

International erfolgreiche Interventionen der Prävention und ihre Übertragbarkeit auf Deutschland

Ein Projekt des Institut für Gesundheitsökonomie und Klinische Epidemiologie (IG-KE) der Universität zu Köln, gefördert von der Hans Böckler Stiftung (Projekt-Nr. S-2006-832-4). Mai 2006- Juli 2007

1 Zusammenfassung

In Deutschland konzentriert sich die Diskussion in der Prävention und Gesundheitsförderung derzeit wesentlich auf die Mittelherkunft, also darum, welche Beteiligten welche Summen aufbringen müssen. Kaum eine Diskussion gibt es um die entscheidendere Frage, für welche Projekte die Mittel ausgegeben werden sollten. Diese Frage hat zwei Facetten:

- 1.) Welche Krankheitsbilder oder Zustände sollen verhindert oder in ihrem Verlauf abgeschwächt oder verzögert werden?
- 2.) Welche Programme sind überhaupt wirksam?

Die erste Frage muss politisch beantwortet werden, beispielsweise durch Setzen von nationalen Präventionszielen auf der Grundlage von betroffenen Gruppen, Häufigkeiten und auch erwartbaren Folgekosten einer Erkrankung.

Die zweite Frage soll durch das vorliegende Gutachten beantwortet werden. Grundlegender Gedanke war, dass die Erfahrungen des Auslands systematisch analysiert werden, um die Kernpunkte wirksamer Programme zu erkennen. Diese Idee entstammt der Evidenz-basierten Medizin. Auch dort werden klinische Studien systematisch gesichtet und aufbereitet, um schließlich Vorschläge (Leitlinien) an den Arzt geben zu können, welche Therapie Wirksamkeit verspricht und welche nicht.

Die Übertragung der Methode der Evidenz-basierten Medizin auf die Prävention ist nicht unumstritten. So ist es schwierig, in einem sehr heterogenen Umfeld Programme zu übertragen, etwa von einem Stadtteil auf einen anderen. Ebenso decken Präventionsprogramme eine Vielzahl von Typen ab, vom Aufhängen von Plakaten bis zum Coaching einzelner Versicherter durch einen Arzt. Die Vergleichbarkeit beziehungsweise Identifizierung der für den Erfolg oder Misserfolg ausschlaggebenden Kernpunkte ist dann schwierig. Drittens schließlich ist der Zusammenhang zwischen Programm und späterem Auftreten einer Erkrankung beziehungsweise Zustandes generell schwierig, da die betrachteten Zeiträume sehr lang sind und die betroffene Fallzahl (beispielsweise Neuerkrankungen) sehr klein.

Ein weiteres Problem in der Bewertung von Präventionsprogrammen besteht auch darin, dass man auf das Einhalten einer so genannten guten Strukturqualität vertraut, und damit eine gewünschte Wirkung des Programms garantiert wird. Die Effekte selbst werden dann oftmals gar nicht mehr gemessen beziehungsweise geprüft. Solche Strukturmerkmale sind beispielsweise die Beteiligung der Betroffenen an der Entwicklung des Programmes, die Ausbildung von Multiplikatoren und Kriterien der Nachhaltigkeit. Sofern diese Qualitätskriterien gesichert sind, sollte auch das Programm durchgeführt werden. Unser Ansatz hingegen berücksichtigt gleichermaßen die Effektivität, das heißt: Programme, die lediglich gute Strukturqualität bieten, ohne die Effektivität nachweisen zu können, sind letztendlich nicht zwangsläufig empfehlenswert.

Das Gutachten untersuchte systematisch Studien in den Bereichen Bewegung, Ernährung, Rauchen und Depression. Es konnten keine generellen Aussagen zu Merkmalen getroffen werden, welche die Wahrscheinlichkeit erhöhen würden, dass ein Programm auch wirkungsvoll ist. Wesentlich dafür war mehrere Gründe:

- Die gefundenen Programme waren sehr heterogen, so dass abschließend nicht festgestellt werden konnte, welche Faktoren der Studien Wirksamkeit (oder Nicht-Wirksamkeit) verursachten.
- Viele Programme wurden nicht oder nicht ausreichend evaluiert. Das wissenschaftliche Niveau der Evaluation war oftmals so schwach, dass die Ergebnisse des Programms nicht zweifelsfrei auf das Programm zurückgeführt werden konnten.

