlung der Programmaßnahmen verschiedene Multiplikatoren eingesetzt wurden. Zur Klärung dieser und ähnlicher Fragen bieten sich verschiedene Verfahren der **Mets-Evaluation** an. Das grundlegende Ziel von Meta-Evaluationen ist die systematische Zusammenfassung und Integration der verschiedenen Forschungsergebnisse. Dies kann durch Diskussion einzelner ausgewählter Evaluationsstudien, qualitative Literaturübersichten (Reviews) oder durch statistisch-quantitative Auswertungen im Rahmen einer Meta-Analyse erreicht werden (vgl. Lösel & Nowack, 1987; Nöldner, 1990). Gegenüber qualitativen Literaturübersichten und der Diskussion ausgewählter Forschungsarbeiten hat **die Meta-Analyse** den Vorteil, daß subjektive Beurteilungen und Bewertungen weitgehend ausgeschaltet werden und neben der systematischen Zusammenfassung und Integration eine Quantifizierung des aktuellen Forschungsstandes möglich ist. Ein Überblick zur Anwendung und zu den Ergebnissen bereits vorliegender Meta-Analysen im Gesundheitsbereich findet sich bei Nöldner (1990; vgl. auch Jerusalem & Mittag, 1994, sowie Jerusalem, in diesem Band).

Kosten-Effizienz-Analysen

Vor dem Hintergrund knapper finanzieller Mittel und stetig steigender Kosten im Gesundheitswesen gewinnt die Forderung an Bedeutung, Programme nicht nur hinsichtlich ihrer Wirksamkeit, sondern auch unter Kosten-Nutzen-Gesichtspunkten zu evaluieren. Daß sich der überwiegende Teil der publizierten Evaluationsstudien aus dem Gesundheitsbereich ausschließlich auf Wirkungsanalysen beschränkt und Kosten-Effizienz-Analysen bislang eher selten vorzufinden sind, mag unter anderem daran liegen, daß sich viele Präventionsprogramme noch in der Test- und Erprobungsphase befinden und zunächst der Aspekt der Wirksamkeit im Vordergrund steht. Ebenfalls zu berücksichtigen ist, daß der Nutzen eines Programmes prinzipiell immer eine subjektive Größe darstellt, der von verschiedenen Faktoren wie den Rahmen-

bedingungen, dem Untersucher oder dem Zeitpunkt der Bewertung abhängt (vgl. Dlugosch & Wottawa, 1994). Gleichwohl kann die Bewertung des Nutzens - insbesondere bei etablierten Programmen - aus verschiedenen Gründen sinnvoll sein, wie z. B. für die Entscheidung über Fortsetzung und Weiterfinanzierung eines Programmes oder als Kriterium für die Auswahl zwischen verschiedenen Programmalternativen.

Eine Möglichkeit, den Nutzen eines Programmes zu bestimmen, besteht darin, die Programmwirkungen in Geld bzw. monetäre Einheiten zu überführen und diese anschließend den Programmkosten gegenüberzustellen. Bei solchen **Kosten-Nutzen-Analysen** (cost-benefit analyses) ist jedoch zu berücksichtigen, daß sich die Wirkungen gesundheitspsychologischer Präventionsmaßnahmen oft nicht unmittelbar in Geldeinheiten ausdrücken lassen oder der monetäre Nutzen erst langfristig und im Rahmen großflächig angelegter Programme deutlich wird. Auch dies durften Gründe dafür sein, warum Kosten-Nutzen-Analysen im Gesundheitswesen bis auf Ausnahmen in der betrieblichen Gesundheitsförderung (vgl. z.B. Murphy, Gasparotto & Opatz, 1987; Terborg, 1988) bisher kaum durchgeführt wurden. Unter gewissen Bedingungen besteht jedoch auch dann die Möglichkeit, die Effizienz von Präventionsprogrammen zu bestimmen, wenn die Programmwirkungen aus den oben genannten Gründen nicht monetär geschätzt werden können. Diese Form der Kosten-Effizienz-Analyse wird **als Kosten-Effektivitäts-Analyse** (cost-effectiveness analysis) oder auch als Effektivitäts- oder Kosten-Wirkungsanalyse bezeichnet (vgl. Bengel & Koch, 1988; Nöldner, 1990; Rossi et al., 1988).

