

## **Evaluation im Gesundheitswesen**

**Wolf Kirschner**

**7.Österreichische  
Gesundheitsförderungskonferenz  
Workshop 5**

**FB + E** GmbH Berlin  
Forschung, Beratung + Evaluation

**Berlin, Mai 2005**

## Einführung

### Evaluation in aller Munde?

Vor noch nicht einmal 30 Jahren war der Begriff der Evaluation - zumindest in Deutschland - zwar in den Sozialwissenschaften bekannt, in der konkreten Gesundheits- und Sozialpolitik und in Beschlüssen politischer Gremien spielte die Evaluation gesundheitlicher oder sozialer Maßnahmen aber noch kaum eine Rolle. Das hat sich gründlich geändert. Heute wird kein auszuschreibendes Interventions- oder Modellvorhaben mehr ohne die notwendige "begleitende Evaluation" vergeben und Mitte der 90er Jahre hatte z.B. das Abgeordnetenhaus von Berlin beschlossen, alle gesundheitlichen und sozialen Programme des Landes schrittweise einer Evaluation zu unterziehen.

Die Gründe für die aktuelle Konjunktur der Evaluationsforschung liegen auf der Hand. Bei zunehmend begrenzteren Ressourcen werden nicht nur allgemeinpolitisch, sondern auch innerhalb verschiedener Institutionen (wie z.B. Krankenkassen, Versicherungsanstalten, Krankenhäuser etc.) Fragen nach der "optimalen Allokation" der begrenzten Ressourcen und damit nach der Wirksamkeit und Wirtschaftlichkeit bestehender oder neu geplanter Maßnahmen und Programme gestellt. Rückblickend mutet es geradezu sonderbar an, dass nicht wenige gesundheitliche und soziale Programme lange Zeit weitgehend ohne jede, jedenfalls fundierte Evaluation durchgeführt worden sind, aber auch dafür gibt es verschiedene Gründe, die Sie noch kennen lernen werden.

Aus der neuen Konjunktur der Evaluationsforschung, die eher eine Konjunktur der Forderungen nach Evaluationen, denn eine Konjunktur qualitativ hochwertiger evaluativer Forschungspraxis ist, darf weder der Schluss gezogen werden, in der Evaluationsforschung sei methodisch Alles zum Besten bestellt, noch der Schluss, dass die Evaluationsforschung zu einer boomenden Branche in der Gesundheitsberatungslandschaft gehöre. Beides ist (leider?) nicht der Fall.

Evaluationen haben in der Regel ganz unterschiedliche Zielindikatoren, an denen die jeweilige Wirksamkeit und Wirtschaftlichkeit eines Programms oder einer Maßnahme überprüft werden soll. **Immer handelt es sich bei den Zielvariablen der Evaluationsforschung aber um Indikatoren der Wirksamkeit und Wirtschaftlichkeit.** Nun ist aber nicht jede Wissenschaftsrichtung und -methode, die sich mit Fragen der Qualität, Wirksamkeit und Wirtschaftlichkeit von Maßnahmen, Interventionen oder Prozessen befasst auch Evaluationsforschung. Deshalb soll i.f. zuerst eine Definition der sozialwissenschaftlichen Evaluation vorgenommen werden.

## 1.1 Definition der sozialwissenschaftlichen Evaluation

Unter Evaluation versteht man ganz allgemein die Bewertung der Wirksamkeit und Wirtschaftlichkeit von gesundheitlichen und sozialen Maßnahmen. Die Programmevaluation untersucht diese Kriterien - wie der Name sagt - im Rahmen gesundheitlicher oder sozialer Programme.

Häufig sind Dienstleistungen im Sozial- und Gesundheitswesen Maßnahmen von Programmen. Was ist ein Programm, was sind Programme?

Ein Programm ist der gezielte und effektive Einsatz von (wirksamen) Maßnahmen unter Verwendung definierter Ressourcen zur Erreichung bestimmter Ziele, die ein bestehendes und klar beschriebenes Problem vermindern oder beseitigen sollen.

