

**Thomas Elkeles**

## **Evaluation von Gesundheitsförderung und Evidenzbasierung?**

### **1. Einleitung**

Das Fragezeichen hinter dem Titel meines Beitrags soll meine Skepsis symbolisieren, inwieweit es angebracht und zielführend ist, die Evaluation von Gesundheitsförderung mit der Forderung nach Evidenzbasierung nach den methodischen Idealen der klinischen Medizin und Epidemiologie in Verbindung zu bringen. Damit sollen nicht etwa verdienstvolle Initiativen konterkariert werden, die bestverfügbare wissenschaftliche Evidenz zur Wirksamkeit und Wirtschaftlichkeit von Gesundheitsförderung zu ermitteln und anzuwenden. Vielmehr geht es um die dabei anzuwendenden bzw. zu entwickelnden Maßstäbe vor dem Hintergrund des Standes in der Evaluation von Gesundheitsförderung, welcher hierzu illustriert und verdeutlicht werden soll, zumal dieser auch in den politischen Forderungen nach einer Evidenzbasierung der Gesundheitsförderung durch Maßnahmen der Krankenkassen im Rahmen des § 20 SGB V unterzugehen droht (vgl. Kirschner/Elkeles/Kirschner 2004a-c).

Hierzu soll entlang von fünf Thesen, beginnend mit einer kurzen begrifflichen Einordnung, die Anwendbarkeit des Evidenzbegriffs auf einige ausgewählte Anwendungsgebiete im Bereich Gesundheit und Pflege dargestellt werden.

### **2. Begriff, Definition und Anwendungsgebiete gesundheitsbezogener Evidenz**

Da sich die sog. Evidenzbasierung auf die Resultate durchgeführter Evaluationen bezieht, sei zunächst kurz die Evaluation selbst definiert. Unter Evaluation verstehen wir die wissenschaftliche Bewertung von Sachverhalten – in unserem Kontext von gesundheitlich-sozialen Programmen – mittels jeweils vom Sachverhalt her angemessener Methoden.

Entsprechend bezieht man sich in der sozialwissenschaftlichen Evaluationsforschung häufig auf Rossi et al. (1988), die Evaluation als „systematische Anwendung sozialwissenschaftlicher Forschungsmethoden zur Beurteilung der Konzeption, Ausgestaltung, Umsetzung und des Nutzens sozialer Interventionsprogramme“ definieren (ebda., S. 3). Programmevaluation unterscheidet sich wie alle anderen Sparten bzw. Perspektiven der Evaluationsforschung dadurch von Alltagsbeurteilungen, dass systematischen, datenbasierten und rational rekonstruierbaren Beschreibungs- und Bewertungsregeln gefolgt wird (vgl. Struening 1975, Hellstern/Wollmann 1984, Wottawa/Thierau 1998, Stockmann 1996, 2004, Elkeles/Georg 2002, Elkeles/Kirschner 2003, Kromrey 2004).

Dass es eine Vielfalt unterschiedlicher Definitionen gibt, ist eine Konsequenz der sehr unterschiedlichen Evaluationsgegenstände. „Da mit Evaluationen sehr unterschiedliche Ziele, Erkenntnis- und Verwertungsinteressen verfolgt werden, werden auch sehr unterschiedliche Ansätze, Modelle und Evaluationsdesigns verwendet“ (Stockmann 2004, S. 15).

Wenn nun von Evidenzbasierung gesprochen wird, sollte zunächst grundsätzlich verdeutlicht werden, was unter Evidenz verstanden bzw. von welchen Prämissen hierbei auch erkenntnistheoretisch ausgegangen wird. Hierzu meine *These 1*:

„Es gibt unterschiedliche Begriffe von Evidenz. Welche Begriffe, Formen und sich daraus ergebende Methoden des Belegs oder Nachweises angemessen sind, hängt vom jeweiligen wissenschaftlichen Gegenstand ab.“

Nach dem Brockhaus ist Evidenz [lateinisch »Augenfälligkeit«] die „unmittelbare, mit besonderem Wahrheitsanspruch auftretende, intuitive Einsicht“. So liege bei Platon Evidenz

vor, wenn es sich um die Ideenschau handelt; bei R. Descartes, wenn eine Vorstellung klar und deutlich erfasst wird; in E. Husserls Phänomenologie im Fall der Schau eines objektiven Wesenssachverhalts“ (Brockhaus 2003). Unabhängig von einem philosophischen oder erkenntnistheoretischen Ansatz kann also zunächst festgestellt werden, dass der Evidenzbegriff als solcher ohne jegliche mathematische Quantifizierung oder gar klinisch-epidemiologische Versuchsanordnung auskommt. Das schließt nicht aus, dass in Teilgebieten die Sachverhalte nicht anders und nur auf eine solche Weise erfaßt und bewertet werden können, jedoch kann dies nur der Maßstab für die Sachverhalte sein, in denen dies so ist, nicht hingegen für die gesamte Welt im Allgemeinen oder die gesamte Gesundheitsförderung im Besonderen.

Genau hierauf zielt meine *These 2*:

„Die Evidenzhierarchie der Biomedizin und vergleichbarer Gebiete ist nicht auf die Gesundheitsförderung übertragbar.“

Auch die Evidenzbewegung in Technik (Health Technology Assessment) und Medizin hat Eingang in den Brockhaus gefunden. Evidenzbasierte Medizin wird dort dahingehend erläutert, sie sei eine „medizinische Vorgehensweise, bei der diagnostische und bzw. oder therapeutische Entscheidungen auf der Basis systematisch zusammengetragener und bewerteter wissenschaftlicher Erkenntnisse getroffen werden, um damit für den einzelnen Patienten das beste Ergebnis möglichst ressourcenschonend zu erzielen. Die erforderlichen ärztlichen Entscheidungen dürfen jedoch nicht allein aufgrund solcher Erkenntnisse und damit schematisch erfolgen, vielmehr sind auch die vielen individuellen Besonderheiten jedes einzelnen Patienten zu berücksichtigen“ (Brockhaus 2003).

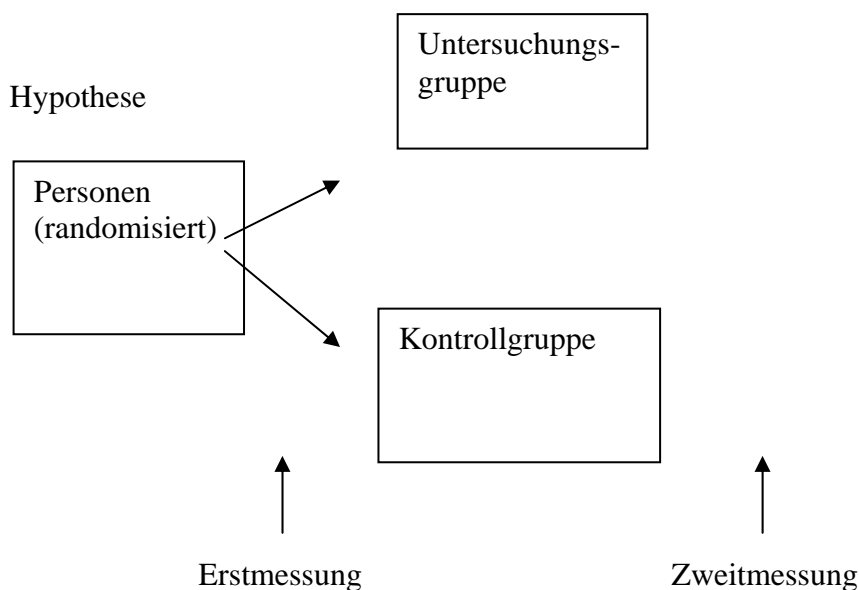
Bei den klinischen Maßnahmen der Biomedizin und anderer therapeutischer Interventionen handelt es sich idealtypisch um situations- und kontextinvariante Applikationen von Produkten oder Verfahren. Hier ist als Wirksamkeitsnachweis ein ‚Goldstandard Randomized Controlled Trial‘ möglich und auch angemessen. Noack (2002) weist allerdings darauf hin, dass auch in klinischen Zusammenhängen die systematische Ausschaltung (der Erfassung) von Interaktionseffekten, wie sie der Randomisierung zugrunde liegt, problematisch und jedenfalls nicht stets sinnvoll sei (ebda., S. 18). So seien z.B. Überlebenszeiten bei schweren chronischen Krankheiten von der Qualität sozialer Bindungen bzw. Interaktionen abhängig.

Bei Gesundheitsförderung handelt es sich idealtypisch um soziale Programme, welche stets kontextabhängig sind. Nur für Einzelbestandteile solcher Programme, wie insbesondere verhaltensbezogene Interventionen, sind überhaupt quantitative Nachweismethoden möglich, andere Programme sind nur mit qualitativen Methoden bewertbar. Eine quantitative Evidenzhierarchie mit einem den Situationskontext bewußt ausblendenden („verblindenden“) Evidenzideal sind dem Gegenstand daher nicht angemessen. „Gesundheit entwickelt sich bzw. wird (re)produziert in der Person-Umwelt-Interaktion (das Geheimnis der Salutogenese und wahrscheinlich auch der Heilung liegt in der Interaktion)“ (Noack 2002, S. 18).

Allerdings muss eingeräumt werden, dass auch die sozialwissenschaftliche Evaluationsforschung in ihrer Geschichte nicht stets frei war von einem „methodologischen Rigorismus“ (Stockmann 2004, S. 20), der experimentellen Designs grundsätzlich den Vorzug gab. Die Tendenz insbesondere der US-amerikanischen Evaluationsforschung zu experimentellen oder quasi-experimentellen Designs steht in der Tradition des logischen Positivismus (vgl. Hügli/Lübcke 1991, S. 468), welcher die Auffassung ablehnt, die Philosophie könne selbständig zur Erkenntnis von Welt gelangen und statt dessen das Ideal der Erkenntnis in den empirischen Wissenschaften sieht, die sich der Wahrnehmung und kontrollierter Experimente bedient.

So kam es zum klassischen experimentellen Design (Abb. 1), bei dem einer Gruppe der Fälle (Untersuchungsgruppe oder auch Experimentalgruppe) eine vergleichbare Gruppe von Personen gegenübergestellt und im Vorher-Nachher-Vergleich untersucht wird, ob die angenommene Wirkung sich mit der hierfür für geeignet gehaltenen Intervention (Hypothese) bei den Personen der Untersuchungsgruppe einstellt bzw. die Hypothese zu verwerfen ist. Damit die Wirkung wirklich ausschließlich auf die postulierte und zu beweisende Kausalbeziehung zurückgeführt werden kann, ist Voraussetzung, daß sich die Personen der Kontrollgruppe von denen der Untersuchungs- oder Experimentalgruppe ausschließlich in dem Merkmal unterscheiden, dass bei ihnen die Intervention nicht durchgeführt wird. Die Sicherheit dafür läßt sich erhöhen, wenn die Zuteilung der Personen zur Untersuchungs- und zur Kontrollgruppe nach dem Zufallsprinzip geschieht (Randomisierung).

Abb. 1: Fall-Kontrollgruppen-Design



Quelle: Elkeles/Kirschner (2003)

Wo immer ein solches Design nicht nur möglich, sondern auch sinnvoll ist, ist das Fall-Kontrollgruppen-Untersuchungsdesign sicherlich das Mittel der ersten Wahl. Das wird vor allem dort zutreffen, wo es sich um gegenüber sozialen Bedingungen invariante Kausal- und Wirkungszusammenhänge handelt, wie das z.B. bei stofflichen Vorgängen der Fall ist. Daher ist die Arzneimittelforschung auch das häufigste Anwendungsfeld des Fall-Kontrollgruppen-Designs. In der Rolle des (potentiellen) Arzneimittelkonsumenten wird man hier den ‚Goldstandard Randomized Controlled Trial‘ keinesfalls missen mögen.

Zur Fall-Kontrollgruppen-„Philosophie“ wird ein solches Design hingegen, wenn seine Anwendung unhinterfragt auch in Interventionsfeldern geschieht bzw. postuliert wird, deren Wirkungsmechanismen komplexer als einfache lineare Ursache-Wirkungsbeziehungen sind, wie dies regelmäßig bei (gesundheitlich-) sozialen Interventionen der Fall ist.

Hier ist nicht nur das Untersuchungsfeld in ständiger Bewegung und Entwicklung und damit der Untersuchungsgegenstand nicht konstant. Auch sind häufig selbst bei klar definierten Maßnahmenprogrammen die abhängigen Variablen, d.h. die Dimensionen sozialer Wirkungen, bei denen Veränderungen beobachtet werden sollen, nicht von vornherein und eindeutig bekannt. Ferner wird es in Interventionsprojekten, die durch freiwillige Teilnahme (Selbstselektion der Teilnehmerinnen und Teilnehmer) zustande kommen, überhaupt

schwierig werden, Kontrollgruppen-Teilnehmerinnen und –Teilnehmer zu finden, jedenfalls nicht ohne seitens der Untersucher die sozialen Bedingungen des Untersuchungsfeldes massiv zu beeinflussen. Damit aber wären die Rahmenbedingungen des Laborexperiments verletzt.

