

A Allgemeiner Teil

1 Einführung

Eine grundlegende Aufgabe der Evaluation (Erfolgskontrolle) besteht darin, sowohl die Stärken als auch Schwächen von Interventionen aufzuzeigen, wodurch die Grundlagen für angebotsbezogene Entscheidungen optimiert und damit die Effektivität und Effizienz erhöht werden können (vgl. BZgA, 2007, S. 349). Folglich dient die Evaluation auch den gesetzlichen Krankenkassen als wichtiges Instrument zur Bewertung und Verbesserung ihrer primärpräventiven Angebote.

Gemäß dem *Leitfaden Prävention* sind zweckmäßige Verfahren zur Erfolgskontrolle zu entwickeln, die als kontinuierliche Routinen etabliert werden. Damit soll die notwendige Transparenz erreicht werden, die die Realisierung der gesetzten Ziele und Erwartungen überprüfbar macht. Gleichzeitig fließen die erhobenen Daten in die Planung zukünftiger Aktivitäten ein, wodurch ein kontinuierlicher Verbesserungsprozess eingeleitet werden soll (vgl. GKV-Spitzenverband, 2010, S. 17).

Von den Spitzenverbänden der Krankenkassen sind bereits *gemeinsame und einheitliche Evaluationsverfahren* erarbeitet worden, mit denen krankenkassen- und handlungsfeldübergreifend die Wirksamkeit von Gesundheitskursen nachgewiesen werden soll. Hierbei waren das Institut und Poliklinik für Medizinische Psychologie am Universitätsklinikum Eppendorf in Hamburg und GESOMED (Gesellschaft für sozialwissenschaftliche Forschung in der Medizin in Freiburg) für die Erstellung, Erprobung und wissenschaftliche Analyse der Evaluationsinstrumente verantwortlich (vgl. Arge SpiK, 2008, S. 4).

Das *Handbuch zum Individuellen Ansatz (Anwenderhandbuch Teil 1)* beinhaltet alle Materialien (wie z. B. Frage- und Dokumentationsbögen) und Anleitungen für die Evaluation von Bewegungs-, Ernährungs- und Stressreduktionskursen. Methodische Aspekte und Ergebnisse aus den Feldversuchen sind in gesonderten wissenschaftlichen Berichten dargestellt, die jedoch nicht öffentlich zugänglich

sind, aber bei den Herausgebern angefordert werden können (vgl. Arge SpiK, 2008, S. 3).

Die Testergebnisse aus der Erprobung der einheitlichen GKV-Fragebögen in über 200 Gesundheitskursen verschiedener Handlungsfelder bestätigen, dass die eingesetzten Evaluationsinstrumente zur Wirkungsabschätzung gut geeignet sind und den Krankenkassen eine begleitende Qualitätssicherung ohne großen Aufwand ermöglichen (vgl. Wanek, 2009).

Auf der Grundlage der *gemeinsamen und einheitlichen Evaluationsverfahren der Spitzenverbände der Krankenkassen* (2008) wurde im Jahr 2009 damit begonnen, den Evaluationsprozess in den Präventionscentern der IKK Nord in Kooperation mit der Hochschule Neubrandenburg weiterzuentwickeln und an die aktuellen Empfehlungen anzupassen.

In diesem Buch werden die Ergebnisse aus dem erweiterten Evaluationsverfahren, das im Rahmen des Frühjahrsprogramms 2010 der IKK Nord erstmalig erprobt worden ist, für die Bewegungskurse ausführlich vorgestellt.

Die Evaluationsergebnisse liefern dabei einerseits Aussagen über die Zielerreichung sowie Anhaltspunkte zur Verbesserung der Kursangebote. Auf der anderen Seite dienen die Resultate als Argumentationsbasis, um den Wert präventiver Interventionen zu betonen; und sie erweitern das Handlungswissen der Praktiker vor Ort (z. B. Präventionscenterleiter) hinsichtlich empirischer Methoden zur Planung weiterer Evaluationsvorhaben (vgl. König, 2009, S. 309).