Sie sollten die obige Definition eines Programms mindestens drei mal lesen und laut verbalisieren, um zu sehen, dass ein Programm:

- ein klar beschriebenes Problem voraussetzt
- das Ziel hat, dieses Problem zu verringern oder zu beseitigen
- wirksame Maßnahmen umfassen soll, eben dieses Ziel zu erreichen,
- wobei diese Maßnahmen selbst gezielt, wirksam und wirtschaftlich einzusetzen sind

Die Wirksamkeit der Maßnahmen von Programmen ist in der Theorie und Praxis der Interventions- und Evaluationsforschung das entscheidende und am meisten problembehaftete Kriterium, bedeutet es doch, dass zu einer Problemlösung mit Interventionen nicht irgendwelche zufälligen oder irgendwie geeignet erscheinenden Maßnahmen eingesetzt (implementiert) werden dürfen, vielmehr solche die, auf der Grundlage der Faktoren, die das Problem determinieren, eine große Wahrscheinlichkeit haben, das Problem auch zu vermindern oder zu beseitigen. D.h. idealtypisch nichts anderes, als dass wirksame Maßnahmen nur Faktoren sein können, die mit dem Problem (kausal) zusammenhängen. Die mit dem Problem (Zielvariable) am stärksten zusammenhängenden unabhängigen und potentiell kausalen Variablen sind also die primären "Schrauben", an denen "gedreht" werden kann, vorausgesetzt diese sind einer Intervention oder Veränderung überhaupt zugänglich. Bereits an dieser Stelle ist aber darauf hinzuweisen, dass der Nachweis von Kausalitäten in der empirischen Sozial und Gesundheitsforschung häufig mit erheblichen methodischen Problemen und auch Erkenntnisgrenzen zu tun hat.

Der ex-ante Evaluationsforscher, d.h. der Evaluationsforscher der bereits bei der Prüfung und Entwicklung einer Intervention beteiligt ist - was im übrigen recht selten der Fall ist - hat also die Aufgabe, im Sinne der Wirksamkeit optimale Problemlösungsstrategien zu entwickeln.

Um Ihnen die Schwierigkeiten dieser Aufgabe deutlicher zu machen, machen wir einen Ausflug in den winterlichen Harz.

### Ausflug in den winterlichen Harz

Mit zwei Freunden sind Sie an einem Freitag mit Ihrem 10 Jahre alten Auto in den Harz zum Skifahren gefahren. Sie sind spät von Berlin losgekommen, weil der Dozent ausgerechnet im Fach Epidemiologie mal wieder die Zeit richtig überzogen hat. Alle Strassen sind verschneit und vereist, es hat -20 Grad und die funktionierende Uhr im Auto zeigt 20.35 an. Ihr Ziel ist ein Hotel in Alexisbad. Kurz nach Quedlinburg stottert der Motor, um seinen Dienst schließlich ganz einzustellen. Sie haben ein Problem. Was könnten Sie tun?

Spielen Sie alle Möglichkeiten einer Problemlösung unter Annahme unterschiedlicher Voraussetzungen und Handlungsmöglichkeiten durch. Sie werden erkennen, dass die sicherste Methode der Problemlösung darin besteht, die Ursache(n) des Motorproblems herauszufinden und zu beheben. Dafür sind aber erhebliche Qualifikationen erforderlich, über die Sie vielleicht nicht verfügen. Tatsächlich verfügen die drei Studenten der Medizin nicht über diese Qualifikationen, sie haben auch kein Handy und weit und breit kommt kein Auto. Sie beschließen, zu beten, und siehe..... der Motor springt an.

Insgesamt ist die Wahrscheinlichkeit einer Problemlösung auf diese Weise aber sehr gering, wenn auch nicht gleich null. Das bedeutet nichts anderes, als dass man auch mit nicht rationalen Maßnahmen gelegentlich Effekte erzielen kann. Dies kann auch in der Programmdurchführung und Evaluationsforschung vorkommen.

Interventionen und Programme fallen nicht vom Himmel. Es gibt in der Regel kein Programm ohne ein Problem, bzw. genauer ohne ein Problem, über das unter den Beteiligten (z.B. Parteien bzw. Mehrheitsfraktion im Parlament, Geschäftsleitung und Betriebsrat im Unternehmen) insoweit Konsens besteht, dass zur Verringerung dieses Problems "etwas getan" werden muss.

Die Initiierung eines Programms ist also eine politische Maßnahme, die freilich auch von anderen Akteuren und auch der Wissenschaft befördert werden kann.

So mussten Mitte der 80er Jahre vor dem Hintergrund der zu beobachtenden raschen Ausbreitung der HIV-Infektionen auch in Deutschland relativ schnell entsprechende Maßnahmen der gesundheitlichen Aufklärung sowohl zielgruppenspezifisch als auch bevölkerungsweit etabliert werden. Vor dem Hintergrund der in einigen Medien sehr stark übertriebenen Bedrohung der "Volksundheit" war es hier relativ einfach, einen politischen Konsens über die Durchführung der Aufklärungsprogramme und die Bereitstellung entsprechender Finanzmittel zu erzielen.