Welches das angemessene Design zur Untersuchung einer Intervention ist, sollte nicht von einem fixen Wissenschaftsparadigma (also weder der Fall-Kontrollgruppen-Philosophie noch einem anderen erkenntnistheoretischen und/oder methodischen Dogma) abhängig gemacht werden, sondern vom Untersuchungsgegenstand selbst. Ferner hat jede Meßmethode ihre eigenen Stärken und Schwächen und erfaßt Sachverhalte aus je unterschiedlicher Perspektive mit je unterschiedlicher Breite und Tiefe.

Wenn ein Untersuchungsgegenstand komplex ist, kann es sinnvoll bzw. die „Methode der Wahl“ sein, eine Kombination von unterschiedlichen Methoden anzuwenden (Methoden-Mix, z.B. quantitative und qualitative Methoden). In einer eigenen Evaluationsstudie zur Wirksamkeit eines staatlichen Förderprogramms, das betriebliche Gesundheitsförderung einschloss (Elkeles/Georg 2002)<sup>1</sup>, schien uns eine durch Triangulation erreichbare Perspektivenerweiterung dadurch beschreibbar, dass z.B. die Betrachtungsebene der Sachinformationen (Organisationsstrukturen, Tätigkeiten, Dokumentation von Konflikten, Ergebnisse etc.) anhand verschrifteter Aussagen (Akten- und Dokumentenanalysen) durch die Betrachtungsebene subjektiver Sinnstrukturen (Rekonstruktion von teilweise latent enthaltenen Präventionskonzepten, Lernerfahrungen etc.) anhand der Durchführung und Auswertung qualitativer Tiefeninterviews zu ergänzen war. Mit quantitativen Methoden oder gar Kontrollgruppen wäre diese Evaluation nicht durchführbar gewesen. Nachweise der Wirksamkeit sozialer Programme sind ohnehin wissenschaftlich nicht unumstritten, zu vielen Programmen (z.B. Programm Humanisierung der Arbeitswelt, vgl. Volkholz 1991, 1993), wurden keine derartigen Nachweise erbracht, jedoch wurde hier nie auch nur der Versuch unternommen, situationsinvariante Nachweismethoden oder gar entsprechende Evidenzhierarchien anzuwenden.

Ein solches Spektrum der Evaluationsforschung, wie es jedenfalls für einen Großteil der Gesundheitsförderung zutrifft, entzieht sich vom Ansatz her den Beweisregeln und deren Evidenzhierarchie (je mehr Laborsituation pro Studientyp, umso höhere Beweiskraft) für die Ableitung von Evidenz zu medizinischen Interventionen.

Der „methodologische Rigorismus“ – wie er uns nun in der Gesundheitsförderung durch Rekurs auf die Evidenzbewegung in der Medizin neu entgegentritt - ist in der Evaluationsforschung zwar nicht grundsätzlich überwunden, aber zumindest durch neue Ansätze relativiert worden, welche sich insgesamt „als Reaktion auf die Erfolglosigkeit der Versuche, das Experimentaldesign zu realisieren“ ergaben (Stockmann 2004, S. 20). Wie noch zu zeigen sein wird, ist von der Sach- und Materiallage her eine standardisierte Evaluation sehr häufig, wenn nicht überwiegend, nicht möglich. Kromrey (2001) gibt als eine Art Faustregel an, dass die „Anwendbarkeit (...) nicht als der Regelfall, sondern eher als der Ausnahmefall gelten kann. Es muss also zu ‚Ersatzlösungen‘ gegriffen werden, die praktikabel erscheinen und dennoch hinreichend gültige Ergebnisse liefern“ (ebda., S. 122).

---

<sup>1</sup> In dieser Studie war z.B. die Bildung von Kontrollgruppen nicht möglich, da sie ex post nicht mehr zu bilden waren, was auf viele ex post-Evaluationen zutreffen wird. Komplexe und diffuse Programme, deren Bestandteil eine Selbstausbreitung in verschiedensten Settings und Regionen ist, würden selbst bei ex ante-Planung eine permanente Kontrollgruppenbildung erfordern, die kaum praktikabel sein dürfte.

### 3. Evidenzbasierte Pflege

Der folgende Exkurs zum Beispiel des Praxisfelds Pflege, in dem die Evidenzbewegung in Technik und Medizin sich ebenfalls ausgebreitet hat, soll verdeutlichen, in welchem Ausmaß hier Adaptionen erforderlich sind, um die Forderung nach Evidenzbasierung überhaupt diskutierbar zu machen, dabei jedoch auf von der Sach- und Materiallage her gegebene Grenzen stößt.

Hierzu meine *These 3*: ‚Aufgrund der kulturellen Hegemonie der Medizin kommt es in anderen Gebieten des Gesundheitswesens (Pflege, Gesundheitsförderung) zu einer Übernahme des biomedizinischen Evidenzbegriffs, wobei dieser eine beispiellose Inflationierung bei teilweise gleichzeitiger Auflösung seines Sinngehalts erfährt.‘

Mit dem Anschluss an die Evidence-based Medicine-Bewegung ist in der Pflegewissenschaft der Begriff Evidence-based Nursing (EBN) geprägt worden. Fragestellungen sind z.B. nach einer Darstellung des German Center for Evidence-based Nursing: „Kann ich bei bettlägerigen Patienten durch einen 2-stündlichen Lagewechsel im Gegensatz zum jetzigen 4-stündlichen Lagewechsel die Entstehung von Dekubiti verhindern?“ oder „Wird bei beatmeten erwachsenen Patienten durch eine Musiktherapie Angst reduziert und Entspannung gefördert?“

„Ein paar Definitionen“ (zitiert nach German Center for Evidence-based Nursing):

„Kurz und knapp: Evidence-based Nursing ist die Integration der derzeit besten wissenschaftlichen Belege in die Pflege.

Eine erweiterte Definition, basierend auf den vier Komponenten einer pflegerischen Entscheidung (wobei die Komponenten bei jeder Entscheidung in unterschiedlich starkem Ausmaß Einfluß nehmen), könnte lauten: Evidence-based Nursing ist die Integration der derzeit besten wissenschaftlichen Belege in die tägliche Pflegepraxis unter Einbezug theoretischen Wissens und der Erfahrungen der Pflegenden, der Vorstellungen des Patienten und der vorhandenen Ressourcen.

Auch die folgende Variante trifft den Kern von EBN ganz gut: Evidence-based Nursing ist die Nutzung der derzeit besten wissenschaftlich belegten Erfahrungen Dritter im Arbeitsbündnis zwischen einzigartigen Pflegebedürftigen und professionell Pflegenden.“ (vgl. Behrens/Langer 2004, S. 21ff.)

Es geht also wie in der Medizin und anderen therapeutischen Gebieten um im engeren oder weiteren Sinne therapeutische Interventionen zu standardisierten, situationsinvarianten Behandlungsbedarfen. Konsens besteht inzwischen, dass in den Evidenzbegriff von Pflege neben wissenschaftlichen Belegen zu therapeutisch-pflegerischen Interventionen das klinische Erfahrungswissen sowie die Patientenerfahrungen bzw. -vorstellungen eingehen sollen.

Mit dieser in der Pflege offenbar notwendigen Öffnung des Evidenzbegriffs ist gleichzeitig impliziert, dass er ins Grenzenlose ausgeweitet werden kann. Wenn aber in der Extremvariante jedwede pflegerisch-therapeutische Entscheidung unter Bezug auf so verschiedene Wurzeln als „evident“ begründet werden kann, ist m.E. der Geltungsbereich der Evidence-based Medicine-Bewegung überschritten, wofür es ja in der Pflege treffende Gründe geben mag; allerdings müßte dann m.E. auch die Konsequenz gezogen werden, sich auch von ihrem Begriffsgebäude zu entfernen. Dies geschieht jedoch in der Pflege aus Gründen unvollständiger Professionalisierung nicht bzw. zu wenig.

Anhand von einer begrifflichen Auseinandersetzung mit Evidence-based Nursing seitens Gross (2004) möchte ich die Widersprüche und die dort gesuchten Lösungsmöglichkeiten im Folgenden näher darlegen. Unterschiedliche Verwendungsweisen des Begriffs „Evidence“ könnten nicht auf unterschiedliche Bedeutungen im Deutschen und im Englischen

zurückgeführt werden. Bedeutsam sei aber, dass der Begriff Evidenz von Ärzten und Pflegepersonen unterschiedlich verwendet wurde und wird.

„Bei Ärzten wird der Begriff als Effektivität gewertet, und dies hat zu einer Hegemonie von experimentellen Untersuchungsdesigns geführt. Evidenz von anderen Quellen wie Expertenwissen, klinischer Sachkenntnis, Patientenperspektiven etc. wird eher vernachlässigt“ (Gross 2004, S. 197f.). Ergebnisse würden eher akzeptiert, wenn sie quantitativer Art seien. „Erschwerend kam hinzu, dass die von Guyatt et al. (1995) aufgestellten Evidenzhierarchien an erster Stelle der Hierarchie immer die Meta-Analyse ansetzten. Expertenmeinungen standen dagegen immer an letzter Stelle. Ergebnisse aus qualitativen Untersuchungen kamen je nach Evidenzhierarchie entweder gar nicht vor oder waren in der Hierarchie sehr weit unten angesiedelt. Dies hat durchaus seinen logischen Grund, da zur Zeit der Entwicklung der Evidenzhierarchien die Medizin in erster Linie der positivistischen Forschungstradition verhaftet war und nach wie vor verhaftet ist (z.B. Hewitt-Taylor, 2003)“ (Gross 2004, S. 199).

Die Autorin verweist darauf, dass es schon ‚traditionell‘ in der Pflegewissenschaft eine Tendenz zur qualitativen Forschung gebe, da man sich hier eher mit Fragen nach dem Warum und Wieso beschäftige; dies sei aber auch der Grund dafür, dass „der von Ärzten verwendete Evidenzbegriff für die Pflege nie wirklich ausschließliche Gültigkeit hatte, allein schon aus dem Mangel an Metaanalysen und randomisierten klinischen Studien“ (Gross 2004, S. 199).

Daher sei der Evidenzbegriff in der Pflege schon frühzeitig in Frage gestellt worden und die Evidenzhierarchie entweder durch eine eigene, qualitative ergänzt oder aber gänzlich aufgelöst worden.

Gross (2004) selbst plädiert nicht für die Auflösung der Evidenzhierarchie, denn Metaanalysen hätten dort ihren Stellenwert, wo quantitative Forschungsansätze erlaubt seien. Auch bezieht sie sich auf anderweitige klinische Modelle, deren höchste Stufe einem klinischen Experimentaldesign entspricht, was durch andere Aussagen wieder relativiert wird und so insgesamt ambivalent erscheint. Letzteres erscheint kongruent mit ihrer Prognose, „dass beide Berufsgruppen in Sachen Evidenz sich wieder annähern und einen einheitlichen Evidenzbegriff definieren werden.“ (ebda., S. 199)

Eine wissenschaftliche Darstellung zum Thema Evidence-based Nursing (Behrens/Langer 2004), die den allgemeinen methodischen Fragen qualitativer und quantitativer Forschungsdesigns und der kritischen Beurteilung von Studien breiten Raum widmet, ohne dabei etwa bestimmte Studientypen grundsätzlich zu bevorzugen<sup>2</sup>, in typischen Beispielen allerdings sich auf medizinverwandte Fragestellungen beschränkt (z.B. Schlucktraining bei Apoplexie), läßt die von Gross (2004) herausgestellte Frage weitgehend offen, wie es um die Datenlage pflegewissenschaftlicher Interventionsforschung und ihre Eignung für

---

<sup>2</sup> „Ob eine Studie angemessen und geeignet ist, ergibt sich für Evidence-based Nursing aus dem Handlungskontext, in dem ihre Ergebnisse gefragt sind und genutzt werden sollen. Insofern gibt es keinen für alle Handlungsprobleme gleichmäßig geeigneten Studientyp.“ (Behrens/Langer 2004, S. 107) Im übrigen wird hier dargelegt, dass jedes Design jeweils seine Vor- und Nachteile hat (ebda., S. 110) und alle Forschungsdesigns ein gemeinsames Ziel haben: „Die Verfahren sollen die Chance erhöhen, dass unser Vorverständnis widerlegt werden kann, oder anders gesagt: dass die – unseren Erwartungen widersprechende – Entwicklung dieselbe Chance hat, sich als zutreffend zu erweisen, wie das von uns erwartete Ergebnis. Dies gilt für Verfahren, die häufig als ‚qualitativ‘ bezeichnet werden, ebenso wie für die so genannten ‚quantitativen‘ Verfahren.“ (ebda., S. 112)

Sekundärpublikationen (Metaanalysen und systematische Übersichtsarbeiten) nach dem Evidenzideal der Medizin und ihrer Evidenzhierarchie steht.<sup>3</sup>

Das Beispiel Evidence-based Nursing zeigt zusammenfassend:

- Obwohl sich pflegerische und medizinische Interventionen aufgrund ihrer therapeutischen Interventionsart wesentlich näher stehen als therapeutische und gesundheitsförderliche Interventionen, gibt es zumindest starke Zweifel, ob alle pflegerischen Interventionen für quantitative Studiendesigns und daraus abgeleitete Evidenznachweise geeignet sind.
- Die Hierarchieniveaus von Evidenz als Gütekriterien in der Evidenzbasierten Medizin setzen eine Materiallage von Studien voraus, wie sie in der Pflege (ebenso wie in der Gesundheitsförderung, s.u.) nicht existiert.