2 Evaluationsdesign und Evaluationsebenen

Die Evaluation im Bereich der individuellen Gesundheitsförderung wird bei der IKK Nord durch die Leiterinnen der Präventionscenter organisiert und mit Unterstützung der externen Anbieter durchgeführt. Die Erfolgskontrolle wird bei den **kassenseitig** organisierten Gesundheitskursprogrammen in Form einer Maßnahmeevaluation durchgeführt, bei der die konkreten Einzelmaßnahmen (z. B. Rückenschule oder Entspannungskurs) Gegenstand der systematischen Bewertung sind. Dabei werden alle im Rahmen des Gesundheitsprogramms angebotenen Kurse mit mindestens fünf Kurseinheiten einer Bewertung unterzogen.

Der neue Evaluationsansatz besteht aus einer längsschnittlichen Befragung, die drei Erhebungszeitpunkte umfasst. Dabei werden die gleichen Kursteilnehmer zu Kursbeginn, bei Kursende sowie sechs Monate nach Beendigung der Kurse schriftlich befragt (Prä-Post-Design mit Follow up). Zur Gewährleistung eines routinemäßigen Einsatzes und zwecks Vergleichbarkeit der Ergebnisse ist eigentlich vorgesehen, für alle Kurse bzw. Handlungsfelder die gleichen Fragebögen zu verwenden, die sich mit einer Länge von lediglich zwei Seiten auf das Wesentliche beschränken sollen (vgl. Arge SpiK, 2008, S. 4f).

Da allerdings in den einheitlichen Fragebögen der Spitzenverbände der Krankenkassen (im Folgenden: einheitliche GKV-Fragebögen) mehrere Dimensionen bzw. Parameter, die für die Evaluation der individuellen Gesundheitskurse bedeutsam sind, zu kurz kommen oder gänzlich fehlen, sind diese Evaluationsinstrumente an der Hochschule Neubrandenburg überarbeitet und erweitert worden. Auf die Evaluationsziele und daraus resultierenden Forschungshypothesen wird im nächsten Abschnitt eingegangen.

Um die Praktikabilität und Qualität der einheitlichen Fragebögen der Spitzenverbände der Krankenkassen direkt mit den überarbeiteten Fragebögen der Hochschule Neubrandenburg vergleichen zu können, wurden beide Versionen im Frühjahrskursprogramm 2010 der IKK Nord zur Evaluation verwendet. In den

Präventionscentern Kiel/Lübeck, Rostock und Flensburg/Rendsburg wurden die einheitlichen GKV-Fragebögen in leicht modifizierter Form (vgl. Anhänge 1 und 2) und in Neubrandenburg, Schwerin und Itzehoe die Fragebögen der Hochschule Neubrandenburg eingesetzt (vgl. Barkusky/Claßen, 2011).

Die folgenden Auswertungen der Befragungsdaten sollen Aufschluss darüber geben, ob bzw. inwieweit sich die erweiterten Fragebögen der Hochschule von denen der GKV-Spitzenverbände unterscheiden hinsichtlich:

1. der Akzeptanz (Verringert der längere Fragebogen den Rücklauf; gibt es mehr fehlende Werte?);
2. der Gütekriterien (Objektivität, Reliabilität, Validität, Veränderungssensitivität usw.) und
3. der gemessenen Gesundheitswirkungen

(vgl. Barkusky/Claßen, 2011, S. 13f).

Die Kursrahmendaten werden in einem gesonderten Dokumentationsbogen (I-Dok-Bogen) erfasst, der u. a. die Kursdauer, Anzahl der Treffen sowie die Teilnehmerzahl zu Kursbeginn und –ende (für Ermittlung der Rücklaufquote) enthält (vgl. Arge SpiK, 2008, S. 5).