Abb. 1 Beispiele für gesundheitliche Interventionsprogramme

Gesundheitliche Aufklärungsprogramme (AIDS), Krebsfrüherkennungsuntersuchungsprogramme, Gesundheitsförderungsprogramme, Herzkreislaufpräventionsprogramme, Enthospitalisierungsprogramme, Screeningprogramme (HPV/Chlamydien), Drogennachsorgeprogramme, Präventionsprogramme zur Vermeidung von Pflegebedürftigkeit. Programme zur Verringerung der Frühgeburten, Betriebliche Gesundheitsförderung

Die Programmevaluation lässt sich damit wie folgt definieren:

„Systematische Anwendung sozialwissenschaftlicher Forschungsmethoden zur Beurteilung der Konzeption, Ausgestaltung, Umsetzung und des Nutzens sozialer Interventionsprogramme.“ (Rossi 1978)

Als Definition für Evaluationsmaßnahmen im Gesundheitswesen spezifizieren wir wie folgt:

Anwendung sozial- und gesundheitswissenschaftlicher Methoden zur Analyse von Organisations- und Versorgungsproblemen im Gesundheitswesen und der zur ihrer Lösung vorgeschlagenen und / oder eingesetzten Programme mit dem Ziel, die Effektivität und Effizienz der durchgeführten Intervention wissenschaftlich zu bewerten. (Kirschner 1998)

Entsprechend dieser Definition kann Evaluationsforschung prinzipiell zu unterschiedlichen Zeitpunkten zur Anwendung kommen, wobei die Einbeziehung der Evaluationsforschung und v.a. die der Ergebnisevaluation (leider) am häufigsten erst während der Laufzeit oder gar kurz vor Beendigung des Programms erfolgt.

Mögliche Einbeziehung der Evaluationsforschung und Häufigkeit

1. bereits bei der Problemanalyse (sehr sehr selten)
2. bei der Programmplanung (selten)
3. bei der Programmimplementierung und -durchführung (noch selten)
4. kurz vor Beendigung des Programms (am häufigsten)
5. nach der Beendigung des Programms (auch nicht selten)

Die Qualität der Evaluationsforschung kann bereits an drei wesentlichen Kriterien (vorab) eingeschätzt werden:

- a) dem tatsächlichen oder nur legitimatorischen Interesse an der Initiierung von Programmen durch die "Politik"
- b) der Wirkungszusammenhänge im Sinne von hoher oder geringer (oder keiner) Kausalität
- c) dem Zeitpunkt der Einbeziehung des Evaluationsforschers.

Gute Evaluationsstandards - natürlich unter der Voraussetzung der Etablierung optimaler bzw. angemessener Evaluationsdesigns - sind dann zu erwarten, wenn die "Politik" tatsächlich an der Problemreduktion und Evaluation interessiert ist, Zusammenhänge oder gar kausale Wirkungsbeziehungen zwischen den Zielvariablen und unabhängigen Variablen gegeben und Letztere einer Intervention zugänglich sind und die Evaluationsforscher möglichst frühzeitig in den Prozess der Programmplanung eingebunden werden.

„Legitimatorische Programme“ ohne klare Wirkungsbeziehungen, bei denen der Evaluationsforscher - wenn überhaupt - erst gegen oder nach Ende des Programms einbezogen wird, lassen schlechte Evaluationsstandards erwarten und zeigen in der Regel keine oder keine brauchbaren Ergebnisse.

## **1.2 Vom Problem zum Programm, vom Programm zur Evaluation von der Evaluation zur Programmmodifikation**

Idealtypisch sind die verschiedenen Phasen und Entscheidungssituationen im Rahmen der Programmplanung, -durchführung und Evaluation in der folgenden Abbildung beschrieben. Zunächst wird ein Problem nach erster wissenschaftlicher Bearbeitung in der Regel durch Medien thematisiert oder skandalisiert, bis sich auch politische Entscheidungsträger bzw. -instanzen mit dem Problem beschäftigen.

Nach konsensuellem Beschluss, etwas gegen das Problem unternehmen zu wollen (oder zu müssen), erfolgt die weitere Problemanalyse, die Klärung der Interventionsfähigkeit des Problems, die konkrete Programmplanung und dann der Start des Programms (Programmimplementation). Mit dem Start des Programms beginnt die Prozess- und Produktevaluation und zu einem Zeitpunkt, zu dem ausreichende Ergebnisse überhaupt erwartet werden können, die Ergebnisevaluation. Die Ergebnisevaluation prüft in der Wirkungsanalyse, ob bzw. in welchem Umfang das Problem reduziert wurde. Sollten die Veränderungen gering sein, war das Programm "offensichtlich" unwirksam. Hier kommt es darauf an, die möglichen Gründe zu analysieren und zu bestimmen. Sind die Veränderungen so wie geplant oder erwartet, wird in der Regel Wirksamkeit unterstellt. Aber Vorsicht: Hier lauern Fehlermöglichkeiten! Offensichtlich unwirksame Programme werden entweder eingestellt oder weiter modifiziert.