#### 4. Gesundheitsförderung

Nach der Ottawa-Charta zielt Gesundheitsförderung auf einen Prozeß, allen Menschen ein höheres Maß an Selbstbestimmung über ihre Gesundheit zu ermöglichen und sie damit zur Stärkung ihrer Gesundheit zu befähigen. Sowohl die hierfür in der Charta genannten vier Handlungsstrategien (Voraussetzungen für die Gesundheit sichern, Interessen vertreten [advocacy], Befähigen und ermöglichen [enabling], Vermitteln und vernetzen [mediating] wie auch die fünf Aktionsfelder (eine gesundheitsfördernde Gesamtpolitik entwickeln, gesundheitsförderliche Lebenswelten schaffen, die Gesundheitsdienste neu orientieren, gesundheitsbezogene Gemeinschaftsaktionen unterstützen, persönliche Kompetenzen entwickeln) geben gleichzeitig Maßstäbe, die gut für die wissenschaftliche Beurteilung der Gesundheitsförderlichkeit durchgeführter Interventionen geeignet sind (vgl. Lenhardt 1999, Elkeles 2002).

Bei der im § 20 des SGB V verankerten Gesundheitsförderung durch Krankenkassen wird inzwischen dem "Setting-Ansatz" eine bedeutsame Rolle neben der individuenzentrierten Verhaltensprävention eingeräumt. Der Setting-Ansatz ist eine Antwort auf die beschränkten Erfolge traditioneller Gesundheitserziehung, die sich mit Informationen und Appellen an Einzelpersonen wenden. Als Setting wird ein soziales System verstanden, das eine Vielzahl relevanter Umwelteinflüsse auf eine bestimmte Personengruppe umfaßt und in dem die Bedingungen von Gesundheit und Krankheit gestaltet werden können. Besonders sind Setting-Ansätze in Kommunen, Schulen, Krankenhäusern und Betrieben üblich (vgl. Baric 2001, Altgeld/Kolip 2004, Trojan 2004).

Am Beispiel dreier Settings (Kommune, Betriebe, Arbeitslose) soll im folgenden (Abschnitte 4.1 – 4.3) die *These 4* erläutert und begründet werden:

---

<sup>3</sup> Dies zum einen, weil der Band seinen Schwerpunkt auf die Darstellung der Möglichkeiten legt, Wissen wissenschaftlich zu überprüfen, zum anderen aufgrund der wissenschaftstheoretischen Position der Autoren (vgl. Fußnote 2), nach der sich eine solche Frage erst gar nicht stellt. Die Autoren gehen entgegen unserem (philosophischen) Evidenzbegriff (vgl. Abschnitt 2) von einem Evidenzbegriff aus, der empirische Hinterfragung zum Kerninhalt hat. Hierbei ist es ihre Grundthese, dass quantitative Untersuchungsmethoden und Ergebnisse nur im Kontext qualitativer, also hermeneutisch arbeitender Theoriebildungs- und Untersuchungsstrategien sinnvoll seien und im übrigen daher auch der „Streit der Fakultäten“ (ebda., S. 53) über naturwissenschaftlich-objektive und sozialwissenschaftlich-subjektive Erkenntnisse zumindest unmittelbar vor dem Ende stehe. Randomisierte kontrollierte Verlaufsstudien seien besonders aussagekräftig bei der Wirkungsanalyse pflegerischer und therapeutischer Interventionen. „Aber nicht alle Fragen lassen sich mit einem solchen Forschungsplan bearbeiten“ (ebda., S. 63). „Würden wir dabei (in diesem Buch, d.V.) noch die beliebten Ranglisten von Evidenz auflisten, würde unser Kurs vollends antiwissenschaftlich, nämlich dogmatisch. Was aber zwischenmenschlich (,intersubjektiv‘) nachprüfbares und durch Nachprüfung beständig verbessertes Wissen ist, davon handelt dieses ganze Buch.“ (ebda., S. 23)

„Eine quantitative Bewertung der Wirksamkeit von Gesundheitsförderung ist – jedenfalls in Teilgebieten – nicht unmöglich, sondern anzustreben.“

Die diesbezügliche Materiallage an wissenschaftlichen Studien oder anderen Dokumenten ist jedoch bei weitem nicht auf dem Stand, dass eine Evidenzhierarchie nach dem Modell der Cochrane-Methodik möglich oder sinnvoll wäre.

Der Diskurs über die Übertragbarkeit der Cochrane-Methodik auf die Gesundheitsförderung birgt daher auch die Gefahr, den Blick auf die Notwendigkeiten der konzeptionellen Entwicklung einer Evaluierbarkeit und Evaluierung komplexer sozialer Programme der Gesundheitsförderung zu verstellen.‘

#### **4.1 Kommunale Gesundheitsförderung**

Unter Bezug auf die Handlungsprinzipien der Ottawa-Charta handelt es sich bei der Gesundheitsförderung in Kommunen um folgende Aufgabenbereiche: „Befähigen; Kompetenzentwicklung und Empowerment; Organisationsentwicklung und Netzwerkbildung; Kommunikation und Dialoge; Mediation; Kooperations- und Konfliktmanagement; Bürgerbeteiligung; Selbsthilfe- und Netzwerkförderung“ (Trojan 2004, S. 310).

Auch wenn in die kommunale Gesundheitsförderung die gesundheitsbezogene Gemeinwesenarbeit mit einbezogen wird, wird der evaluative Kenntnisstand als mindestens lückenhaft bezeichnet. „Eine systematische Evaluierung der Gesundheitsverträglichkeit von Maßnahmen in Kommunalpolitik und –verwaltung gibt es bisher kaum [Trojan 2001]“ (Trojan 2004, S. 311). Von den – allerdings vor Verabschiedung der Ottawa-Charta konzipierten – gemeindeorientierten Präventionsprogrammen gegen Herz-Kreislaufkrankheiten bzw. deren somatische Risikofaktoren seien viele zwar erfolgreich gewesen, jedoch nicht in dem erwarteten Ausmaß. Hierbei werde dem Aspekt der Bürgerbeteiligung eine bedeutende Rolle zugeteilt, welche in den kardiovaskulären Präventionsstudien zu eng verstanden worden sei. „Als falsch erweist sich die Annahme, dass man mit beschränkten Instrumenten der Bürgerbeteiligung tatsächlich Einflüsse auf die vielfältigen sozialen und politischen Faktoren, die als Rahmenbedingungen für Gesundheitsförderung eine Rolle spielen, in den Griff bekommt“ (Trojan 2004, S. 312).

Daher soll im folgenden am Beispiel eines Ansatzes, bei dem die Partizipation im Mittelpunkt steht, die Frage deren Evaluierbarkeit und Evidenz beleuchtet werden.

Das von mir hierzu gewählte Projekt-Beispiel „Kiezdetektive – Kinderbeteiligung für eine gesunde und zukunftsfähige Stadt“ soll zunächst kurz vorgestellt werden.

„Trotz niedrigstem Sozialindex Berlins, höchster Arbeitslosenrate, zweithöchstem Anteil an Sozialhilfeempfänger/-innen und Migrant(inn)en, Wohnungen mit hoher Belegungsdichte und den daraus resultierenden Problemen, verfügt der Bezirk über viele wertvolle Ressourcen. Hierzu zählen die reiche Projektlandschaft, die Vielfalt der Kulturen, das hohe Potenzial an Selbsthilfe, nachbarschaftliche Kiezstrukturen, gute Modelle von Stadtplanung und Stadtplanungsentwicklung und eine lange Tradition der Bürgerbeteiligung.“

Hier setzt auch das Projekt zur Kinderbeteiligung Kiezdetektive an, eine Idee, die vom Kinder- und Jugendbüro Marzahn entwickelt wurde. In enger Kooperation mit dem Gesunde-Städte-Netzwerk und der Lokalen Agenda 21 wurde 1999 das Kinderbeteiligungsprojekt ins Leben gerufen: Kinder werden als Experten in eigener Sache in Planungs- und Entscheidungsprozesse zur nachhaltigen gesunden Stadtentwicklung und -gestaltung eingebunden.



Bei ihrer Tätigkeit als Kiezdetektive spüren sie vor allem Probleme in ihrem unmittelbaren Lebensraum auf und stoßen bei ihrer Suche auch auf Schätze. Die Ergebnisse werden in einer Ausstellung der Öffentlichkeit und auf einer Kinderversammlung den Bezirkspolitikern vorgestellt. Nach sechs Monaten werden auf einer Folgeversammlung die Umsetzungsergebnisse nachgefragt.

Für die Umsetzung sind die Politiker mit ihren jeweiligen Abteilungen verantwortlich. Die Kinder halten die Probleme auf "Denkzetteln" fest, die an die zuständigen Mitarbeiter zur Bearbeitung weitergegeben werden. Für die Ergebniskontrolle ist die Plan- und Leitstelle Gesundheit als Koordinatorin verantwortlich.

Ein Beispiel:

Bei ihrer Kiezbegehung entdeckten die Kinder der Kita Schlesische Str. die Seniorenfreizeitstätte in der Falckensteinstr. 6. Es fand ein gemeinsames Mittagessen zwischen Alt und Jung statt. Die Kinder haben den Alten einen Hip-Hop Tanz gezeigt; die Senioren haben getöpfert, getanzt und gesungen. Dabei haben die Kinder gelernt, dass auch die "Alten" Spaß haben können.

Weiterhin entdeckten die Kiezdetektive, neben der Senioreneinrichtung, ein verwildertes ungenutztes Grundstück. Sie hatten die Idee, dass man dieses gemeinsam umgestalten und nutzen könnte. Dieses Ergebnis stellten sie auf der Kinderversammlung den Politikern vor.

Die Idee wurde vom Stadtrat für Stadtentwicklung an das Quartiersmanagement zur Bearbeitung übergeben. Inzwischen wurde das Gelände zwischen Kita, Seniorenfreizeitstätte und einer angrenzenden interkulturellen Mädcheneinrichtung zu einer gemeinsamen Nutzfläche ausgebaut. (...)

Das Ziel des Projektes ist es, den Kindern die Chance zu geben, ihren unmittelbaren Lebensraum näher kennenzulernen, sich ihre Lebenswelt aktiv anzueignen und mitzugestalten. Des Weiteren fördert die aktive Teilnahme die allgemeine Entwicklung der Persönlichkeit sowie die Wahrnehmung, das Selbstbewusstsein und die Verantwortlichkeit. Ein weiterer wichtiger Faktor dieses Projekts zielt auf das Erleben demokratischen Handelns ab und stellt somit einen umfassenden Ansatz zur Gesundheitsförderung dar.

Insbesondere in problembelasteten Stadtgebieten, in denen ein hoher Anteil an Migranten lebt, können durch dieses Projekt Kinder erreicht und für gesundheitliche und soziale Belange aktiviert werden.

Eine Partizipation an politischen Prozessen auf Bezirksebene ist ausdrücklich erwünscht und als dauerhafte Einrichtung vorgesehen.“ (Papies-Winkler 2003)

Das überwiegend mit personellen Ressourcen aus kommunalen Einrichtungen und Freier Träger, aber auch u.a. im Rahmen der BKK-Initiative „Mehr Gesundheit für alle“ geförderte Projekt wurde hinsichtlich der Maßnahmen und der Reaktionen aus Politik und Verwaltung dokumentiert (Hänselmann/Schreiter 2004) und hinsichtlich des Partizipationsprozesses evaluiert. Kliemke/Daubitz/Grenner (2005, S. 7) erläutern zu Letzterem: „Betrachtet man die Zielsetzungen der partizipativen Verfahren von Kinderbeteiligung, z.B. das Erleben von demokratischem Handeln sowie die Sensibilisierung für gesundheits- und umweltbezogene Fragestellungen, so bieten die qualitativen Methoden einen guten Ansatz der Evaluation, da sie komplexe Lernprozesse oder Entwicklungen der Persönlichkeit sichtbar werden lassen“.

Es erübrigt sich, anhand dieses Beispiels nochmals herzuleiten, dass derartige Gesundheitsförderungsprogramme, die doch einen Kerninhalt der Ottawa-Charta treffen, hinsichtlich ihrer Evaluationsergebnisse nicht paßfähig für Evidenzhierarchien nach dem Modell der Cochrane Collaboration sein können und daher ausgeschlossen werden müßten,

würde Evidenzen von Gesundheitsförderungsprogrammen nach diesem Modell zu bewerten versucht.