Die Evaluation erfolgt nach dem Eingruppenplan, da ausschließlich die an der Maßnahme teilnehmende Gruppe (Experimentalgruppe) getestet wird. Auf ein Evaluationsdesign mit Kontrollgruppen wird verzichtet, weil es sich aufgrund des damit verbundenen Aufwands nicht für ein Routinesystem eignet (vgl. Arge SpiK, 2008, S. 4). Damit orientiert sich das Untersuchungsdesign an den Realitäten der Intervention und nicht ausschließlich an den methodischen Vorgaben. Die Evaluation soll die Interventionen so wenig wie möglich beeinträchtigen – der Fokus auf aufwendige Methoden der Fehlerkontrolle kann die Effektivität und Effizienz der Maßnahmen durchaus gefährden (vgl. BZgA, 1999, S. 70 ff).

Aufgrund der fehlenden Kontrollgruppen kommt der Verfügbarkeit von Normdaten eine besondere Bedeutung zu, um Ausgangslage und Wirkungen abschätzen zu können. Vergleiche zur nationalen Gesundheitsberichterstattung sind möglich,

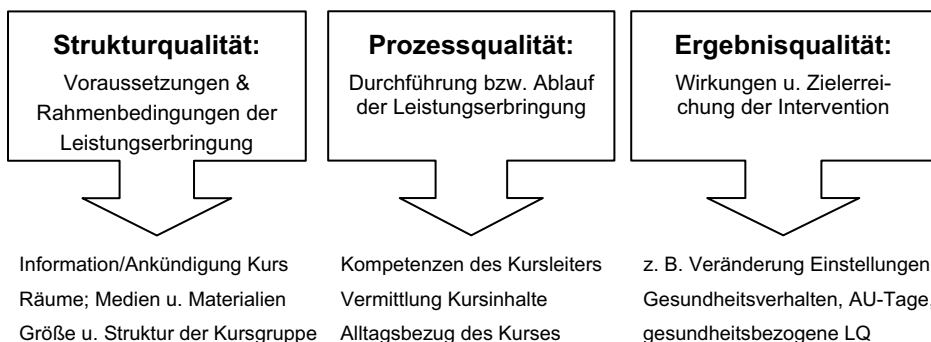
wenn die Normdaten aus großen Studien gewonnen werden (vgl. Arge SpiK, 2008, S. 4).

Zur Gewährleistung des Datenschutzes sind die Fragebögen der IKK Nord anonymisiert und enthalten eine „Code-Nr.“, die sich aus Geburtsdatum des Befragten, Geschäftsstellen-Nummer des Präventionscenters und Kursnummer zusammensetzt. Dadurch können die Befragungsdaten der einzelnen Messzeitpunkte einer Person zusammengeführt werden, was einen längsschnittlichen Vergleich ermöglicht.

Aufgrund der wesentlich größeren Datenmenge, die durch das neue Evaluationsverfahren entsteht, und der Erprobung neuer Fragebögen sind die Daten extern durch die Hochschule Neubrandenburg unter Anwendung des Statistikprogramms SPSS ausgewertet worden.

Evaluationsebenen

In den einheitlichen GKV-Fragebögen und in den überarbeiteten Fragebögen der Hochschule Neubrandenburg werden mehrere Qualitätsebenen (Struktur-, Prozess- und Ergebnisqualität) der Präventionskurse evaluiert und verschiedene Gesundheitsdimensionen (physisch, psychisch, Ressourcen usw.) sowie relevante Confounder (Individualmerkmale der Zielgruppen) berücksichtigt, wobei die Bewertung der Gesundheitseffekte und der Zielerreichung im Vordergrund steht.