Die Fehlermöglichkeiten in der Evaluationsforschung ergeben sich aus der Logik des Hypothesentestens und entsprechen den Fehlern beim generellen Testen von Hypothesen. Auch in der Evaluationsforschung werden also Hypothesen getestet. (Beispiel:  $H_0$ : Bewegungstherapie wirkt gleich wie Balneotherapie;  $H_1$ : Wirkung von Bewegungstherapie ist besser).

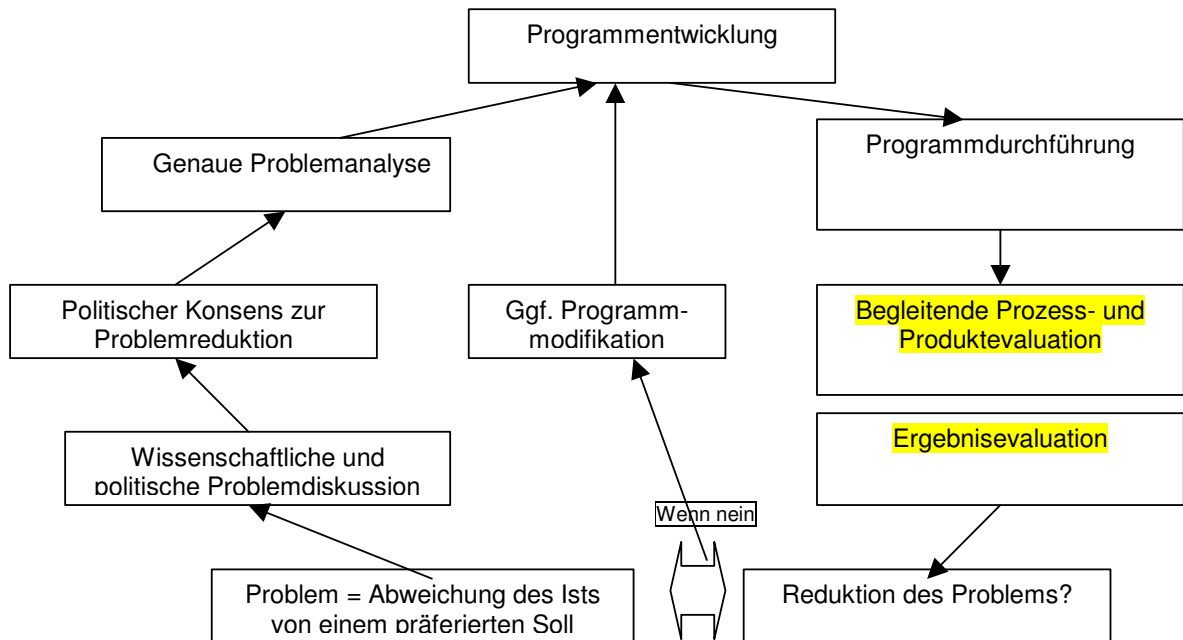


Abb. Fehlermöglichkeiten beim Testen von Hypothesen

Wahrheit unbekannt	Hypothese Ho= normal    H1=erhöht	
Blutdruck in A Stadt normal	Richtig	Falsch 1
Blutdruck in A Stadt erhöht	Falsch 2	Richtig
Falsch 1 = Fehler 1. Art	Hypothese H1 akzeptiert und Ho abgelehnt	
Falsch 2 = Fehler 2. Art	Hypothese H0 akzeptiert und H 1 abgelehnt	

Nur die empirisch gestützten Entscheidungen in der Vierfeldertafel für A und C erbringen die zutreffenden Lösungen, in den beiden anderen Fällen werden Maßnahmen für wirksam erklärt, obwohl sie es nicht sind, bzw. Maßnahmen für unwirksam, obwohl sie es sind. Beide Fehlermöglichkeiten sind in der Logik gleich negativ zu bewerten, in der Realität ist der Nachweis wirksamer, tatsächlich aber unwirksamer Maßnahmen dagegen problematischer, da Programmen Wirksamkeit attestiert wurde, die tatsächlich unwirksam sind, womit sie häufig weiterlaufen oder sogar noch ausgeweitet werden.

### 1.3 Arbeitsschritte bei der Evaluation von Programmen

Aus der Definition der Evaluationsforschung leiten sich auch die Arbeitsschritte im Rahmen der Evaluation ab.