Stattdessen soll hier abschließend auf ein anderes Problem hingewiesen werden. Wright (2004), der den partizipativen Ansatz in der Qualitätssicherung und Evaluation vertritt (d.h. mittels möglichst starker Beteiligung der Projektmitarbeiter/innen und der Zielgruppen), weist selbst darauf hin: „Mit der Konzentration auf ortsspezifische Lösungen stellt der partizipative Ansatz den üblichen wissenschaftlichen Anspruch auf allgemein gültige Evidenz für die Wirksamkeit einer Intervention grundsätzlich in Frage. (...) Das bedeutet konkret, dass die Beurteilung der Beweislage für die Wirksamkeit einer Präventionsstrategie letztendlich bei den Akteuren vor Ort liegt. Aus dieser Sicht kann es keine standardisierten Interventionen geben, die überall verwendbar sind; Evidenzen der Wirksamkeit sind immer ortsgebunden und prozessabhängig. Dennoch können auf Grund der Erfahrung in verschiedenen spezifischen Settings Aussagen über bestimmte Interventionsformen und Herangehensweisen getroffen werden, von denen die Akteure in vergleichbaren Settings lernen können. Mit anderen Worten: Nach dem partizipativen Ansatz wird Evidenz lokal erzeugt, die Diffusion von erprobten Innovationen findet durch einen Austausch zwischen vergleichbaren Settings statt“ (ebda., S. 315f.).

Auch für solche, an die Tradition von Aktionsforschung einerseits, Selbstevaluation in sozialen Projekten andererseits anknüpfende Ansätze werden Vorgehensweisen diskutiert, mittels externer Verfahren, wie z.B. Plausibilitäts-Peer-Reviews analog zu einem Gerichtsurteil (ebda., S. 316f.), der Gefahr der Geschlossenheit der Entscheidungsprozesse zu begegnen. Aus Sicht der Evaluationsforschung erscheinen solche Ansätze – wie insbesondere der der Selbstevaluation – im Grenzbereich zum Qualitätsmanagement zu liegen. Es erweist sich hier erfahrungsgemäß als schwierig, überhaupt Dokumentationen zu erzielen, die extern nachvollziehbar sind.<sup>4</sup> Für eine Diskussion der Übertragbarkeit des Evidenzmodells der Cochrane Collaboration auf die Gesundheitsförderung wird man jedoch Dokumentationen als Mindestvoraussetzungen zur Bewertung von Gesundheitsförderungsmaßnahmen fordern müssen, als Evaluator zudem eher einem externen Ansatz der Evaluationsforschung anhängen, der sich von Qualitätsmanagement abgrenzen läßt (Stockmann 2002, Elkeles/Kirschner 2003).

## **4.2 Betriebliche Gesundheitsförderung**

Es finden sich zur Betrieblichen Gesundheitsförderung (BGF) eher Übersichtsarbeiten zur Wirksamkeit verhaltensbezogener Programme bzw. entsprechender präventiver Elemente, also derjenigen Teilmenge von Gesundheitsförderung, die therapeutischem Vorgehen vergleichsweise nahe steht. Kreis und Bödeker (2003) fanden bei ihrer Sichtung von zwanzig Übersichtsartikeln zu verhaltenspräventiven Maßnahmen, nur in den wenigsten Fällen sei hierbei versucht worden, die Größe des Effekts in einer metaanalytischen Herangehensweise zu quantifizieren. Die deutschsprachige unter diesen drei ist die Metaanalyse von 27 vorwiegend als Stressimmunisierungs-Trainings durchgeführten Interventionen (Bamberg/Busch 1996) mit einer als gering zu bezeichnenden mittleren Effektstärke von 0,41. Vor allem scheint die Langfristigkeit der Wirkungen solcher Programme unsicher zu sein. Holling (1999) berichtet als wesentliches Ergebnis unter kontrollierten Bedingungen

---

<sup>4</sup> Auch Selbstevaluationen brauchen, um extern als glaubhaft und wirkungsvoll gelten zu können, wiederum externe Standards (vgl. Müller-Kohlenberg/Beywl 2003).

durchgeführter Streßbewältigungstrainings bei Polizisten<sup>5</sup> signifikante, jedoch nicht besonders hohe Effekte. Die Frage der Langfristigkeit solcher Effekte sei jedoch noch völlig offen.

Anhand von zwei Übersichtsarbeiten zur Wirksamkeit arbeitsweltbezogener Prävention von Rückenbeschwerden soll im folgenden dargestellt werden, welche Methoden bei der gegebenen Materiallage zur Bewertung entwickelt wurden und möglich sind. Da sich die untersuchten Maßnahmen nicht auf verhaltensbezogene Elemente beschränkten, war zu erwarten, dass recherchierbare Interventionen nur in sehr wenigen Fällen strengen methodischen Anforderungen epidemiologischer Interventionsstudien entsprechen. „Die Erfüllbarkeit (und damit: die Angemessenheit) solcher harter Kriterien im Rahmen sozialer – hier: betrieblicher – Interventionen ist ohnehin grundsätzlich anzweifelbar“ (Lenhardt/Elkeles/Rosenbrock 1997, S. 135). Epidemiologen begründen diese Skepsis mit der nicht streng zu kontrollierenden Komplexität von betrieblichen Einflüssen auf die Zielvariablen (Silverstein 1992) und den multidimensional auf verschiedenen Problemebenen ansetzenden Interventionen (Kilbom 1988, Wickström et al. 1993).

Osterholz (1993) führte eine „synoptische Evaluation“ von betrieblichen Maßnahmen zu Rückenschmerz durch. Es handelt sich um eine internationale Übersichtsarbeit zu 66 Studien, die in der Mehrzahl von positiven Effekten berichteten. Der Autor gliederte die Studien zu vier Maßnahmentypen (Ergonomie, Organisationsentwicklung, Rücken-/Nackenschulen<sup>6</sup> sowie Trainingsmaßnahmen) und kam zu der synoptischen Bewertung, die größten Erfolge hätten sich bei ‚ergonomischen‘ Maßnahmen gezeigt, insbesondere unter partizipativen Bedingungen, ergänzt durch Verhaltenstrainings, wobei es sich eher nicht um Effekte reiner Ergonomie gehandelt habe, sondern synergetische Effekte wahrscheinlich seien. Organisationsentwicklungs-Maßnahmen waren allerdings weniger gut dokumentiert. Die „rudimentären methodischen Anforderungen“ seien von vielen Studien erfüllt worden.

Um diese zu prüfen, wurde eine Liste von 11 Bewertungskriterien erstellt, nach der die Studien geprüft wurden (Abb. 2).

*Abb. 2: Prüfkriterien zur Einhaltung methodologischer Standards in Interventionsstudien*

- 
- Definition einer Kontrollgruppe
  - Größe der Gruppen ( $\geq$  50 Personen)
  - follow-up ( $\geq$  6 Monate)
  - loss to follow-up (weniger als 20% der Ausgangspopulation)
  - Randomisierung
- 

<sup>5</sup> Im Rahmen eines Simulationsexperiments wurde ermittelt, inwieweit die im Training (einwöchiges Programm) vermittelten Verhaltensweisen in fünf realitätsnahen Rollenspielen unter Feldbedingungen (Familienstreit, Fahrzeugkontrolle, aufdringlicher Bürger, Verdächtigung eines Pkw-Diebstahls und Provokation durch Studenten) Anwendung fanden. Eine Gruppe trainierter und eine Gruppe nicht trainierter, zufällig ausgewählter Beamter führten hierzu Rollenspiele auf einen Polizeigelände durch. Die Ergebnisse ergaben hinsichtlich der meisten vermittelten Trainingsinhalte bedeutsame Unterschiede zugunsten der trainierten Beamten. Jedoch konzentrierten sich diese Unterschiede vorwiegend auf die im Training behandelte Situation „Familienstreit“. Die Trainingseffekte waren bei den übrigen, insbesondere komplexeren Situationen bedeutend schwächer ausgeprägt, d.h. der Transfer der Lerninhalte auf nicht trainierte Situationen war unmittelbar nach der Durchführung der Integrierten Fortbildung gering.

<sup>6</sup> Vgl. als hierzu neuere Untersuchung auch: Schneider et al. (2004)

- Blinding
- Information über Compliance
- Grund für Wahl der Maßnahme angegeben
- adäquate Beschreibung der Maßnahme
- Grund für die Wahl des outcome
- adäquate Beschreibung des outcome

---

*Kriterium erfüllt: 1 Punkt; nicht erfüllt: 0 Punkte; Maximum :11 Punkte*

---

*Quelle: nach Osterholz (1993)*

Nach Ausschluß experimenteller Studien ergab die Analyse zu den betrieblichen Studien, dass keine die Maximalzahl von 11 erreichte, die besten Untersuchungen erhielten 5 Punkte (ebda., S. 83). Im einzelnen wurden zum Maßnahmebereich Ergonomie: 27 Punkte in 7 Studien (Durchschnitt 3,96 Punkte) vergeben, zum Maßnahmebereich Organisationsentwicklung: 20 Punkte in 5 Studien (Durchschnitt 4,0 Punkte), zum Maßnahmebereich Rücken-/Nackenschule: 26 Punkte in 5 Studien (Durchschnitt 5,2 Punkte) und zum Maßnahmebereich Fitneßtraining: 32 Punkte in 6 Studien: (Durchschnitt 5,33 Punkte). Die verhaltensbezogenen Studien hätten eher die klinischen Standards beachtet. Häufiger erfüllt worden seien die Forderungen nach Einbeziehung einer Kontrollgruppe, einer größeren Anzahl von Probanden und ein längeres follow-up: „Die wichtigsten Kriterien werden aber von einer erfreulich großen Zahl der Studien beachtet“, gibt der Autor an, nachdem er zuvor darauf hingewiesen hat, dass insgesamt nicht mehr Punkte erreicht worden seien, „liegt natürlich an dem Kriterienkatalog, der sich an dem traditionellen Kanon der Meta-Analyse anlehnt“ (ebda., S. 83).

Dieses Problem von Materiallage und Angemessenheit der Evaluationskriterien läßt sich abschließend nochmals verdeutlichen, reanalysiert man die Punktzahlen für alle untersuchten Studien insgesamt: In den 23 von Osterholz (1993) analysierten betrieblichen Präventionsstudien wurden insgesamt 104 von 253 möglichen Gütepunkten erreicht, das sind durchschnittlich 4,5 bei 11 möglichen Punkten bzw. 40,9% Erreichungsquote. Fragen einer „Evidenzhierarchie“ diskutiert Osterholz (1993) nicht. Anhänger harter epidemiologischer Kriterien könnten aber ableiten, die 11 möglichen Gütepunkte seien in dem Schema nicht als Maximum zu werten, sondern als Minimum zu fordern. In dieser Hinsicht hat die zweite Übersichtsarbeit, die eine auf Osterholz (1993) aufbauende Bewertungsmethodik verwendet, ein zusätzliches Kriterium für den Mindeststandard eingeführt (s.u.).

Diese zweite Übersichtsarbeit (Lenhardt 2003) ist eine Bewertung der Wirksamkeit betrieblicher Gesundheitsförderung durch Krankenkassen. Der Publikation lag eine im Auftrag des Sachverständigenrats für die Konzertierte Aktion im Gesundheitswesen erstellte Kurzstudie in Bezug auf Rückenbeschwerden bzw. diesbezügliche Arbeitsunfähigkeit (AU wegen Dorsopathien oder Muskuloskeletalen Erkrankungen [MSE]) zugrunde. Das Material bestand aus 13 Dokumenten, die auf Anfrage von 4 Trägern der gesetzlichen Krankenversicherung zur Verfügung gestellt worden waren. Dabei handelte es sich um 11 Einzelbetriebe, 1 Handwerksinnung, sowie eine 2000köpfige Beschäftigtenpopulation aus 16 Betrieben. Weitere Einschränkungen waren: eine Verzerrung durch Überrepräsentanz „positiver Fälle“ sowie die Struktur und Qualität der Materialien selbst: Die Daten waren überwiegend nicht zu wissenschaftlichen Zwecken erhoben worden (AU) und die meisten erhaltenen Materialien bezogen sich nicht speziell auf Rückenleiden. „Bei enger Auslegung der Zielstellung hätten somit fast alle Materialien aus der Betrachtung ausgeschlossen werden müssen“ (ebda., S. 25).

Der Autor weist eingangs auf einige Grundbedingungen hin: die Durchführung einer Einzelmaßnahme sei noch keine BGF. Betriebliche Gesundheitsförderung sei Intervention in ein komplexes, dynamisch sich entwickelndes Sozialsystem. Zumindest partiell beinhalte sie die Veränderung der dort vorherrschenden Strukturen, Organisationsroutinen, Handlungsmuster, Sichtweisen, Kooperationsformen und Steuerungsstrukturen. Es gebe beträchtliche Hindernisse, Veränderungen gesundheitsbezogener Zielvariablen kausal auf eine solche Intervention zurückzuführen.

Daraus leitet Lenhardt (2003) ab, auch unter ‚Alltagsbedingungen‘, d.h. bei Material, wo die methodischen Ansprüche bescheidener als bei wissenschaftlichen Studien zu sein hätten, seien zumindest bestimmte Kernanforderungen an eine BGF-Evaluation zu stellen, wie z.B.:

- wenigstens grobe Bestandsaufnahme der Interventionsbedingungen;
- Aussagen über Ursache-Wirkungsbeziehungen im strengen Sinne seien kaum erzielbar, dennoch sei zumindest eine Vorher-Nachher-Messung mit Nachverfolgung und mit geeigneten Wirkungsindikatoren zu fordern, d.h. nicht ausschließlich auf AU-Zahlen gestützt. Verfahren der (retrospektiven) subjektiven Veränderungseinschätzung könnten eine „second best“-Lösung darstellen, wenn Vorher-Nachher-Messungen nicht durchführbar sind.
- Eine Bildung „echter“ Kontrollgruppen sei in der Regel nicht möglich, dennoch sollte eine pragmatische Ermittlung von Referenzwerten zur Plausibilisierung von Effekten vorgenommen werden.