Quelle: eigene Darstellung

Abbildung 1: Evaluationsebenen der Kursfragebögen

Im Bereich der Ergebnisqualität sind zur Erfassung des *Gesundheitsverhaltens* Skalen für Ernährung, Bewegung und Stress/Entspannung entwickelt worden. Für jedes der drei Handlungsfelder sind jeweils vier Items formuliert worden, die sich an den Zielen des GKV-Leitfadens Prävention orientieren (vgl. Arge SpiK, 2008, S. 11).

Die Fragebögen der Hochschule Neubrandenburg sind ähnlich aufgebaut wie die einheitlichen GKV-Fragebögen. Nach der Einstiegsfrage folgen Fragen zum Gesundheitsverhalten, Gesundheitszustand und zur gesundheitsbezogenen Lebensqualität. Darüber hinaus werden ebenfalls AU-Tage, Gewicht, Größe, soziodemographische Daten und Kontrollvariablen erfasst. Allerdings erfordert jedes Handlungsfeld zusätzlich spezifische Fragestellungen, weshalb handlungsfeldbezogene Erhebungsinstrumente erstellt wurden. Insofern gibt es für jedes der vier Handlungsfelder Bewegung, Ernährung, Stress/Entspannung und Suchtmittelkonsum je einen Eingangs-, Schlussfragebogen und Nachbefragungsbogen (I-1, I-2; I-3).

Der theoretische Hintergrund zur Überarbeitung der einheitlichen GKV-Fragebögen und die Erhebungsinstrumente sind im Handbuch zur Evaluation von Präventionsangeboten nach dem individuellen Ansatz enthalten (vgl. Barkusky/Claßen, 2010).

Um bestimmte Perspektiven zu vertiefen und die aus den schriftlichen Befragungen gewonnenen Erkenntnisse zu erweitern, sind im Rahmen der Evaluation von Gesundheitskursen der IKK Nord auch *qualitative Datenerhebungsmethoden* angewendet worden. Hierzu wurden Versicherte, die an Gesundheitskursen in Neubrandenburg teilgenommen haben, gebeten, in narrativen Interviews ihre Lebensgeschichte zu erzählen. Dies ermöglicht einen genaueren Einblick in die Bedeutung der besuchten Kurse für die Gesundheit und das Leben der Kursteilnehmer. Hieraus können wichtige Rückschlüsse für die weitere Gestaltung von Kursangeboten gezogen werden, indem z. B. die Maßnahmen noch mehr auf die Bedarfe der zukünftigen Teilnehmer zugeschnitten werden.

3 Evaluationsziele und die Ableitung von Hypothesen

Bereits bei der Evaluationsplanung ist zu klären, aus welchen Gründen Interventionen evaluiert werden sollen, da die Zwecke bzw. Funktionen der Maßnahmenbewertung unterschiedlich sein können. Daraus sind konkrete Evaluationsziele ableitbar, die als Ausgangspunkt für die Formulierung spezieller Forschungshypothesen dienen.

Für die Krankenkassen sind im Bereich der individuellen Gesundheitsförderung folgende Evaluationsziele vordergründig:

- Kontrolle des Erfolgs bzw. der Wirksamkeit von Maßnahmen;
- Legitimation von Arbeitsbereichen (d. h. für die IKK Nord die Legitimation der Präventionscenter);
- Qualitätsverbesserung der Angebote bzw. ihrer Struktur (u. a. durch die Bewertung von Prozessen);
- Qualifizierung der Mitarbeiter/Innen hinsichtlich ihrer empirischen Kompetenz

(vgl. hierzu auch König, 2009, S. 303).