Einige Erkenntnisse zeigen aber zumindest einen Trend. So sind zur Förderung der **Bewegung** Plakate mit der Aufforderung zum Gebrauch der Treppe statt des Aufzugs in Betrieben nicht effektiv. Bei Jugendlichen kann eine Zunahme der Bewegung nur mit kombinierten Angeboten, welche Schule und/oder Eltern einbeziehen, erreicht werden, nicht mit Programmen, die rein edukativ ansetzen. Die Änderung des **Ernährungsverhaltens** konnte von isolierten Programmen kaum nachgewiesen werden, lediglich die Kombination mit verstärkter Bewegung unter Einbeziehung der Familien/Eltern zeigt vermehrt Effektivität. Die Änderung des **Rauchverhaltens** bei Schwange-

ren wird am ehesten dann erreicht, wenn als Autoritäts- oder Fachpersonen wahrgenommene Gruppen im Programm beteiligt werden, etwa Ärzte oder Hebammen. Auch im Bereich Depression scheint sich zu zeigen, dass rein edukative Ansätze kaum Wirkung zeigen, jedoch Programme mit verhaltenstherapeutischen Ansätzen eher effektiv sind.

Als Schlussfolgerung aus diesen Erkenntnissen kann gezogen werden:

- Generell muss zwingend eine Evaluation von Programmen durchgeführt werden. Ansonsten können effektive Programme und nicht-effektive Programme nicht unterschieden werden.
- Die Methoden zur Evaluation von Präventionsprogrammen müssen weiter entwickelt werden. Dazu gehören Definitionen von Zielgrößen, Randomisierung, Kontrollgruppen, Nachbeobachtungszeiträume und die Benennung von Störgrößen.
- Die überprüfbaren Ziele der Programme müssen im Vorfeld festgelegt werden.
- Da der Erfolg der Programme derzeit unsicher ist, sollten alle Programme zeitlich befristet werden.
- Da die erfolgversprechenden Strukturen von Programmen derzeit nicht bekannt sind, sollte möglichst ein Wettbewerb um beste Effektivität und Kosteneffektivität entstehen. Dies bedeutet, dass eine Begrenzung auf Programme, die lediglich Strukturqualität aufweisen, nicht angebracht ist.

2 Hintergrund und Fragestellungen

Das Krankheitsgeschehen in industrialisierten Ländern ist durch eine Zunahme chronischer Krankheiten geprägt, die zum großen Teil lebensstilbedingt sind. Über die Notwendigkeit einer Stärkung der Prävention besteht in Deutschland vor diesem Hintergrund Einigkeit. Obwohl die Knappheit der verfügbaren Ressourcen eine Priorisierung präventiver Maßnahmen erfordert, bleibt das Wie der Priorisierung gegenwärtig unbeantwortet.

Deutschland wurde durch die mit dem Beitragsentlastungsgesetz vom 01.11.1996 erfolgte zwischenzeitliche Herausnahme primärpräventiver Leistungen aus dem SGB V beim Aufbau präventiver Strukturen zurückgeworfen. Erst mit der erneuten Aufnahme von Satzungsleistungen zur Primärprävention in § 20 SGB V durch das Gesundheitsreformgesetz 2000 und der Ankündigung eines Präventionsgesetzes wurde der Prävention wieder vermehrte Aufmerksamkeit gewidmet. Es ist zu erwarten, dass auch ein neuer Gesetzesentwurf keinen abschließenden Katalog von Präventionsmaßnahmen vorsehen wird, sondern vielmehr Anforderungen festlegt, welche die von den Präventionsträgern zu erbringenden Präventionsleistungen zu erfüllen haben. Vor diesem Hintergrund hatte unser Projekt folgende Fragestellungen:

- Können im Ausland durchgeführte und als effektiv bewertete Projekte der Prävention und Gesundheitsförderung auf Deutschland übertragen werden?
- Für welche Projekte der Prävention und der Gesundheitsförderung konnte in den ausgewählten Ländern ein Effekt nachgewiesen werden?
- Welche Erfahrungen liegen im Ausland mit der Messung der Effektivität und Qualität von Präventionsmaßnahmen vor?
- Lassen sich diese Präventionsmaßnahmen nach Kriterien untersuchen und kategorisieren, um Rückschlüsse auf mögliche Gründe für den Erfolg spezifischer Maßnahmen zu formulieren?