Lenhardt (2003) benutzte ein quantifizierendes Bewertungsverfahren, bei dem im ersten Schritt sieben Studien-Methodenkriterien geprüft wurden. Wenn hier mindestens 50% der maximalen Punkte erreicht wurden (Mindeststandard, s.o.), wurden in einem zweiten Schritt zwei Ergebniskriterien geprüft (maximal 12 Punkte). Hierzu wurde definiert: Wurden mindestens sechs Punkte erreicht, sei die Annahme eines Maßnahmeeffektes plausibel; Wurden 75% der möglichen Punktzahl in beiden Kriterienbereichen erreicht, könne von erhöhter Plausibilität für eine starke Interventionswirkung ausgegangen werden. Damit solle aber kein Niveau wissenschaftlicher Exaktheit vorgespiegelt werden, sondern lediglich der Anspruch eingelöst werden, überhaupt Ansatzpunkte für eine transparente Bewertung zu schaffen und zu benutzen. Das Kriterienschema zeigt Tab.1.

*Tab. 1: Methoden- und Ergebniskriterien zur Bewertung der Wirksamkeit der dokumentierten betrieblichen Gesundheitsförderungsprojekte durch Krankenkassen*

A) Methodenkriterien	Punkte	A) Methodenkriterien	Punkte
1) <i>Beschreibung der Maßnahmen</i>		2) <i>Ausgewiesene Wirkungsindikatoren</i>	
- gar nicht	0	- Arbeitsunfähigkeit	1
- grob	0,5	- Beschwerden	1
- detailliert	1		
3) <i>Zusammenhang Veränderungen Belastungen/Krankh.indikatoren</i>		4) <i>Größe der Population</i>	
- nicht untersucht	0	- unter 50	0,5
- untersucht	2	- 50 bis 100	1
		- 100 und mehr	1,5
5) <i>Vorher-Nachher-Messung</i>		6) <i>Beobachtungszeitraum</i>	
- nein	0	- unter 6 Monate	0,5
- ja	2	- 6 Monate bis 1 Jahr	1
		- über 1 Jahr	1,5
7) <i>Vergleichswerte</i>			
- keine	0		

- andere Kollektive (Kassenmitglieder, andere Betriebe) 2
- Kontrollgruppe i.e.S. 3

B) Ergebniskriterien	Punkte	B) Ergebniskriterien	Punkte
8) Ausmaß der Beschwerde (AU-Reduktion)		9) Reduktion im Vergleich zu Referenzwerten	
- unter 5%	0	- bis 1,5-fach	2
- 5 bis 19%	2	- bis 2-fach	4
- 20 bis 39%	4	- mehr als 2-fach	6
- 40% und mehr	6		

*Quelle: nach Lenhardt (2003)*

Anhand dieser Kriterien ergaben sich folgende Ergebnisse: 12 der 13 Projekte überschritten die 50%-Mindestgrenze bei den Methodenkriterien (zwei davon nur knapp). Von den 12 verbleibenden Dokumenten erfüllten 10 mindestens 50% der Ergebnis-Maximalpunktzahl, d.h. hatten sie hatten Effekte in einer für eine positive Wirksamkeitseinschätzung erforderlichen Größenordnung. Im weit überwiegenden Teil der Fälle (77%) sprachen die Ergebnisse somit für die Annahme einer Reduzierung Dorsopathie- bzw. MSE-bedingter Beschwerden oder AU durch die BGF-Aktivitäten. Der strengeren Bedingung (75% der Gesamt-Höchstpunktzahl (= starker und hochplausibler Interventionseffekt) entsprach kein Projekt, aber drei Projekte lagen knapp unter 75% der Maximalpunktzahl.

Insgesamt sei mit dieser Evaluation ein fortgeschritteneres Niveau der Interventionen gegenüber den Beschreibungen bis Mitte der 90er Jahre feststellbar, allerdings hätten überwiegend nur relativ grobe Informationen über die durchgeführten Projekte und Maßnahmen vorgelegen, so dass eine Einschränkung der Bewertbarkeit gegeben sei. Hinsichtlich der AU-Zahlen berichteten die Dokumente von Rückgängen großer Spannweite. Die Annahme, dass es sich um Effekte der BGF-Aktivitäten handele, habe ein beachtliches Maß an Plausibilität. Effekte auf das Beschwerdeniveau waren überwiegend positiv, aber weniger eindeutig.

Lenhardt (2003) zieht als Fazit, eine Wirksamkeitsanalyse sei aufgrund der Zugänglichkeit und Qualität von Informationen über BGF-Maßnahmen der Kassen außerordentlich schwierig (selektierte Materialbasis, sehr heterogene Daten von recht begrenzter Aussagekraft). Hohen methodischen Ansprüchen genügende Verfahren der Wirksamkeitsbewertung nach Art der Metaanalyse seien auf dieser Basis nicht möglich, aber es seien doch wenigstens grobe Plausibilitätsabschätzungen vornehmbar. Dies sei ein "Niveau relativ 'weicher' Evidenz" (ebda., S. 35).

Hervorgehoben werden soll, dass die Materiallage bei der Betrieblichen Gesundheitsförderung durch Krankenkassen<sup>7</sup> es erforderlich machte, eine Grundvoraussetzung der externen Evaluierbarkeit, die Beschreibung der durchgeführten Maßnahme(n) mit einem Gütekriterium zu versehen, für das es 0 Punkte gibt, wenn gar keine Beschreibung vorliegt und bereits 0,5 Punkte, wenn wenigstens eine grobe Beschreibung der Maßnahme vorliegt, deren Wirksamkeit wissenschaftlich geprüft werden soll.

Fazit der Ausführungen zum Stand der Wirksamkeitsforschung der Betrieblichen Gesundheitsförderung ist somit insgesamt, dass – bis auf wenige Ausnahmen, die jedoch nur einzelne Maßnahmen wie z.B. Stressbewältigungstrainings, nicht hingegen komplexe Programme darstellen - quantifizierte Angaben, wo sie überhaupt aufgrund des

<sup>7</sup> Nach aller Evaluations- und Praxiserfahrung gilt das nicht nur für Krankenkassen, sondern für die Praxis der Gesundheitsförderung überhaupt.

Sachgegenstands einerseits, der Materiallage andererseits möglich sind, eher Versuche darstellen, erste Transparenzraster einer Bewertung zu schaffen. Keinesfalls geeignet sind sie für Evidenzhierarchien. Vielmehr ist es erforderlich, noch viel Entwicklungsarbeit zu leisten, um Betriebliche Gesundheitsförderung überhaupt evaluierbar zu machen.

### **4.3 Gesundheitsförderung für Arbeitslose**

In sozialepidemiologischer Hinsicht stellen Arbeitslose eine Gruppe mit besonders ausgeprägtem Bedarf für gesundheitsförderliche Interventionen dar. Verfügbare Untersuchungen weisen auf einen durchgängig schlechteren Gesundheitsstatus von Arbeitslosen im Vergleich zu Erwerbstätigen hin. Dies betrifft alle verfügbaren Indikatoren. Allerdings handelt es sich meist um Erkenntnisse auf der Basis von Querschnittsdaten, so dass Kausalaussagen zur Richtung von Ursache-Wirkungsbeziehungen bei ihnen nicht möglich sind. In Bezug auf bestimmte Krankheiten imponiert der „mental health“- Bereich und die Sucht mit erheblichen Prävalenzunterschieden zu Nichtarbeitslosen, aber auch in anderen Krankheitsgruppen ergeben sich erhebliche Prävalenzunterschiede. Nahezu alle gesundheitlichen Risiken der allgemeinen Lebensbedingungen und des Verhaltens sind bei Arbeitslosen in höherem Maße ausgeprägt.

Die tendenziell höhere Inanspruchnahme des medizinisch-kurativen Sektors von Arbeitslosen korrespondiert nicht mit einer höheren Aufgeschlossenheit gegenüber präventiven bzw. gesundheitsförderlichen Angeboten sowie generellen Hilfsangeboten. Gerade das Gegenteil ist der Fall. Eine Konzeptionierung von gesundheitlichen Interventionen und zielgruppenspezifischen Gesundheitsförderungsprogrammen für Arbeitslose, die allerdings kein eigentliches Setting darstellen, da ihnen eine institutionelle Verortung gerade fehlt, steht also vor besonderen Herausforderungen.

Vor dem Hintergrund erstellten wir für den Bundesverband der Betriebskrankenkassen ein Gutachten (Elkeles/Kirschner 2004), das eine empirisch verlässliche Grundlage für die Konzipierung und spätere Umsetzung von Maßnahmen legen sollte, die das Ziel verfolgen, den gesundheitlichen Zustand und die gesundheitlichen Kompetenzen von Arbeitslosen zu verbessern. Ein wesentlicher Bestandteil war, Projekte mit gesundheitlichen Interventionen in dieser Zielgruppe in Deutschland und anderen Ländern hinsichtlich ihrer Evidenz zu recherchieren, zu beschreiben und zu bewerten.

Das Gutachten basiert im wesentlichen auf fünf Methoden: der Analyse und Aufbereitung der wissenschaftlichen Literatur, umfangreichen Projektrecherchen zur Identifikation von Interventionen, der einzelnen Beschreibung und Bewertung der Interventionen, der kritischen Gesamtbewertung der Interventionen sowie ergänzenden Gesprächen und Interviews mit verschiedenen Experten. Materialbasis waren Förderprogramme, Datenbanken (z.B. BZgA), Internet und eigene Recherchen. Reviews waren nicht zu finden, Evaluierungen nur in Ausnahmen, sondern meist Projektbestandsaufnahmen. Die recherchier- und untersuchbare Basis bestand aus 36 Projekten. Zusätzlich wurden 15 Projekte aus einem EU-Programm (1999-2000) mit 55 ‚models of good practice‘ hinsichtlich ihrer Evaluationsergebnisse im Rahmen einer Bachelorarbeit an der Fachhochschule Neubrandenburg nachbefragt (Simon 2003, 2004).

Von den 55 ‚models of good practice‘ hatten im Abschlußbericht 64% eine Evaluation angegeben, wobei sich diese auf Selbstevaluationen zu heterogenen Parametern beschränkten (Teilnehmerzahlen, Wiederbeschäftigte). Es lagen jedoch keine Informationen zu outcomes psychischer Gesundheit vor. Diese Evaluationsdefizite seien erkannt und daher geplant worden „to make an official evaluation in the future“. Für die 2003 durchgeführte Nachrecherche wurden 28 Projekte als innovativ ausgewählt, realisierbar war sie bei 15

Projekten (Telefoninterviews und schriftliche Befragung). Die Nachrecherche ergab 13 interne Bewertungen für den Gesamtzeitraum (Teilnehmer-Feedback mittels schriftlicher Zufriedenheitsbefragungen zu unterschiedlichen Zeitpunkten, Gruppendiskussionen während und/oder am Projektende, individuelle Interviews mit Kursteilnehmern, Beobachtungen und Eigenreflexionen durch Projektverantwortliche und Träger), jedoch keine Modifikation des Vorergebnisses, dass keinerlei outcomes vorlagen (Simon 2003, 2004). Insbesondere hatten die Projektdokumente auch meist kaum extern nachvollziehbar machen lassen, worin die Massnahmen konkret bestanden hatten.

Von allen in unserem Gutachten recherchierten 36 Projekten hatten fünf Projekte ein Kontrollgruppen-Design (13,9%).

Beschränkt man sich auf Projekte mit primären bzw. noch erkennbaren gesundheitlichen Interventionen (viele hatten gegenstandsgemäß einen arbeitsmarktbezogenen Ansatz), verblieben 22 hinsichtlich ihrer Evidenz untersuchbarer Projekte. Auch für diese galt, dass Zielgruppen, Ziele und Maßnahmen häufig heterogen waren und Teilnehmerzahlen häufig klein (häufig auch keine Angaben).

Auch wenn auf der Grundlage der Projektrecherchen wirklich belastbare Quantifizierungen nirgends vorgenommen werden konnten, war bereits die Qualität der Dokumentation und Evaluation von Projekten sehr heterogen. Bei den insgesamt 22 Vorhaben gab es bei 8 keine Dokumentation und Evaluation (Tab. 2).

*Tab. 2: Standards der Dokumentation and Evaluation*

Keine Dokumentation verfügbar <sup>a</sup>	8
Evaluationsbericht	2
Externe Evaluation	1
Selbstevaluation	3
Vorläufige Analysen	1
Publication verfügbar	7
Total	22
<sup>a</sup> im Rahmen des Gutachtens (6 Monate)	

*Quelle: Elkeles/Kirschner (2004), S. 214*

Insoweit die Projekte evaluiert worden sind oder evaluiert werden sollten, stand aber offenbar nach wie vor die Methode der internen Evaluation durch die Projektnehmer selbst auf der Tagesordnung. Dies ist aber (s.o.) ein eher suboptimales Verfahren.