Ausgehend von diesen Evaluationszielen sind mehrere Qualitätsebenen in die Bewertung von Interventionen einzubeziehen, denen konkrete Zielvariablen bzw. Indikatoren zugeordnet werden können (vgl. Barkusky/Claßen, 2010, S. 16 ff):

Ergebnisqualität

- Veränderungen in *Gesundheitsverhalten* bzw. *Einstellungen*
- Veränderungen in *Gesundheitszustand* und *gesundheitsbezogener Lebensqualität* (z. B. Verbesserung des allgemeinen Gesundheitszustands und des Wohlbefindens; Reduzierung von Beschwerden und Schmerzen)
- *nachhaltige Effekte* der Interventionen (längerfristige Anwendung des neu Erlernten im beruflichen/privaten Alltag)

Strukturqualität

- Erfassung der Zufriedenheit teilnehmender Versicherter mit *Rahmenbedingungen* des Kurses (z. B. Räumlichkeiten, verwendete Materialien, Teilnehmergruppe) und
- *Kursorganisation* (Information und Ankündigung bzgl. Kursangebot, Organisation der Anmeldungen usw.)

Prozessqualität

- Erfassung der Zufriedenheit teilnehmender Versicherter mit *Kurs* bzw. *Kursleiter* (z. B. Informationsgehalt der Kurse; Vermittlung der Kursinhalte und Alltagsbezug)
- *Zielgruppenerreichung*: d. h. zielgruppengerechte Inanspruchnahme der Präventionsangebote (Übergewichtige in Ernährungskursen, Stressbelastete in Entspannungskursen usw.); Beitrag zur *Verminderung sozial bedingter Ungleichheit von Gesundheitschancen* (werden Angehörige sozial schwächerer Schichten erreicht?) sowie Erreichung speziell oder besonders stark belasteter Zielgruppen

Die zu den Evaluationszielen generierten *Hypothesen* bzw. *Forschungsfragen* über die zu erwartenden Veränderungen bzw. Effekte im Kursverlauf orientieren sich eng an den Zielen des GKV-Leitfadens Prävention, die für die Präventionsangebote der einzelnen Handlungsfelder formuliert worden sind (vgl. hierzu Baruksky/ Claßen, 2010, S. 50 ff).

Im Folgenden sind die allgemeinen Annahmen sowie die spezifischen Hypothesen für die Bewegungskurse aufgeführt:

I. handlungsfeldübergreifende Annahmen:

1. *Durch die Teilnahme an Gesundheitskursen/Präventivangeboten verbessert sich der allgemeine Gesundheitszustand der Teilnehmer.*
2. *Durch die Teilnahme an Gesundheitskursen/Präventivangeboten verbessert sich die Lebensqualität der Teilnehmer (umfasst u. a. Verbesserung des psychischen/seelischen und körperlichen Wohlbefindens sowie Verringerung von Schmerzen).*

3. *Durch die Teilnahme an Gesundheitskursen/Präventivangeboten fühlen sich die Teilnehmer seltener krank bzw. arbeitsunfähig.*
4. *Durch die Teilnahme an Gesundheitskursen/Präventivangeboten reduzieren sich bei den Teilnehmern Risikofaktoren (Übergewicht/Rauchen).*
5. *Inwieweit führen die Teilnehmer nach Kursende die erlernten Übungen selbstständig fort bzw. integrieren die erworbenen Fähigkeiten in ihren (beruflichen) Alltag?*

II. Hypothesen zu Bewegungsangeboten:

1. *Durch die Teilnahme an Bewegungskursen verbessern sich die physischen Gesundheitsressourcen (z. B. Ausdauer, Kraft u. Entspannungsfähigkeit) der Versicherten.*
(Kernziel 1 im Leitfaden Prävention)
2. *Durch die Teilnahme an Bewegungsangeboten werden die psychosozialen Gesundheitsressourcen – hier insbesondere die Selbstwirksamkeit – gestärkt.*
(Kernziel 2 im Leitfaden Prävention)
3. *Versicherte werden durch die Teilnahme an Bewegungsangeboten sportlich aktiver bzw. treiben regelmäßiger/häufiger Sport.*
(zentrales Kernziel 5 im Leitfaden Prävention)
4. *Aufgrund der Teilnahme an Bewegungskursen verringern sich Beschwerden (wie z. B. Rückenbeschwerden u. Muskelverspannungen).*
(Kernziel 4 im Leitfaden Prävention)
5. *Ferner wird angenommen, dass sich aufgrund der Inanspruchnahme von Bewegungsmaßnahmen der wahrgenommene Stress verringert.*

Die obigen Hypothesen sind anhand statistischer Testverfahren überprüft worden, deren Ergebnisse ausführlich in Teil B und C dieses Buches dargestellt sind.