- Zunächst ist mit Verfahren der Problem- und Defizitanalyse das gegebene gesundheitliche und soziale Problem zu beschreiben und zu analysieren.
  - Danach ist zu untersuchen, ob das Problem interventionsfähig ist, d.h. durch den Einsatz spezifischer Maßnahmen das Problem reduziert werden kann. Hierzu ist ein Wirkungsmodell zu entwickeln.
  - Auf der Grundlage des Wirkungsmodells sind Programmziele, Zielpopulation, Programmmaßnahmen zu definieren.
  - Danach erfolgt die Phase der Programmkonzeption und -implementierung
  - Ist das Programm implementiert, erfolgt die Phase der Programmüberwachung, Prozess- und Produktevaluation und Qualitätssicherung
  - Zu einem Zeitpunkt, zu dem Programmwirkungen überhaupt erwartet werden können, ist mit den dann durchzuführenden Wirkungsanalysen die Ergebnisevaluation einzuleiten und
  - mit vergleichenden Kostenuntersuchungen oder Kosten-Nutzenuntersuchungen abzuschließen. (Programmeffektivität und -effizienz)

Diese Arbeitsschritte sind bei der Planung und Einrichtung neu zu etablierender Programme augenfällig und unerlässlich. Aus den bereits genannten Gründen besteht die Aufgabe des Evaluationsforschers allerdings seltener in der sog. ex-ante Evaluation, d.h. der Forscher wird in die Planung des Programms von Beginn an einbezogen, häufiger besteht seine Aufgabe darin, ein laufendes oder abgeschlossenes Programm ex-post zu evaluieren. Hierbei ergeben sich allerdings dieselben Arbeitsschritte, d.h. der Forscher muss den Prozess der Programmentstehung differenziert rekonstruieren. Es dürfte aber klar sein, dass eine derartige ex-post Evaluation ungleich schwieriger ist als eine ex-ante oder begleitende Evaluation, da selbst bei guter Dokumentation der Entstehungsgeschichte eines Programms oft viele wichtige Informationen nicht mehr zugänglich sind.

#### **1.4 Zusammenfassung und Abgrenzung**

Evaluationsforschung untersucht die Wirksamkeit und Wirtschaftlichkeit von Interventionen, die darauf gerichtet sind, gesundheitliche oder soziale Problemlagen zu vermindern. Häufig erfolgen derartige Interventionen in Programmen.

Von der Evaluationsforschung abzugrenzen sind Verfahren der Kontrolle und Optimierung von Produkten, Dienstleistungen oder Prozessen, wie:

- das betriebswirtschaftlichen Controlling
- die Qualitätssicherung
- das Qualitätsmanagement
- die evidence based medicine
- das diseases management
- die epidemiologischen Überwachungsmethoden

Die Evaluationsforschung ist Teil rationaler Interventionen oder Politiken, die sich im Rahmen einer evaluativen „Schleife“ bewegt. In der Evaluationsforschung werden Hypothesen getestet, womit die generellen Fehlermöglichkeiten beim Testen von Hypothesen bestehen.

## **1.5 Zwei Beispiele zur Evaluation von Programmen**

### **1.5.1 Evaluation des Programms zur Verringerung von Frühgeburten in Deutschland BabyCare<sup>1</sup>**

#### **Beispiel für die Ableitung eines Wirkungsmodells**

Im Rahmen der Entwicklung des Programms zur Reduktion von Frühgeburten in Deutschland BabyCare wurde zunächst die internationale epidemiologische Datenlage zu Frühgeburten und zu Interventionen zur Verringerung von Frühgeburten analysiert.

Abb. Risikofaktoren der Frühgeburt

---

<sup>1</sup> Informationen unter [www.baby-care.de](http://www.baby-care.de)

**Untersuchte Risikofaktoren der Frühgeburt**

Lebensbedingungen und Verhalten	Primär medizinische Risikofaktoren
Soziale Lage, Bildung, Arbeitsbelastungen	Multiple Schwangerschaft
Alter	Vorausgegangene Frühgeburt
Region	Vorausgegangenen Fehlgeburt
Rauchen	Infertilitätsbehandlung
Andere Drogen (Alkohol, ill. Drogen)	Induzierte Abtreibungem
Ernährung	Anämie
Body Mass Index	Vaginalinfektionen
Sport	Krankheiten (Hypertonie, Diabetes)
Stress, psych. Belastungen	Blutungen
Informationsstand	Andere mediz. Komplikationen
Qualität und Inanspr. der gyn. Versorgung	Parodontitis

Dabei wurden für die medizinisch-klinischen Risikofaktoren der Frühgeburt folgende Assoziationsmaße gefunden:

Multiple Schwangerschaften	RR 8
Vorausgegangene Frühgeburt(en)	RR 3-4 / OR 5,7 - 9
Infertilitätsbehandlung	OR 3,5
Vorausgegangene Tot- oder Fehlgeburt	RR 1,6 - 2
Induzierte Abtreibungen (>2 oder 3)	RR 1,6 - 2
Plazentaablösungen	OR 2,8
Anämie	OR 1,3
Aszendierende Vaginalinfektionen	RR 1,3 -1,5 / OR 2,0
Blutungen (vor der 24.SSW)	OR 1,4-2,5
Schwangerschaftshypertonie	OR 1,7-1,8

Im Rahmen der Analyse der Assoziationen mit dem BabyCare Datensatz ergaben sich folgende Zusammenhänge.