3 Können Empfehlungen für Präventionsprogramme ausgesprochen werden?

Generelle Erfolgsfaktoren für Präventionsprojekte können aus der Literatur nicht abgeleitet werden. Jedoch können einzelne, herausragende Ergebnisse ausgewählter Studien in den vier Handlungsfeldern

- Bewegung
- Depression
- Ernährung
- Rauchen

dargestellt werden:

3.1 Bewegung

Im Handlungsfeld Frauen und Mädchen und Bewegung existieren zwei effektive bzw. teilweise effektive Studien, die in den Bereich stark empfehlenswert gruppiert wurden. Sie waren multistrategisch ausgerichtet, schlossen eine vorherige Befragung der Zielgruppe ein und integrierten möglichst alle Beteiligten wie Schule, Gemeinde etc. (Wen et al. 2002, Pate et al. 2005). Unter den Studien im Bereich eingeschränkter Empfehlung zeigte eine Studie mit einem sehr engen verhaltenspräventiven Ansatz keinen Effekt (Robbins et al 2006). Zwei weitere Studien mit einem gemischten nachweisbaren Effekt haben ebenfalls einen sehr engen Ansatz und sind von der Anlage her eher dem Feld efficacy (Schofield et al. 2005) bzw. einem Pilotcharakter (Williams et al. 2005) zuzuordnen.

Für den Bereich Betrieb und Bewegung zeigt sich, dass alle Aktionen, die mit Plakaten etc. das Treppensteigen isoliert steigern wollen, nicht mehr durchgeführt werden sollten, da sie scheitern (Adams and White 2002, Marshall et al. 2002). Nur zwei der acht Studien zum Treppensteigen im betrieblichen Setting erreichten in unserer Evaluation

das nötige Qualitätsniveau. Die einzige Studie, die nach Qualität in den Bereich empfehlenswert kam, und einen positiven Effekt hat, entspricht zwar eher einem von oben verordneten Programm (Harrell et al. 1996). Sie weist damit aber in der Tendenz auf, dass Bewegung im Betrieb, hier zwar das spezielle Klientel von Polizeiangehörigen, „einfach“ durchgeführt werden muss, so dass man im Betrieb alle Mitarbeiter erreicht. Weiterhin zeigten die weiteren gescheiterten Studien Faktoren, die man als Misserfolgskriterien beurteilen könnte: a) Arbeitsbefreiung wird nicht gewährt (Gerdle et al. 1992). b) Die Intervention ist isoliert auf Beratung oder Fitnessförderung als Anreiz ausgelegt (Aittasalo et al. 2004).

3.2 Depression

Die zur Prävention der Depression analysierten Studien konzentrieren sich auf wenige Programme. Am häufigsten wurden die Programme Leichtigkeit im sozialen Alltag (LISA), das Programm FRIENDS sowie Resourceful Adolescent Programm (RAP) evaluiert, die auch alle unter den Studien mit überdurchschnittlichen Qualitätsscores vertreten sind. Für alle drei Programme ist mindestens eine der Studien mit überdurchschnittlichem Qualitätsscore teilweise effektiv. Besonders hervorzuheben sind die beiden mit überdurchschnittlichen Qualitätsscores durchgeführten Studien zu dem Programm RAP, die beide als teilweise effektiv bewertet worden sind. Weitere Studien für dieses Programm mit unterdurchschnittlicher Studienqualität liegen nicht vor (Merry et al. 2004; Shochet et al. 2001). Die Studie von Merry et al. 2004 zu dem Programm RAP erreicht als einzige Studie den Qualitätslevel stark empfehlenswert.

Für das Programm LISA (Pössel et al. 2004; Groen et al. 2003; Pössel et al. 2006a) und für das Programm FRIENDS (Lowry-Webster et al. 2003) liegen auch Studien mit unterdurchschnittlicher Studienqualität vor. Die Studien für das Programm FRIENDS sind aber alle zumindest teilweise effektiv und weisen damit eine einheitliche Tendenz auf. Für das Programm LISA liegen dagegen sowohl ineffektive als auch teilweise effektive Studien vor.