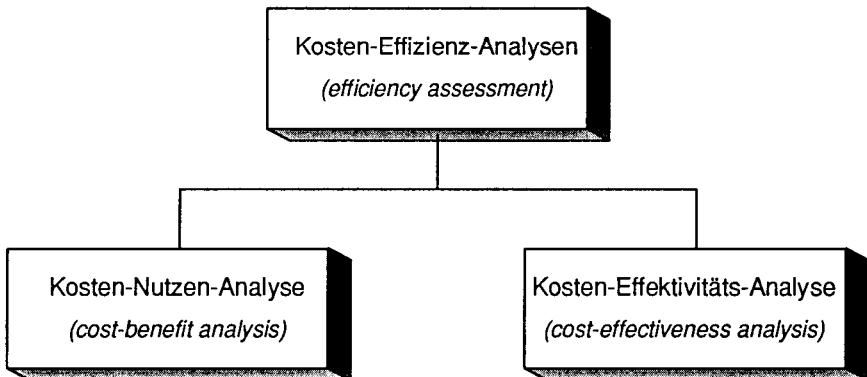


Abbildung 4. **Kosten-Nutzen- und Kosten-Effektivitäts-Analyse als Unterformen der Kosten-Effizienz-Analyse.**

Während bei einer Kosten-Nutzen-Analyse sowohl die Kosten als auch der Nutzen eines Programmes in monetäre Einheiten überführt und beide Größen in Beziehung zueinander gesetzt werden, werden bei einer Kosten-Effektivitäts-Analyse lediglich die Programmkosten monetär bestimmt und im Verhältnis zu den erzielten Wirkungen des Programmes betrachtet. Die entscheidende Frage ist hier also, in welchem

Verhältnis die Programmkosten zu den erzielten Wirkungen in der untersuchten Zielvariablen stehen. So könnte beispielsweise die Kosten-Effektivität eines Programmes zur Raucherentwöhnung so aussehen, daß jeweils 2000 DM für die langfristige Entwöhnung eines Rauchers aufgewendet werden müssen. Es ist offensichtlich, daß die Kosten-Effektivität eines einzelnen Programms wenig aussagekräftig ist. Nur wenn mehrere alternative Programme mit der gleichen Zielsetzung vorliegen, lassen sich die jeweiligen Kosten-Effektivitätswerte verschiedener Programmalternativen sinnvoll miteinander vergleichen. Ein alternatives Programm zur Raucherentwöhnung wäre zum Beispiel mit einer Kosten-Effektivität von 1700 DM effizienter als das oben genannte mit 2000 DM. Programme mit unterschiedlichen Zielsetzungen können mit Hilfe einer Kosten-Effektivitäts-Analyse nicht verglichen werden. Dies ist nur durch Kosten-Nutzen-Analysen möglich. Trotz ihrer Einschränkung ist die Kosten-Effektivitäts-Analyse bei vielen gesundheitsbezogenen Präventionsprogrammen eine sinnvolle und mitunter die einzige Alternative zur Bestimmung der Programmeffizienz, denn gerade im Gesundheitsbereich ist es oft sehr schwierig oder sogar unmöglich, den Geldwert von Programmwirkungen zu bestimmen (Welcher Geldwert soll beispielsweise für ein Menschenleben oder für eine höhere Lebenserwartung zugrundegelegt werden?). Daß Kosten-Effektivitäts-Analysen in der einschlägigen Literatur dennoch kaum anzutreffen sind, liegt vermutlich auch daran, daß die publizierten Evaluationsstudien oft nur sehr schwer oder gar nicht zu vergleichen sind, weil die Programmziele häufig nicht genau genug expliziert werden oder keine vergleichbaren Zielindikatoren vorliegen. Darüber hinaus ist auch zu bedenken, daß sich die Entwicklung von gesundheitspsychologischen Präventionsprogrammen zum Teil noch immer in der Testphase bzw. im Stadium der wissenschaftlichen Erprobung befindet und von daher Kosten-Nutzen-Überlegungen an dieser Stelle verfrüht oder eher von nachrangiger Bedeutung sind.

Literatur

- Bengel, J. & Koch, U. (1988). Evaluationsforschung im Gesundheitswesen. In U. Koch, G. Lucius-Hoene & R. Stegie (Hg.), *Handbuch der Rehabilitationspsychologie* (S. 321-347). Berlin: Springer.
- Botvin, G. J., Baker, E., Filazzola, A. D. & Botvin, E. M. (1990). A cognitive-behavioral approach to substance abuse prevention: One-year follow-up. *Addictive Behaviors*, *15*, 47-63.
- Bühringer, G. & Hahlweg, K. (1988). Kosten-Ergebnis-Analysen im Gesundheitswesen: Einführung in die Methodik und in die Analyse psychotherapeutischer Interventionen. In F. Lösel & H. Skowronek (Hg.), *Beiträge der Psychologie zu politischen Planungs- und Entscheidungsprozessen* (S. 98-104). Weinheim: Deutscher Studien Verlag.
- Cook, T. D. & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Chicago, IL: Rand McNally.
- Cook, T. D. & Matt, G. E. (1990). Theorien der Programmevaluation - Ein kurzer Abriß. In U. Koch & W. W. Wittmann (Hg.), *Evaluationsforschung: Bewertungsgrundlage von Sozial- und Gesundheitsprogrammen* (S. 15-38). Berlin: Springer.