Abb. Assoziationsmaße zwischen Risikofaktoren und der Frühgeburt (Basis: Primiparae n=803)

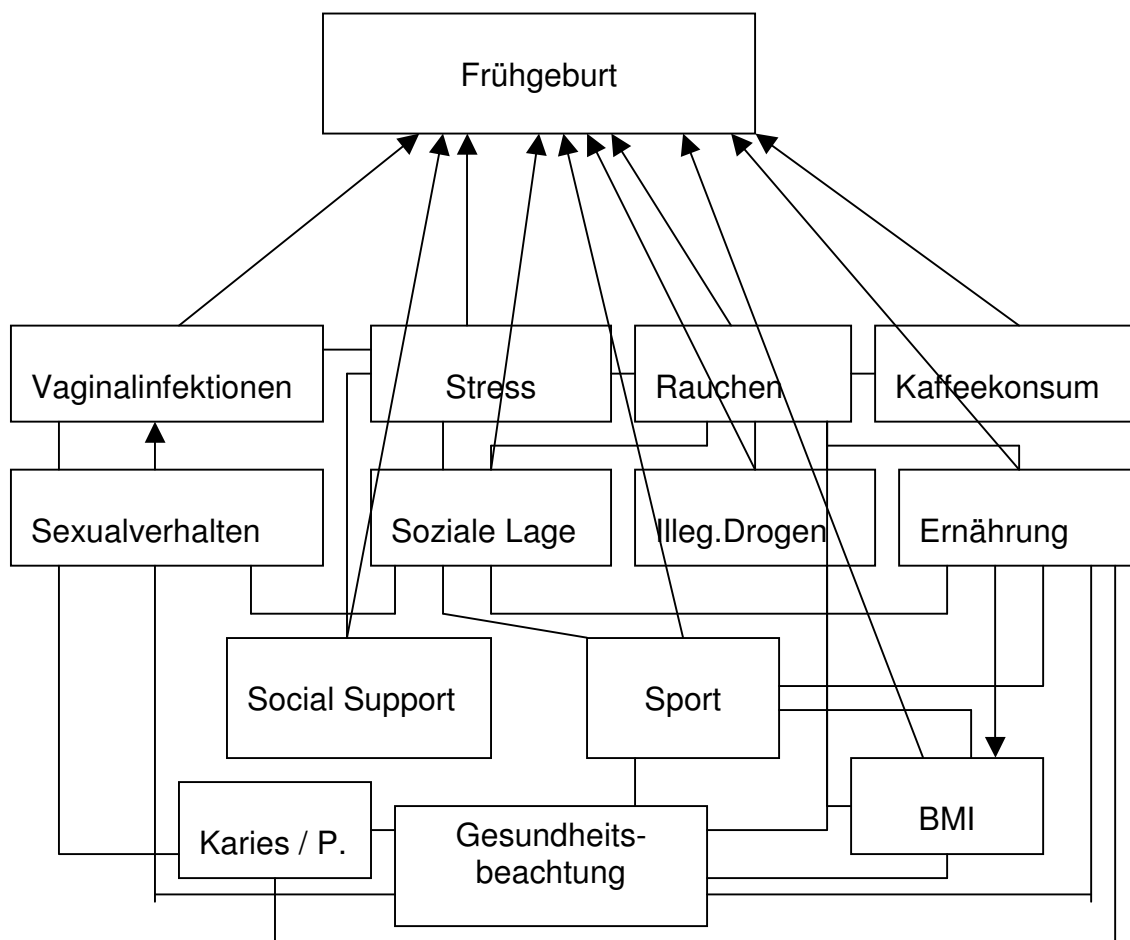
<b>Assoziationsmaße für Primiparae</b>							
Frühgeburtenrate: 7,5%							
Gewichtet nach Alter	Gesamt	FG	NFG	OR	KONF 95%	RR	(RR-1)in%
n	803	60	743				
Expositionsprävalenz							
Diabetes	1,0%	8,5%	0,4%	22,8	8,3	63,2	9,1 810%
Mehrlingsschwangerschaft	1,4%	10,8%	0,7%	17,9	7,4	43,6	8,4 736%
Verwendung illegaler Drogen in den letzten letzten 12 Monaten	3,9%	10,8%	3,4%	3,5	1,5	8,2	3,0 196%
Mind. 1 Tag Krankenhaus, letzte 12 Monate	13,6%	28,0%	12,4%	2,8	1,5	4,9	2,5 149%
Hypertonie	5,9%	12,5%	5,4%	2,5	1,1	5,6	2,3 128%
Rauchen bei Diagnose der Schwangerschaft	25,8%	41,8%	24,5%	2,2	1,3	3,8	2,1 107%
Nichtkontinuierliche Teilnahme an der Krebsfrüherkennungsuntersuchung	31,4%	45,3%	30,3%	1,9	1,1	3,2	1,8 81%
Soziale Lage nicht protektiv	44,0%	58,6%	42,8%	1,9	1,1	3,2	1,8 80%
Starke Belastungen letzte 12 Monate	39,5%	53,6%	38,4%	1,8	1,1	3,0	1,7 71%
Migräne	21,1%	31,3%	20,3%	1,8	1,0	3,2	1,7 70%
Viel emotionale Unterstützung	91,7%	84,6%	92,3%	0,4	0,9	0,2	0,5 -54%
Multivitaminensaft (mehrmals) täglich	27,6%	14,6%	28,7%	0,4	0,9	0,2	0,4 -55%
Gynäkologische Operationen inkl. Schwangerschaftsabbrüche (lifetime)	14,7%	23,2%	14,0%	1,9	1,0	3,5	1,8 75%
Karies, Parodontose letzte 12 Monate	35,7%	45,7%	34,9%	1,5	0,9	2,6	1,5 47%
Betacarotin >110%	37,2%	24,5%	38,2%	0,6	1,0	0,3	0,6 -43%
Betacarotin <90%	45,9%	55,4%	45,1%	1,5	0,9	2,5	1,4 43%
Gesundheitsbeachtung sehr stark / stark	55,4%	43,5%	56,4%	0,6	1,0	0,4	0,6 -38%
Krebsfrüherkennungsuntersuchung jemals	76,9%	66,9%	77,7%	0,6	1	0,3	0,6 -40%