Es gibt Anhaltspunkte, dass die Ausgestaltung der Interventionen über die in den Qualitätsscores erfassten Aspekte hinaus großen Einfluss auf die Ergebnisse haben (Pössel et al. 2006 b). Nach unserer Auffassung ist im Bereich der Prävention von Depressionen und Angststörungen noch weiterer Forschungsbedarf gegeben. Trotz der positiven Tendenzen der kognitiv-verhaltenstherapeutischen Ansätze sind gegenwärtig keine klaren Empfehlungen möglich. Sie scheinen jedoch dem rein edukativen Ansatz überlegen, für den die zwei analysierten Studien – weder die mit unterdurchschnittlichen noch die mit überdurchschnittlichen Qualitätsscores – keine Wirksamkeit zeigen konnten. (Clark et al. 1993) Eine flächendeckende Umsetzung kognitiv verhaltenstherapeutischer Interventionen zur universalen Prävention der Depression erscheint gegenwärtig aufgrund der offenen Fragen zur optimalen Ausgestaltung verfrüht.

3.3 Ernährung

Für die 24 evaluierten Studien über alle Qualitätsstufen hinweg ergab sich, dass bei den acht Interventionen, die kombiniert Ernährung und Bewegung in den Blick nahmen, nur der Ernährungsbereich einen positiven Effekt hatte, der Bewegungsbereich hingegen einen negativen. Alle drei Interventionen, die in den Bereich empfehlenswert fielen, hatten entweder einen Ansatz, der sowohl Ernährung und Bewegung einbezog (Gortmaker et al. 1999a) oder einen Ansatz, in dem die Eltern eng eingebunden waren (Reynolds et al. 2000, Townsend et al. 2006).

3.4 Rauchen

Die internationale Studienlage zur Prävention des Rauchens bei Schwangeren erlaubt keine eindeutigen Empfehlungen. In der Tendenz sind effektive Interventionen mit hoher Qualität in der internationalen Studienlage multimodal angelegt, jedoch sind auch einige multimodale Programme gescheitert. Erfolgreiche Programme wurden eher durch Hebammen, Ärzte oder Pflegepersonal während der routinemäßigen Schwangerschaftsuntersuchungen durchgeführt.

3.5 Empfehlungen

Unsere Studie hat gezeigt, dass die Situation zum Nachweis von Effektivität in Prävention und Gesundheitsförderung auch im internationalen Umfeld in etwa den Stand der kurativen Medizin von vor 25 Jahren aufweist. Belastbare und ausreichend erprobte Kataloge zur Bewertung und anschließenden Priorisierung von Präventionsangeboten fehlen weitgehend. Vorhandene Kataloge haben eine spezifische Zielrichtung (etwa Beratung der Projektdurchführenden).

In Deutschland wurden in den letzten Jahren Gesundheitsziele erarbeitet.¹ (GVG 2003) Aufgrund des erwartbaren und notwendigen Ressourcenvolumens für Prävention und Gesundheitsförderung ist nun zu entscheiden, welchen Beitrag Prävention und Gesundheitsförderung dazu leisten können. Nach unserer Auffassung müssen die Erfahrungen der Evidenz-basierten Medizin aufgegriffen werden, um rationale Entscheidungen zu treffen.

Um die Rationalität aufbauen zu können muss der Gesetzgeber eine verbindliche Evaluation der geförderten Interventionen, gemessen an den Zielen, vorsehen. Da die derzeit nicht ausreichende empirische Studienlage dazu führt, dass absehbar auch nicht

¹ Die sechs Gesundheitsziele auf Bundesebene sind: „Diabetes mellitus Typ2: Erkrankungsrisiko senken, Erkrankte früh erkennen und behandeln“, „Brustkrebs: Mortalität vermindern, Lebensqualität erhöhen“, „Tabakkonsum reduzieren“, „Gesund aufwachsen“, „Patient(inn)ensouveränität stärken“, „Depressive Erkrankungen: verhindern, früh erkennen, nachhaltig behandeln“.

wirksame Interventionen durchgeführt werden, sollten darüber hinaus alle Interventionen zunächst zeitlich befristet werden.