- Dlugosch, G. E. & Wottawa, W. (1994). Evaluation in der Gesundheitspsychologie. In P. Schwenkmezger & L. R. Schmidt (Hg.), *Lehrbuch der Gesundheitspsychologie* (S. 149-168). Stuttgart: Enke.
- Fuchs, R. & Schwarzer, R. (1994). Gesundheitserziehung und Gesundheitsförderung. In K. Schneewind (Hg.), *Psychologie der Erziehung und Sozialisation* (S. 403-432). Göttingen: Hogrefe.
- Jerusalem, M. & Mittag, W. (1994). Gesundheitserziehung in Schule und Unterricht. *Zeitschrift für Pädagogik*, **40**, 851-869.
- Lösel, F. & Nowack, W. (1987). Evaluationsforschung. In J. Schultz-Gambard (Hg.), *Angewandte Sozialpsychologie* (S. 57-87). München: PVU.
- Murphy, R. J., Gasparotto, G. & Opatz, J. P. (1987). Current issues in the evaluation of worksite health Promotion programs. In J. P. Opatz (Ed.), *Health promotion evaluation: Measuring the organizational impact* (pp. 1-14). Steven Point: National Wellness Association.
- Nöldner, W. (1990). Evaluation in der Gesundheitspsychologie. In R. Schwarzer (Hg.), *Gesundheitspsychologie. Ein Lehrbuch* (S. 475-487). Göttingen: Hogrefe.
- Rossi, P. H. & Freeman, H. E. (1993). *Evaluation. A systematic approach* (5th ed.). Newbury Park, CA: Sage.
- Rossi, P. H., Freeman, H. E. & Hofmann, G. (1988). *Programm-Evaluation: Einführung in die Methoden angewandter Sozialforschung*. Stuttgart: Enke.
- Schaps, E., DiBartolo, R., Moskowitz, J., Palley, C. S. & Churgin, S. (1981). A review of 127 drug abuse prevention program evaluations. *Journal of Drug Issues*, **12**, 17-43.
- Schwarzer, R. (1996). *Psychologie des Gesundheitsverhaltens* (2. Aufl.). Göttingen: Hogrefe.
- Schwenkmezger, P. & Schmidt, L. R. (Hg.). (1994). *Lehrbuch der Gesundheitspsychologie*. Stuttgart: Enke.
- Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, R. M. Gagne & M. Scriven (Eds.), *Perspectives of curriculum evaluation* (pp. 39-83). Chicago: Rand McNally.
- Suchman, E. A. (1967). *Evaluative research: Principles and practice in public service and social action programs*. New York: Russell Sage Foundation.**
- Telch, M. J., Miller, L. M., Killen, J. D., & Cooke, S. (1990). Social influences approach to smoking prevention: The effects of videotape delivery with and without same-age peer leader participation. *Addictive Behaviors*, **15**, 21-28.
- Terborg, J. R. (1988). The organization as a context for health promotion. In S. Spacapan & S. Oskamp (Hg.), *The socialpsychology of health* (S. 129-174). Beverly Hills, CA: Sage.
- Troschke, J. von, Dirscherl, C., Gerber, U., Kluge, M., Kotek, H., Köhler, S., Kupke, R., Stünzer, W. von, Wagner, A., Wiche, E. & Worth, M. (1990). Die gemeindenahe Begleitforschung der Deutschen Herz-Kreislauf-Präventionsstudie (DHP). In U. Koch & W. W. Wittmann (Hg.), *Evaluationsforschung - Bewertungsgrundlage von Sozial- und Gesundheitsprogrammen* (S. 227-247). Berlin: Springer.
- Wittmann, W. W. (1985). *Evaluationsforschung. Aufgaben, Probleme und Anwendungen*. Berlin: Springer.
- Wottawa, H. & Thierau, H. (1990). *Lehrbuch Evaluation*. Bern: Huber.