Prüft man nun, welche dieser Risikofaktoren (und auch protektiven Faktoren) einer Intervention zugänglich sind, bleiben:

- Vaginalinfektionen
- Verwendung illegaler Drogen
- Rauchen
- Belastungen / Stress
- Social support
- Karies, Parodontose
- Ernährung (β-Karotin)
- Gesundheitsbeachtung
- Sportliche Aktivität
- Kaffeekonsum
- BMI

Unter Berücksichtigung der insgesamt gefundenen Assoziationen könnte ein Wirkungsmodell wie folgt aussehen.

Abb. Wirkungsmodell zu Zusammenhängen zwischen der Frühgeburt und unabhängigen Variablen

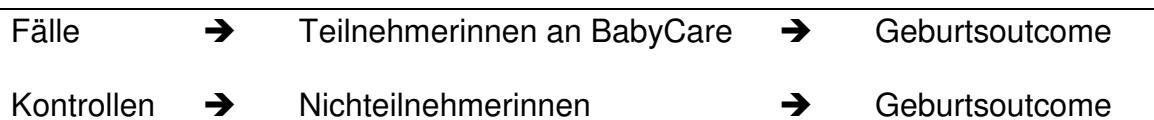


Zu beachten sind die zahlreichen wechselseitigen Abhängigkeiten und Interkorrelationen. Das Schaubild dürfte klar machen, dass zur Analyse von Reduktionseffekten der Frühgeburt erheblich große Fallzahlen erforderlich sind, da die jeweiligen Interkorrelationen durch geschichtete Analysen zu kontrollieren sind.

**Interventionen zur Verringerung von Frühgeburten stützen sich also auf Ergebnisse kausalanalytischer epidemiologischer Forschung, sie sind kausalanalytisch fundiert.**

### Beispiel für die Ergebnisevaluation

Klassisches Fall-Kontrolldesign



### 1.5.2. Evaluation von Maßnahmen der Gesundheitsförderung bei Arbeitslosen

In der Mehrzahl der Programme und Maßnahmen der Gesundheitsförderung liegt aber kaum oder nur wenig bzw. keine kausalanalytische Evidenz vor. Häufig nur multiple Korrelationen bis hin zu lediglich plausiblen Vermutungen.

**Intervention und Evaluation sind methodisch immer abhängig von der Evidenz in der sozialepidemiologischen Forschung und der Evaluationsforschung.**

Hierbei lassen sich idealtypisch die in der folgenden Abbildung beschriebenen Situationen unterscheiden.