Auf Grund der Heterogenität der Erhebungsmethoden und der Effektparameter sollte der Gesetzgeber dafür Sorge tragen, dass bundesweit einheitliche Evaluationsstandards erarbeitet werden. Nach dem Ablauf einer genügend großen Zahl von Interventionen und nachfolgender Evaluation können dann Empfehlungen ausgearbeitet werden, welche Interventionstypen Erfolg versprechend sind. Der Gesetzgeber sollte daher auch vorsehen, dass in einem angemessenen Zeitraum eine umfassende Zusammenführung der durchgeführten Interventionen beziehungsweise ihrer Evaluationen erfolgt.

Die bisherigen Entwürfe zum Präventionsgesetz schrieben für Maßnahmen der Verhaltensprävention Effektivitätsnachweise, für Maßnahmen der Verhältnisprävention ein angemessenes Qualitätsmanagement vor. Es hat sich gezeigt, dass international zwischen Qualität des Studiendesigns, der Intervention und dem Effekt keine Korrelation besteht. Dies ist wissenschaftstheoretisch zunächst nicht erstaunlich. Es führt aber in der Konsequenz dazu, dass eine Beschränkung auf eine reine Einhaltung vorgegebener Qualitätskriterien, wie es der Gesetzgeber im Präventionsgesetz für Interventionen der Verhältnisprävention vorgesehen hatte, nicht ausreichend ist. Mit anderen Worten reicht es beispielsweise nicht aus, eine Partizipation der Betroffenen vorzusehen oder eine Schulung von Multiplikatoren. Trotz der Einhaltung dieser Qualitätskriterien kann eine Intervention gleichermaßen am Ende einen Effekt aufweisen oder wirkungslos bleiben.

4 Methode des Projekts

Der empirische Teil des Projektes gliedert sich in die Teile systematische Literaturrecherche sowie Bewertung der gefundenen Studien und Entwicklung eines Kriterienkatalogs zur Evaluation von Interventionen.

Für die Bewertung der Studien bzw. Interventionen wurde ein Kriterienkatalog entwickelt, der auf einer kritischen Durchsicht und Zusammenschau der derzeit international gebräuchlichen Qualitätsinstrumente beruht. Dazu wurden unter anderen die Instrumente PREFFI (PREvention EFFect-management-Instrument), Closing the Gap, IDM (Interactive Domain Model), QSP (Qualitätssicherung Prävention), Quintessenz etc. herangezogen. Die Erstellung des Kriterienkatalogs erfolgte in einem begründeten Verfahren nach dokumentierten Kriterien. Anschließend wurden Korridore für Empfehlungsstärken festgelegt, wie sie ähnlich in der Evidenz-basierten Medizin und auch der Public-Health-Forschung in Konsensusprozessen entwickelt werden.

Der Kriterienkatalog bewertet Programme hinsichtlich ihrer Qualität und Effektivität. Er besteht insgesamt aus 68 Fragen. Die im systematischen Suchverfahren gefundenen Studien wurden im Gutachten hinsichtlich Qualität des Studiendesigns und Qualität

der Intervention evaluiert. Unter **Qualität des Studiendesings** werden Gütekriterien aus der evidenzbasierten Medizin herangezogen wie statistische Aussagekraft, Vergleichbarkeit von Interventions- und Kontrollgruppe etc. Diese Gütekriterien würden für eine Implementierung wegfallen, wenn für eine Intervention nach einer Studienphase die Wirksamkeit belegt worden ist. Unter **Qualität der Intervention** werden Gütekriterien der Best practice sowie des Qualitätsmanagements zusammen geführt, die die inhaltliche Gestaltung der Intervention bzw. eines Programms bewerten können. Für jedes Qualitätskriterium wurden nach einem festgelegten Verfahren zwischen null und einem Punkt vergeben. Die maximale Punktzahl im Score war abhängig vom Studiendesign. Die unterschiedlich hohen erreichbaren Punktzahlen ergeben sich, da einige Kriterien nicht anwendbar waren, etwa die Güte der Randomisierung.