Insgesamt ist anzumerken, dass die Qualität der Projektdokumentation und -evaluation dringend verbesserungsbedürftig ist. Mag man bei nur knapp ausfinanzierten Projekten eine akzeptable Evaluation schon gar nicht erwarten können, so doch jedenfalls eine saubere Projektdokumentation. Überraschend war, dass selbst bei durchgeführten Evaluationen auch im universitären Umfeld zum Teil deutliche Evaluationsmängel festzustellen waren.

Dieser Befund wurde durch die Bachelorarbeit an der Fachhochschule Neubrandenburg (Simon 2003) bestätigt: „Mit Blick auf die befragten Projekte bezüglich dieser Voraussetzungen zur Evaluationsmessung wurden bereits hinsichtlich der Ausgangsanalyse und der genauen Zieldefinierung Defizite sichtbar. In wissenschaftlichen Begleitforschungen zu anderen Projekten wurden Längsschnittuntersuchungen mit Eingangs-, Abgangs-, Verbleibs- und Abbrecherbefragungen bei den Teilnehmern sowie in Kontrollgruppen durchgeführt, um beispielsweise Veränderungen in den Demoralisierungs- und Selbstwirksamkeitswerten der Teilnehmer über den Kursverlauf vergleichen zu können.



Diese Bedingungen waren nach dem zu Grunde liegenden Informationsstand in diesen befragten Projekten nicht gegeben.“ De facto wurde bei den 15 von Simon (2003) nachuntersuchten Projekten gerade ein Projekt gefunden, dem ein einigermaßen akzeptables Evaluationsdesign zugrunde lag.

Die Ursachen hierfür müssen hier im Einzelnen nicht beleuchtet werden. Für die möglichen Auftraggeber von Maßnahmen der Gesundheitsförderung und deren Evaluation sowie für die Diskussion über eine Evidenzbasierung der Gesundheitsförderung muss aber darauf hingewiesen werden, dass belastbare Evaluationen zunächst eine ausreichende finanzielle Ausstattung benötigen, wobei der inzwischen diffundierende Anteilswert von 10% der Projektkosten für die Evaluation allenfalls ein grobes Raster sein kann und bei Projekten unter 500.000 Euro häufig zu nicht ausreichenden Evaluationsbudgets führen wird.

Zur Effektivität und Effizienz der Interventionsmaßnahmen ließen sich nur Ergebnisse zur Wirksamkeit von Interventionen zugrunde legen, die – wenn auch in ganz unterschiedlicher Weise – auf positive „mental health“ Effekte zielen (z.B. Stressabbau, Depressivität in sieben Projekten).

Die Erfahrungen in diesem Interventionsfeld basieren im wesentlichen auf den Arbeiten der MRPC-Gruppe mit dem Job Search Program (JOBS) in den USA (Caplan et al. 1997), dem methodisch vergleichbaren Vorhaben von Vuori in Finnland (Vuori et al. 2002) und dem CBT-Ansatz von Proudfoot in Großbritannien (Proudfoot et al. 1997). Der Kern des Job Search-Programms besteht aus einem einwöchigen Training, welches sowohl praktische Fähigkeiten der eigenaktiven Arbeitssuche als auch psychosoziale Bewältigungsstrategien für die Situation des Arbeitsplatzverlustes vermitteln soll. Die Ergebnisse der JOBS-II-Evaluation zeigten bei den Teilnehmenden eine schnellere Wiederbeschäftigung in höherer Qualität (nach 1,5 Monaten eine um 7% höhere, knapp signifikante Wiederbeschäftigungsquote in der Fallgruppe, nach 4 Monaten 59% in der Fallgruppe, 51% in der Kontrollgruppe). Hervorzuheben ist, dass es sich bei der Zielgruppe um Kurzarbeitslose handelte. Die insgesamt hohe Wiederbeschäftigungsquote spiegelt sicher auch die Dynamik des US-amerikanischen Arbeitsmarktes wieder.

Mit dem Työhön Job Research Program wurde der Ansatz auf ein Gebiet in Finnland übertragen, wo nicht nur Kurzarbeitslose einbezogen wurden. Im Fall-Kontrollgruppenvergleich konnten keine signifikanten Effekte auf die Wiederbeschäftigung, die Arbeitszufriedenheit und depressive Symptome gezeigt werden (vgl. Vuori et al. 2002). Die Autoren führen dies darauf zurück, dass Arbeitslose in Finnland aufgrund einer längeren Bezugsdauer und höheren Leistungen beim Arbeitslosengeld nicht im vergleichbaren Maß an Stress und depressiven Symptomen leiden wie Arbeitslose in den USA, so dass mit den identischen Interventionen hier keine Effekte gezeigt werden konnten.

In einer britischen Interventionsstudie an langzeitarbeitslosen Angestellten fanden in der Interventionsgruppe kognitionsorientierte Trainings statt, in der Kontrollgruppe Coaching-Maßnahmen zur Mobilisierung sozialer Unterstützung. Für beide Gruppen ließen sich signifikante Verbesserungen der psychischen Gesundheit, der Leistungsfähigkeit sowie des Arbeitsmarktverhaltens zeigen. Der Kontrollgruppe gelang eine Wiederbeschäftigung häufiger und stabiler. Allerdings war die Stichprobe mit n=199 recht klein (vgl. Proudfoot et al. 1997).<sup>8</sup>

---

<sup>8</sup> Für eine ausführlichere und eingehendere Bewertung zu den verhaltenstherapeutischen Interventionsstudien zu Arbeitslosen vgl. Pröll (2004) und Kirschner (2004).

In synoptischer Form zeigt Abb. 3 die Ergebnisse in diesem vergleichsweise gut durch Studien bearbeiteten Interventionsbereich.

Abb. 3: *Hinweise zur Wirksamkeit von Interventionen bei "mental health"-Indikatoren*

	Hinweise zur Wirksamkeit	
	Positiv	Negativ
Aktiva Dresden (58)		X
EVA Siegen (59)		X
MPRC Michigan (60,61)	X	
Työhön (62)	X	
Proudfoot et al. (63)	X	
Muller (64)	X	

*Quelle: Elkeles/Kirschner (2004), S. 215*

Das Ergebnis ist also unentschieden. Berücksichtigt man auch noch die geringen Fallzahlen des Projekts 36, so ist der Erkenntnisstand zur Wirksamkeit dieser Interventionen nahezu als dürftig zu bezeichnen. Dieser Befund wird u.a. von Kuhnert (1999) bestätigt: „Die bisherige Forschungslage gibt jedoch kaum Hinweise auf die Effektivität des SMT (Stressmanagementtrainings) für die Zielgruppe der Langzeitarbeitslosen“ (ebda., S. 230), um sogleich die Studie von Proudfoot et al. (1997) als „eine der wenigen sorgfältig evaluierten Studien eines durchgeführten SMT“ zu bezeichnen (ebda., S.231).

Eine Gesamtbewertung der Projekte zeigte folgende Ausrichtungen sowie Schwächen und Probleme:

- Eine starke Konzentration auf „mental health“ und Stressbewältigung einerseits, wie aber auch auf unspezifische Maßnahmen andererseits;
- Eine unklare Bedarfslage für (weitere) gesundheitliche Interventionen und Indikationen;
- Eine unklare Lage hinsichtlich der Akzeptanz und Inanspruchnahme entsprechender Angebote;
- Eine – von Pflichtberatungen abgesehen – insgesamt geringe Nutzung;
- Zum Teil erhebliche Implementationsschwächen;
- Eine höchst defizitäre Dokumentations- und Evaluationspraxis;
- Eine unklare Evidenzlage hinsichtlich der Wirksamkeit und Wirtschaftlichkeit von Stressbewältigungsstrategien (Situation von conflicting results) bzw. von gesundheitsfördernden Maßnahmen überhaupt.

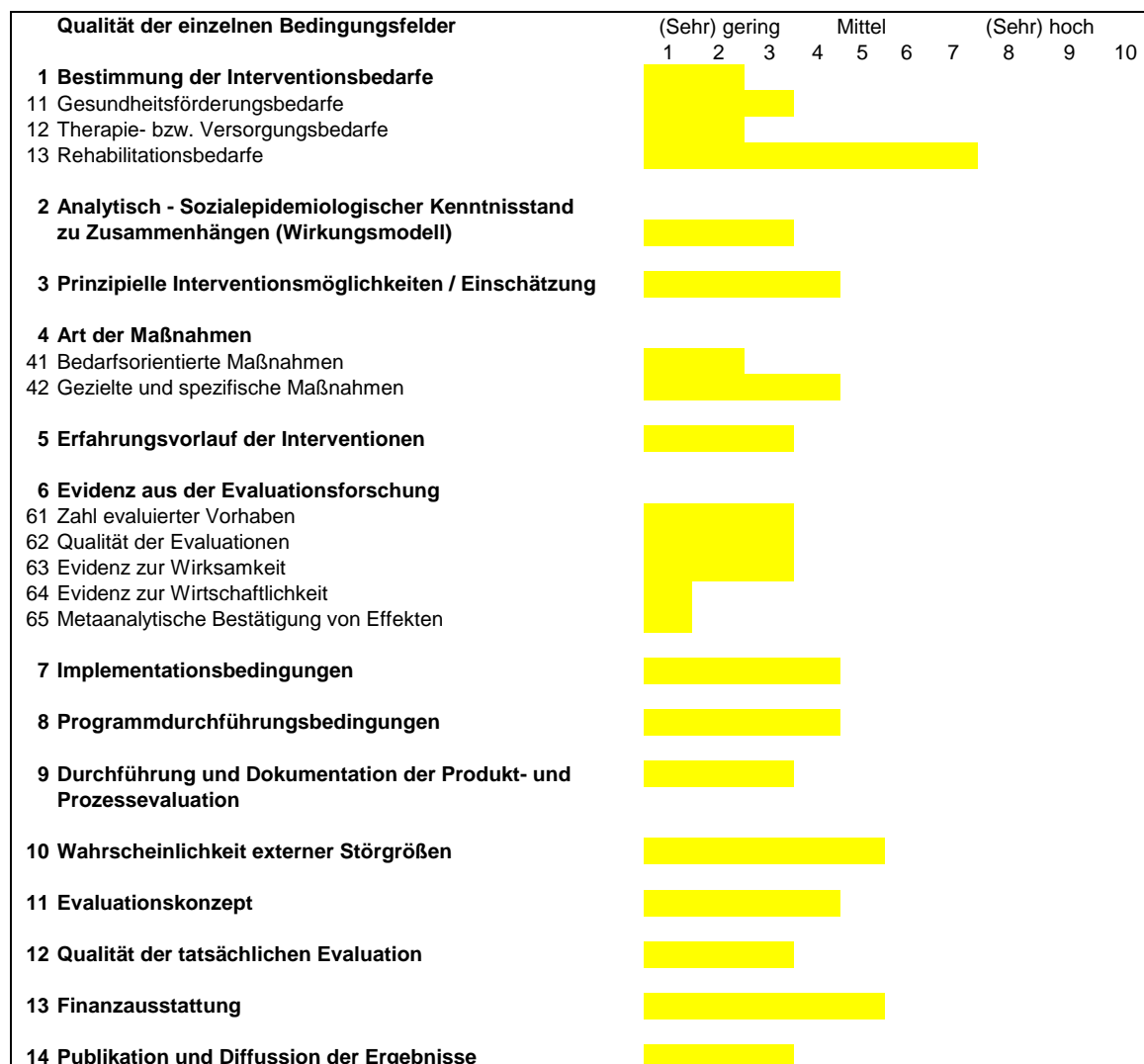
Programme, d.h. der gezielte Einsatz (wirksamer) Maßnahmen zur Erreichung bestimmter Ziele, sind oft komplexe sozialwissenschaftliche Experimente mit dem Ziel, gesundheitliche oder soziale Problemlagen in angebar positiver Richtung zu verändern.

Der Erfolg derartiger Interventionen ist zunächst generell unsicher, da sie häufig sozialökonomisch bedingte und oft über Jahre verfestigte Problemlagen, die sich mit je individuellen Lebensgewohnheiten und Verhaltensweisen mischen und kumulieren, mit in der Regel vergleichsweise einfachen, zeitlich begrenzten und wenig kostenintensiven Interventionsmaßnahmen zu verändern suchen. Zugleich müssen die erzielbaren Effekte möglichst zeitstabil wirken. Gleichwohl lassen sich bestimmte Bedingungen ableiten, die erfüllt sein müssen, um entsprechende Programme schon ex-ante als eher erfolgreich oder eher erfolglos einzuschätzen.

In Abb. 4 sind die wesentlichen Bedingungen hierfür benannt. Der Erfolg einer vorzunehmenden Intervention ist vom sozialepidemiologischen Kenntnisstand (Wirkungsmodell), der prinzipiellen Interventionsfähigkeit, der Bedarfsgerechtigkeit und Qualität der Interventionsmaßnahmen, dem Erfahrungsvorlauf ähnlicher Interventionen und den daraus sich ergebenden Befunden zur Evidenz in der Evaluationsforschung bis hin zu den jeweiligen Implementations- und Durchführungsbedingungen der konkreten Intervention und Evaluation und deren Qualitäten und schließlich auch der Finanzausstattung abhängig.