Die günstigste Entscheidungssituation ist D. Hier liegt hohe Evidenz sowohl aus der analytischen Forschung als auch aus bereits durchgeführten Interventions- und Evaluationsstudien vor. Die Situation B ist praktisch nicht sehr wahrscheinlich, impliziert sie doch, dass viele evaluierte Interventionen auf der Grundlage ungesicherter analytischer Zusammenhänge durchgeführt worden sind.

Abb. Idealtypische Situationen zum Evidenzniveau in der analytischen Forschung und der Evaluationsforschung

Niveau der Evidenz	Evaluationsforschung	
Analytische Forschung (z.B. Sozialepidemiologie)	Gering	Hoch
Gering	<b>A</b>	(B)
Hoch	C	D

Die Situation C markiert die Situation, dass bei hohem analytischen Evidenzniveau nur geringe Evidenz aus der Evaluationsforschung vorliegt, was unterschiedliche Gründe haben kann:

1. die Interventionen waren unwirksam
2. die Interventionen wurden nicht, oder nicht angemessen evaluiert
3. die Interventionen wurden als unwirksam bewertet, obwohl sie wirksam waren ( $\beta$ -Fehler)

Die Situation A markiert einen ungesicherten Wissenstand über Zusammenhänge zwischen einem Problem und die dieses (mit)verursachenden Faktoren und einen sehr geringen Kenntnisstand aus der Evaluationsforschung, womit sie die aktuelle Lage auf dem Gebiet der Arbeitslosigkeit recht genau beschreiben dürfte.

In der Situation A muss ein Programmplaner und ex-ante beteiligter Evaluationsforscher der Finanzierungseinrichtung deutlich machen, dass positive Interventionswirkungen – aus zu benennenden Gründen – plausibel sind, dass sich die Intervention aber nicht auf ein analytisch abgesichertes Modell begründet, vielmehr **ein prinzipiell ergebnisoffenes Experiment darstellt**, ein Forschungsansatz der häufig auch die Methoden der Handlungsforschung umfasst. Die Konzeptionalisierung, Umsetzung und Evaluation derartiger Experimente impliziert z.T. völlig andere Verfahren und Methoden, als dies bei klassischen Interventionen der Fall ist. Entsprechend hoch ist hier der Forschungs-, Entwicklungs- und Lernbedarf.

### 1.5.3 Zusammenfassung

Interventionen basieren auf einem bestimmten wissenschaftlichen Kenntnisstand hinsichtlich des zu intervenierenden Problems und der damit korrelativ oder kausal in Verbindung stehenden Faktoren (Wirkungsmodell). Je höher die Zahl der Einflussfaktoren und je geringer die Assoziationsmaße zwischen der zu intervenierenden Zielvariablen und den mit dieser in Verbindung stehenden Faktoren ist, desto komplexer werden sowohl Intervention als auch Evaluation. Noch komplexer werden diese bei eher unspezifischen Interventionen, z.B. bei Maßnahmen der Erhöhung von Ressourcen. Mit steigendem Komplexitätsgrad

nimmt die Möglichkeit ab, Interventionserfolge auf der Grundlage klassischer Fall-Kontrolldesigns nachzuweisen.

Da Interventionen meist aus politischen Gründen implementiert werden, beruhen sie häufig nicht auf einem exakten Wirkungsmodell, sondern auf Plausibilitätsüberlegungen, was die Evaluation zusätzlich erschwert.

Das im Zuge der zunehmenden Diskussion um evidence based medicine auch in die Primärprävention und Gesundheitsförderung in Deutschland eingeführte Evidenzpostulat in Zusammenhang mit dem § 20 SGB V zum ex-ante Nachweis der Effektivität und Effizienz von Interventionen ist 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 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.

#### Literatur:

Friese, K., Dudenhausen, W., Kirschner, W., Schäfer, A., Elkeles, T.: Risikofaktoren der Frühgeburt und ihre Bedeutung für Prävention und Gesundheitsförderung – Eine Analyse auf der Grundlage der Daten des Babycare.-Programms. Das Gesundheitswesen, 9/2003, S. 477-485

Elkeles, T., Kirschner, W.: Arbeitslosigkeit und Gesundheit – Interventionen durch Gesundheitsförderung und Gesundheitsmanagement – Befunde und Strategien. BKK Bericht GuS Nr. 3, NW-Verlag, Bremerhaven, 2004

Elkeles, T., Kirschner, W. : Health promotion for the unemployed – Needs, strategies and evidence on effectiveness and efficiency. 3 rd International Conference on Unemployment and health, Bremen, September 23-25, 2004, In publication <http://www.ipg.uni-bremen.de/icoh/bremen2004/>

Kirschner, W.: Babycare – Ein Beispiel erfolgreicher Prävention mit Krankenkassen, Die BKK, 11/2004, S. 497-501