Gegenstand des Projektes waren Interventionen der Prävention und Gesundheitsförderung für die Risikofaktoren Bewegung, Depression, Ernährung und Rauchen. In die Untersuchung wurden präventive und gesundheitsfördernde Interventionen aus den USA, Kanada, Australien, Neuseeland, Großbritannien, Frankreich, Niederlanden, den skandinavischen Ländern, Österreich und der Schweiz einbezogen. Die Recherche wurde auf Publikationen der Jahre 1990 bis 2006 beschränkt. Recherchiert wurde in den Literaturdatenbanken Medline, Amed, Biosis, Embase, Somed, Cinahl, Gopal Health und Psychinfo. Eingeschlossen wurden Publikationen, die nicht nur eine Intervention, sondern auch deren Evaluation enthielten. Ausgeschlossen wurden Studien, welche ausschließlich auf die Multiplikatorenebene, also z. B. Ärzte/innen ausgerichtet waren. D.h., die Maßnahme sollte in ihrer Wirkung auf die Betroffenen evaluiert sein, deren Gesundheit durch die präventiven Maßnahmen beeinflusst werden sollte. Ein weiteres Ausschlusskriterium war, dass die Evaluation nicht nur Veränderungen im Wissen oder in den Einstellungen gemessen hatte, sondern auch Veränderungen im Verhalten oder klinischen beziehungsweise physischen Parametern.

Die Zahl der in der Literaturrecherche erzielten Treffer zu den verschiedenen Handlungsfeldern lag zwischen ca. 12.400 (Rauchen) und ca. 5.000 Studien (Depressionen). Nach Berücksichtigung der Ein- und Ausschlusskriterien verblieben immer noch rund 1.500 bzw. 500 Studien für die genannten Handlungsfelder.

Auf Grund dieser großen Menge an Studien zu Interventionen innerhalb der Felder Bewegung, Depression, Ernährung und Rauchen mussten Fokusbereiche ausgewählt werden. Es wurden daher im Bereich Bewegung das Setting Betrieb ausgewählt (26 Interventionen) sowie Intervention bei Mädchen und Frauen (14 Interventionen). Im Handlungsfeld Depression wurden 18 Studien der universellen Prävention der Depression in Schulen bewertet. Desweiteren wurden 24 Studien zur Ernährung in der Zielgruppe Kinder und Jugendliche im Setting Schule sowie 34 Interventionen zur Raucherentwöhnung bei Schwangeren analysiert.

Die Verbindung der Ergebnisse der Qualitätsevaluation mit den Ergebnissen der erzielten Effekte der Intervention machte es möglich, Empfehlungen auszusprechen.

Dabei wurden Interventionen, die mehr als 80% der Punkte in den beiden Dimensionen Qualität des Studiendesigns und Qualität der Intervention erzielten, als „in hohem Maße empfehlenswert“ eingestuft. Das heißt wenn die Intervention einen Effekt hatte, liegt eine hohe Wahrscheinlichkeit vor, dass dieser so wieder eintritt und umgekehrt, wenn diese Intervention keinen Effekt hat, liegt eine hohe Wahrscheinlichkeit vor, dass er bei erneuter Durchführung ausbleibt. Interventionen deren Qualität einen Wert von 60% erreichte, wurden als „stark empfehlenswert“ bezeichnet. Interventionen die noch über dem Durchschnittswert von der Qualität des Studiendesigns und der Qualität der Intervention lagen, konnten nur „eingeschränkt empfohlen“ werden. Alle darunterliegenden Interventionen wurden „nicht weiter in die Empfehlung“ aufgenommen. Bei dieser Einstufung ist zu berücksichtigen, dass für eine Empfehlung auch ein positiver Effekt nachgewiesen sein muss. Bei einer nicht-effektiven Intervention, die mit hohen Qualitätsscores durchgeführt wurde, ist die Wahrscheinlichkeit hoch, dass sie in eine Negativliste aufgenommen werden muss. Nicht-effektive Studien mit geringeren Qualitätsscores können dagegen bei einer Wiederholung der Studie mit besserer Qualität gegebenenfalls zu anderen (besseren) Ergebnissen führen. Daher bedeutet eine fehlende Empfehlung nicht, dass die Intervention generell zu verwerfen ist. Es bedeutet lediglich, dass die Qualität der Studiendurchführung keine Verlässlichkeit für eine Empfehlung bietet.