Gleichzeitig ist in Abb. 4 als eine Gesamtbewertung angegeben, wie der jeweilige Stand im Interventionsfeld Arbeitslosigkeit und Gesundheit auf Grundlage der analysierten Materialien von uns eingeschätzt wurde. Hierzu wurde der – im Einzelnen sicher diskutierbare – Versuch gemacht, die in den verschiedenen Kontextbereichen vorfindlichen Standards zwischen den Polen einer Intervention mit ganz schlechten Erfolgsbedingungen und einer Intervention mit hervorragenden Erfolgsbedingungen auf einer 10er Skala zu markieren.

Abb. 4: Einschätzung der Bedingungen für den Erfolg von Interventionen



Quelle: Elkeles/Kirschner (2004), S. 236

Es handelt sich also um eine qualitative Evidenzbewertung in quantifizierter Darstellungsweise, wie es dem Gegenstand angemessen erschien. Die Bewertung begründet sich besonders damit, dass:

- bedarfsorientierte Herangehensweisen nur in geringem Maße vorkamen, die wenn überhaupt auch nur die gesundheitsfördernde Seite berücksichtigten und nicht den Bereich der Therapie,
- der sozialepidemiologische Kenntnisstand zur Interventionsfähigkeit als noch relativ gering bezeichnet werden muss,
- die prinzipiellen Interventionsmöglichkeiten etwas optimistischer eingeschätzt werden können, wobei diese Bewertung auf der Einschätzung einer Wirksamkeit gerade unspezifischer Maßnahmen der Gesundheitsförderung basiert,
- der Erfahrungsvorlauf entsprechender Interventionen als recht gering einzuschätzen ist und
- entsprechend die Evidenz in der Evaluationsforschung recht gering ist und sich metaanalytisch gesicherte Wirksamkeitsnachweise in diesem Feld nicht finden lassen,
- die Implementations- und Durchführungsbedingungen der Projekte sich jeweils unterschieden, im Mittel aber auch als nur durchschnittlich zu bezeichnen waren,
- aus der insgesamt unzureichenden Ergebnisevaluation auch auf Evaluationsmängel in der Produkt- und Prozessevaluation geschlossen werden kann,
- die Wahrscheinlichkeit externer Störgrößen als durchschnittlich einzuschätzen ist,
- die Evaluationskonzepte und die Evaluationsstandards nur unterdurchschnittlich, im Einzelfall aber auch höchst defizitär waren,
- die Finanzausstattung recht unterschiedlich war, insgesamt aber sicher als durchschnittlich bezeichnet werden kann,
- Dokumentationen und Publikationen defizitär sind.

Insgesamt sind damit die Erfolgsbedingungen und möglichen Erfolge von gesundheitlichen Interventionen und Maßnahmen der Gesundheitsförderung bei Arbeitslosen beim bisherigen Niveau von Intervention und Evaluation als eher kritisch denn als auf einem Stand, der Quantifizierungen von Evidenz i.S. von Hierarchisierungen möglich und sinnvoll macht, einzuschätzen.

## **5. Schlußfolgerungen**

Aus allem dürfte deutlich geworden sein, dass eine stärkere Ausrichtung schon der Gesundheitsförderung auf die Zielgruppe der Arbeitslosen, aber auch auf andere Bereiche der Gesundheitsförderung, die Lösung einer ganzen Vielzahl inhaltlicher, methodischer und konzeptioneller Probleme erfordert. Dies gilt in noch stärkerem Maße für gesundheitliche Interventionen, die den rein gesundheitsfördernden Bereich überschreiten und therapeutische und rehabilitative Maßnahmen in Ergänzung oder als Voraussetzung mit umfassen, wie es bei einem großen Teil der Arbeitslosen erforderlich erscheint.

Ein Ingangsetzen derartiger Maßnahmen, so *These 5*, erfordert schließlich auch eine angemessene Revision des 'Evidenzpostulats' mit dem Ziel, die Projektförderungsmöglichkeiten von einer ex ante-Einschätzung einer hohen Plausibilität von positiver Programmwirkung abhängig zu machen und entsprechende Hypothesen in einer fundierten Evaluation zu testen und nachzuweisen (vgl. Elkeles/Kirschner 2004, Kirschner/Elkeles/Kirschner 2004a).

Es muss darauf ankommen:

- die in den einzelnen Feldern erkannten offenen Fragen, Schwachstellen und Probleme zu lösen bzw. einer Lösung näher zu bringen,
- die Qualität von bereits laufender und geplanter Intervention und Evaluation zügig zu verbessern,
- die auf der Grundlage des gegebenen wissenschaftlichen Kenntnisstandes vorzunehmende Plausibilisierung, dass mit den vorgeschlagenen Interventionen positive Effekte in einem grob abschätzbaren Umfang erreichbar erscheinen, was in der Regel die Skizzierung eines Wirkungsmodells umfassen wird,
- zu testende Hypothesen zu formulieren,
- Qualität in der Produkt- und Prozessevaluation durch qualitätssichernde Maßnahmen sicherzustellen,
- Projekte und Programme auf adäquate Methoden der Ergebnisevaluation, in der die abgeleiteten Hypothesen zu überprüfen sind, zu verpflichten.

Das im Zuge der zunehmenden Diskussion um Evidenzbasierung auch in die Primärprävention eingeführte faktische 'Evidenzpostulat' in Zusammenhang mit dem § 20 SGB V zum ex ante-Nachweis der Effektivität und Effizienz von Interventionen ist jedoch am klassischen Fall-Kontrolldesign recht einfacher Interventionen und Evaluationen orientiert. Die Rigorosität, mit der es eingefordert wird, geht an der Realität der theoretischen Fundierung, instrumentellen Umsetzung und evaluativen Nachweissicherung von Wirkungen durch gesundheitliche und gesundheitsfördernde Interventionen jedoch völlig vorbei. Auch wenn es ggf. als Anreiz gedacht ist, die Evaluationspraxis zu verbessern, scheint es die Weiterentwicklung von Interventionen und Evaluationen eher zu behindern denn zu befördern. Die ex-ante Evidenzforderung ist daher zugunsten einer ex-ante Plausibilitätsschätzung positiver Wirkungen bei Verpflichtung auf rigorose Qualitätssicherung und Evaluation zu revidieren (vgl. Elkeles/Kirschner 2004, Kirschner/Elkeles/Kirschner 2004a).

Die Tatsache, dass sich in der Cochrane Library zunehmend auch Reviews zu Public Health und Gesundheitsförderung finden, steht dieser These nicht entgegen. Denn entweder decken diese zu einem großen Teil Impfmaßnahmen und die Applikation anderer präventiver Mittel ab und stellen damit gerade keine (komplexen) gesundheitsförderlichen Programme dar oder es handelt sich zwar um Interventionen mit komplexen Mitteln, jedoch dann zu einem sehr eingeschränkten, mit traditionellen Ergebnisindikatoren meßbaren Einzelthema (Unfallraten, Tabakprävention, Kondomgebrauch). Das jedoch ist nicht der Kern, welchen es an Bedarf gibt, Gesundheitsförderungsprogramme zu evaluieren.

Maßnahmen der primären Prävention und Gesundheitsförderung, die durch ein epidemiologisches oder gesundheitswissenschaftliches Wirkungsmodell begründet sind, haben – je nach dem Stand der wissenschaftlichen Erkenntnis – ein angebbar klares Niveau von Evidenz. Bei Maßnahmen der sogenannten salutogenetischen Gesundheitsförderung im Sinne der Stärkung von Ressourcen (z.B. Antonovsky [1997], auf den sich die Gesundheitsförderer gerade in Deutschland häufig alleinig berufen), ist die Evidenz gesichert, wenn der Nachweis erbracht wurde, dass durch Maßnahmen der Ressourcenstärkung sich Gesundheitsniveaus verbessert haben. Dieser Nachweis ist aber z.B. gerade für den „sense of coherence“ von Antonovsky bisher nicht erbracht worden (Bengel et al. 1998, Geyer 2000).

Die Rigorosität, mit der das 'Evidenzpostulat' gerade in der Prävention und Gesundheitsförderung eingefordert wird, nimmt den hier insgesamt geringen Stand wissenschaftlicher Erkenntnis nicht zur Kenntnis (vgl. Elkeles/Kirschner 2004, Kirschner/Elkeles/Kirschner 2004a). Bei aller Berechtigung, die die Forderungen nach einer sozialwissenschaftlichen und ökonomischen Evaluation auch von Maßnahmen der Gesundheitsförderung und Prävention auch haben, darf nicht übersehen werden, dass noch

erhebliche theoretische und instrumentelle Defizite bestehen, die Praxis gesundheitlicher Interventionen und auch der Gesundheitsförderung gerade bei komplexen, settingbezogenen Ansätzen belastbar zu evaluieren.. Vor diesem Hintergrund ist davon auszugehen, dass sich eine angemessene und belastbare Evaluation von Maßnahmen der Gesundheitsförderung – jedenfalls im Bereich der Ergebnisevaluation – bei komplexeren Programmen erst im Laufe einiger Jahre entwickeln lassen. Insgesamt ist zu befürchten, dass durch das ‚Hochschrauben‘ der Erwartungen an die Standards von Evaluationen und deren Evidenz im Bereich der Gesundheitsförderung und Prävention die notwendige Kreativität in der Konzeption, Durchführung und Evaluation insbesondere sozialkompensatorischer Projekte eher gelähmt als gefördert wird.

## **Literaturverzeichnis**

- Antonovsky, Aaron (1997): Salutogenese. Zur Entmystifizierung der Gesundheit. DGVT-Verlag: Tübingen.
- Altgeld, Thomas/Kolip, Petra (2004): Konzepte und Strategien der Gesundheitsförderung. In: Hurrelmann, Klaus/Klotz, Theoder/Haisch, Jochen (Hg.): Lehrbuch Prävention und Gesundheitsförderung. Verlag Hans Huber: Bern, Göttingen, Toronto, Seattle, S. 41-51.
- Badura, Bernhard/Siegrist, Johannes (Hg.) (2002): Evaluation im Gesundheitswesen. Ansätze und Ergebnisse. 2. Aufl., Juventa: Weinheim und München.
- Bamberg, Eva/Busch, Christine (1996): Betriebliche Gesundheitsförderung durch Stressmanagementtraining: Eine Metaanalyse (quasi-)experimenteller Studien. Zeitschrift für Arbeits- und Organisationspsychologie 40: 127-137.
- Baric, Leo/Conrad, Günter (2001): Gesundheitsförderung in Settings. Konzept, Methodik und Rechenschaftspflichtigkeit zur praktischen Anwendung des Settingansatzes der Gesundheitsförderung. Verlag für Gesundheitsförderung G. Conrad: Gamburg.
- Behrens, Johann/Langer, Gero (2004): Evidence-based Nursing. Vertrauensbildende Entzauberung der Wissenschaft. Verlag Hans Huber: Bern, Göttingen, Toronto, Seattle.
- Bengel, Jürgen/Strittmatter, Regine/Willmann, Hildegard (1998): Was erhält Menschen gesund? Antonovskys Modell der Salutogenese – Diskussionsstand und Stellenwert. Bd. 6 der Schriftenreihe Forschung und Praxis der Gesundheitsförderung. Bundeszentrale für gesundheitliche Aufklärung: Köln.
- Caplan, R.D./Vinokur, A.D./ Price, R.H. (1997): From job loss to reemployment: field experiments in prevention-focused coping. In: Albee, G.W./Gullotta, T.P. (Eds.): Primary prevention works. Thousand Oaks, CA: Sage Publications, Issues in children's and families' lives (16), 341-379.
- Clemens, Wolfgang/Strübing, Jörg (Hg.) (2000): Empirische Sozialforschung und gesellschaftliche Praxis. Leske + Budrich: Opladen.
- Der Brockhaus in Text und Bild (2002). Das Lexikon in der PC-Bibliothek. Version 3.0. F.A.Brockhaus: Leipzig, Mannheim.
- Elkeles, Thomas (2002): Qualitätsicherung und –entwicklung im betrieblichen Gesundheitsmanagement. Arbeit und Sozialpolitik 56, 7/8: 33–40.
- Elkeles, Thomas/Georg, Arno (2002): Bekämpfung arbeitsbedingter Erkrankungen. Evaluation eines Modellprogramms. Juventa: Weinheim/München.
- Elkeles, Thomas/Georg, Arno (2005): Evaluation komplexer Modellprogramme am Beispiel eines gesundheitlichen Präventionsprogramms. In: Loidl-Keil, Rainer/Zapotoczky, Klaus

(Hg.): Evaluationen im Gesundheitswesen. Konzepte, Beispiele und Erfahrungen. Im Erscheinen.

Elkeles, Thomas/Kirschner, Wolf (2003): Evaluation im Gesundheitswesen. Lehrmodul. Hochschule für Angewandte Wissenschaft Hamburg, Forschungsschwerpunkt Public Health. <http://www.elearning.rzbd.haw-hamburg.de>

Elkeles, Thomas/Kirschner, Wolf (2004): Arbeitslosigkeit und Gesundheit. Intervention durch Gesundheitsförderung und Gesundheitsmanagement – Befunde und Strategien. Band 3 der Reihe Gesundheitsförderung und Selbsthilfe (Hg: Bundesverband der Betriebskrankenkassen). Wirtschaftsverlag NW, Verlag für neue Wissenschaft: Bremerhaven.

Elkeles, Thomas, Kirschner, Wolf (2005): Unemployment as a determinant of health. In: Burazeri, Genc/Georgieva, Lidia (Eds.): Determinants of health. In press.

German Center for Evidence-based Nursing: <http://www.medizin.uni-halle.de/pflegewissenschaft/index.php?id=347> (04.05.2005)

Geyer, Siegfried (2000): Antonovskys Sense of Coherence - ein gut geprüftes und empirisch bestätigtes Konzept? In: Wydler Hans/Kolip, Petra/Abel, Thomas (Hg.): Salutogenese und Kohärenzgefühl. Juventa: Weinheim, S. 71-84.

Gross, Dorothea (2004): Evidence Based Nursing - der umfassende Begriff. Pflege 17: 196 – 207.

Hänselmann, Kathrin/Schreiter, Itke (2004): Gesundheitsförderung bis zum Schulalter in Friedrichshain-Kreuzberg. Bachelorarbeit zur Erlangung des akademischen Grades Bachelor of public health and administration. Fachbereich Gesundheit und Pflege, Fachhochschule Neubrandenburg.

Hellstern, Gerd-Michael/Wollmann, Hellmut (Hg.) (1984): Handbuch zur Evaluierungsforschung. Bd. 1. Westdeutscher Verlag: Opladen.

Holling Heinz (1999): Evaluation eines Stressbewältigungstrainings. In: Holling, Heinz/Gediga, G. (Hrsg.). Evaluationsforschung. Hogrefe: Göttingen, S. 35-57.

Hügli, Anton/Lübcke, Poul (Hg.) 1991): Philosophielexikon. Personen und Begriffe der abendländischen Philosophie von der Antike bis zur Gegenwart. Rowohlt: Reinbek bei Hamburg.

Kilbom, A. (1988): Intervention programs for work-related neck and upper limb disorders: strategies and evaluation. Ergonomics 31: 387-747.

Kirschner, Renate/Elkeles, Thomas/Kirschner, Wolf (2003): Evaluation der Tätigkeit des Fonds Gesundes Österreich 1998-2001. Ergebnisbericht. Zusammenfassung, Schlußfolgerungen und Empfehlungen. Fonds Gesundes Österreich, Eigenverlag: Wien.

Kirschner, Wolf (2004): Arbeitsmarktintegration und Gesundheitsförderung. Die Bedeutung von Gesundheitskompetenzmodulen. State of the art, Evidenz, offene Fragen. Vortrag auf der 1. Tagung des Forums Arbeitsmarktintegration und Gesundheitsförderung am 18.11.2004, Bundesverband der Betriebskrankenkassen, Essen..

Kirschner, Wolf/Elkeles, Thomas/Kirschner, Renate (2004a): Fonds/Stiftung Prävention und Gesundheitsförderung – Plädoyer für eine evidenzbasierte Weiterentwicklung präventiver und gesundheitsfördernder Interventionen auf der Grundlage der Erfahrungen in Deutschland, Österreich und der Schweiz. Gesundheits- und Sozialpolitik 58, 3-4: 37 – 45.

Kirschner, Wolf/Elkeles, Thomas/Kirschner, Renate (2004b): Evaluation der Tätigkeit des Fonds Gesundes Österreich (FGÖ) - Was können wir für Deutschland lernen? Die Krankenversicherung 56, März: 65 – 68.

- Kirschner, Wolf/Elkeles, Thomas/Kirschner, Renate (2004c): GKV-Stiftung für Prävention und Gesundheitsförderung – Vadis et quo vadis? Impulse, H. 43: 21.
- Kliemke, Christa/Daubitz, Stefan/Grenner, Gerd (2005): Modellprojekt Stadtteilnetz „Soziale Ökonomie für Kinder, Umwelt und Gesundheit“. Evaluationsbericht der Projekteffekte auf die gesundheits- und wohnumfeldbezogene Lebensqualität von Kindern, Jugendlichen und Familien. Gefördert im Rahmen des Aktionsprogramms Umwelt und Gesundheit des Bundesministerium für Gesundheit und Soziale Sicherung, Förderkennzeichen: AZ: 122-1720/71.
- Kreis, Julia/Bödeker, Wolfgang (2003): Gesundheitlicher und ökonomischer Nutzen betrieblicher Gesundheitsförderung und Prävention. IGA-Report 3. Bundesverband der Betriebskrankenkassen, Hauptverband der gewerblichen Berufsgenossenschaften, Berufsgenossenschaftliches Institut Arbeit und Gesundheit. Essen und Dresden.
- Kromrey, Helmut (2001): Evaluation - ein vielschichtiges Konzept. Begriff und Methodik von Evaluierung und Evaluationsforschung. Empfehlungen für die Praxis. Sozialwissenschaften und Berufspraxis 24, 2: 105 – 132.
- Kromrey, Helmut (2004): Evaluation, Evidenzbasierung, Qualitätssicherung – Worüber reden wir? In: Luber, Eva/Geene, Raimund (Hg.): Qualitätssicherung und Evidenzbasierung in der Gesundheitsförderung. Mabuse-Verlag: Frankfurt am Main, S. 75-85.
- Kuhnert, Peter (1999): Bewältigungskompetenzen und Beratung von Langzeitarbeitslosen. Dissertation. Universität Dortmund.
- Lenhardt, Uwe (1999): Betriebliche Gesundheitsförderung durch Krankenkassen. Rahmenbedingungen, Angebotsstrategien, Umsetzung. Bd. 18 der Reihe Forschung aus der Hans-Böckler-Stiftung. Edition sigma: Berlin.
- Lenhardt, Uwe (2003): Bewertung der Wirksamkeit betrieblicher Gesundheitsförderung. Zeitschrift für Gesundheitswissenschaften 11, 1: 18 – 37.
- Lenhardt, Uwe/Elkeles, Thomas/Rosenbrock, Rolf (1997): Betriebsproblem Rückenschmerz. Eine gesundheitswissenschaftliche Bestandsaufnahme zu Verursachung, Verbreitung und Verhütung. Juventa: Weinheim und München.
- Müller-Kohlenberg, Hildegard/Beywl, Wolfgang (2003): Standards der Selbstevaluation. Zeitschrift für Evaluation 2, 1: 79-93.
- Noack, R. Horst (2002): Evidenzbasierte Gesundheitsförderung als Strategie einer QM-gestützten kommunalen Gesundheitspolitik. Vortrag auf dem Akademie-Forum, Akademie für öffentliches Gesundheitswesen, Düsseldorf, 20.November 2002.
- Osterholz, Uwe (1993): Kritische Bewertung der Wirksamkeit verschiedener Maßnahmen zur Lösung des Problems 'Rückenschmerz'. Veröffentlichungsreihe der Forschungsgruppe Gesundheitsrisiken und Präventionspolitik, Wissenschaftszentrum Berlin für Sozialforschung, Nr. P93-204. Berlin.
- Papies-Winkler, Ingrid (2003): Kiezdetektive – Kinderbeteiligung für eine gesunde und zukunftsfähige Stadt. Projektbericht, Deutsches Institut für Urbanistik, <http://www.sozialestadt.de/praxisdatenbank/suche/ausgabe.php?id=85> (08.03.2005).
- Pröll, Ulrich (2004): Arbeitsmarkt und Gesundheit. Gesundheitliche Implikationen der neuen Arbeitsmarktpolitik und Ansätze zur Prävention. Veröffentlichungsreihe der Bundesanstalt für Arbeitsschutz und Arbeitsmedizin, Nr. Fb 1018. Wirtschaftsverlag NW: Dortmund/Berlin/Dresden.
- Proudfoot, Judith/Guest, David/Carson, Jerome/Dunn, Graham/Gray Jeffrey (1997): Effect of



cognitive behavioural training on job finding among long-term unemployed people. *The Lancet* 350: 96-100.

Rosenbrock, Rolf (2004): Evidenzbasierung und Qualitätssicherung in der primären Prävention. *Zeitschrift für Evaluation* 3, 1: 71-80.

Rossi, Peter H./Freeman, Howard E./Hofmann, Gerhard (1988): *Programm - Evaluation. Einführung in die Methoden angewandter Sozialforschung*. Enke: Stuttgart.

Schneider, S./Hauf, C./Schiltenswolf, M. (2004): Ineffektive Rückenschmerzprävention wegen mangelhafter Zielgruppenerreichung – Eine bundesweite Repräsentativstudie zu Nutzerstruktur und Teilnahmefaktoren von Rückenschulen. *Der Schmerz (Suppl. 1)*, S94.

Silverstein, Barbara (1992): Design and evaluation of interventions to reduce work-related musculoskeletal disorders. In: Hagberg, M./Kilbom, A. (Eds.): *International Scientific Conference on Prevention of Work-related Musculoskeletal Disorders (PREMUS)*, Sweden, May 12-14, 1992. *Book of Abstracts*. Arbetsmiljöinstitutet: Solna, pp. 1-7.

Simon, Wenke (2003): *Arbeitslosigkeit und psychische Gesundheit. Eine Recherche nach evaluierten Gesundheitsförderungsprojekten für Arbeitslose*. Bachelorarbeit zur Erlangung des akademischen Grades Bachelor of public health and administration. Fachbereich Gesundheit und Pflege, Fachhochschule Neubrandenburg.

Simon, Wenke (2004): Unemployment and Mental Health-Interventionstrategies. In: Geene, Raimund, Halkow, Anja (Hg.): *Armut und Gesundheit – Strategien der Gesundheitsförderung*. Mabuse – Verlag: Frankfurt, S. 213-219.

Stockmann, Reinhard (1996): *Die Wirksamkeit der Entwicklungshilfe. Eine Evaluation der Nachhaltigkeit von Programmen und Projekten der Berufsbildung*. Westdeutscher Verlag: Opladen.

Stockmann, Reinhard (2002): Qualitätsmanagement und Evaluation - Konkurrierende oder sich ergänzende Konzepte? *Zeitschrift für Evaluation* 1, 2: 209-244.

Stockmann, Reinhard (2004): Evaluation in Deutschland. In: Stockmann, Reinhard (Hg.): (2004): *Evaluationsforschung. Grundlagen und ausgewählte Forschungsfelder*. 2. Aufl., Leske + Budrich: Opladen, S. 13 - 43.

Trojan, Alf (2000): Evaluation komplexer Ansätze: das Healthy City Projekt. *Public Health Forum*, H. 28: 13.

Trojan, Alf (2001): Qualitätsentwicklung in der Gesundheitsförderung. In: Dierks, Marie-Luise/Walter, Ulla/Windel, Irmtraut/Schwartz, Friedrich-Wilhelm: *Qualitätsmanagement in Gesundheitsförderung und Prävention. Grundsätze, Methoden und Anforderungen*. Bd. 15 der Reihe *Forschung und Praxis der Gesundheitsförderung*. Bundeszentrale für gesundheitliche Aufklärung. Köln, S. 51-72.

Trojan, Alf (2004): Prävention und Gesundheitsförderung in Städten und Gemeinden. In: Hurrelmann, Klaus/Klotz, Theodor/Haisch, Jochen (Hg.): *Lehrbuch Prävention und Gesundheitsförderung*. Verlag Hans Huber: Bern, Göttingen, Toronto, Seattle, S. 305-315.

Volkholz, Volker (1991): *HdA-Bilanzierung*. Gesellschaft für Arbeitsschutz- und Humanisierungsforschung. Eigenverlag: Dortmund.

Volkholz, Volker (1993): Für eine phantasieanstiftende Sozialwissenschaft. Erfahrungen aus den Forschungsprogrammen „Humanisierung des Arbeitslebens“ und „Arbeit und Technik“. *Arbeit* 2, 3: 264-278.

Vuori, Jukka/Silvonen, Jussi/Vinokur, Amiram D./ Price, Richard H. (2002): The Työhön Job Search Programm in Finland: Benefits for the Unemployed With Risk of Depression or Discouragement. *Journal of Occupational Health Psychology* 1, 2: 5-19.

Wickström, Gustav/Hyytiäinen, K./Laine, M. et al. (1993): A Five Year Intervention Study to reduce Low Back Disorders in the Metal Industry. *International Journal of Industrial Ergonomics* 12: 25-33.

Wottawa, Heinrich/Thierau, Heike (1998): *Lehrbuch Evaluation*. 2. Aufl. Verlag Hans Huber: Bern, Göttingen, Toronto, Seattle.

Wright, Michael T. (2004): Partizipative Qualitätssicherung und Evaluation für Präventionsangebote in Settings. In: Rosenbrock, Rolf/Bellwinkel, Michael/Schröer, Alfons (Hg.): *Primärprävention im Kontext sozialer Ungleichheit. Wissenschaftliche Gutachten zum BKK-Programm „Mehr Gesundheit für alle“*. Bd. 8 der Reihe Gesundheitsförderung und Selbsthilfe (Hg: Bundesverband der Betriebskrankenkassen). Wirtschaftsverlag NW, Verlag für neue Wissenschaft GmbH: Bremerhaven, S. 297–346.

Struening, Elmar L. (Ed.) (1975): *Handbook of evaluation research*. 2. Print. Sage Publ.: Beverly Hills.