Darüber hinaus sind mehrere Annahmen zu Kontrollvariablen bzw. potenziellen Einflussfaktoren auf die Kurseffekte formuliert worden (vgl. Barkusky/Claßen, 2010, S. 53 ff) – die Ergebnisse der statistischen Analyse werden voraussichtlich nach Durchführung der Nachbefragung in den Abschlussbericht einfließen.

4 Vorgehensweise bei der Datenauswertung (Methodik)

Ausgehend von anderen Evaluationsstudien bzw. Reviews im Bereich der Prävention und Gesundheitsförderung (Walter et al., 2002; Ogilvie et al., 2007; Goldgruber/Ahrens, 2009; WIdO, 2003; Müller-Riemenschneider et al., 2008; Lenhardt, 2001; Stöbel et al., 1998) und den Ergebnissen aus den Feldtests zu den einheitlichen GKV-Fragebögen (Wanek, 2009) wurden auch bei der Entwicklung und Erprobung der Hochschul-Fragebögen bezüglich der Gesundheitswirkungen von individuellen Präventionsangeboten schwache bis mittlere Effektgrößen erwartet.

Zuerst sind die Erhebungsinstrumente hinsichtlich verschiedener Gütekriterien geprüft worden, wozu z. B. die Reliabilitätsanalyse gehört, um die Zuverlässigkeit der Skalen zu untersuchen oder die Analyse fehlender Werte zur Ermittlung der Akzeptanz.

Anschließend sind alle Skalenwerte ermittelt worden, d. h. für das Gesundheitsverhalten im Bereich Bewegung, Ernährung und Stress (einheitliche GKV-Fragebögen) sowie für die allgemeine Selbstwirksamkeitserwartung (SWE), die Stresswahrnehmung (PSS) und die physischen Gesundheitsressourcen, indem die angekreuzten Itemwerte addiert und aus der Summe das arithmetische Mittel gebildet wurden. Bei allen Skalen sind in die Berechnung der Skalenwerte nur Fälle eingegangen, in denen über die Hälfte der Items beantwortet wurden, d. h. für die Selbstwirksamkeitserwartung mindestens sechs gültige Items und für das Gesundheitsverhalten, die Stresswahrnehmung und die physischen Gesundheitsressourcen mindestens drei gültige Items.

Bei der Auswertung der einheitlichen GKV-Fragebögen sind die drei Skalen zum Gesundheitsverhalten dahingehend umkodiert worden, dass ein höherer Wert eine positivere Ausprägung der Skala bedeutet. Dadurch können die Testergebnisse leichter interpretiert werden. Die drei Itemgruppen sind entweder in Mittelwer-

ten für die jeweiligen Handlungsfelder oder als Gesamtwert für Gesundheitsverhalten auswertbar (vgl. Arge SpiK, 2008, S. 11).

Bezüglich der Fragen zur gesundheitsbezogenen Lebensqualität (SF-12) wurden in den HS-Fragebögen die körperliche und die psychische Summenskala mit der speziellen Auswertungssoftware für den SF-12 berechnet. Die Summenskalen werden als fehlend betrachtet, wenn eines der zwölf Items des SF-12 nicht beantwortet wurde (vgl. Bullinger/Kirchberger, 1998, S. 68).

In den einheitlichen GKV-Fragebögen, die den SF-36 nur auszugsweise enthalten, sind die Subskalen „körperliche Schmerzen“, „Vitalität“ und „Psychisches Wohlbefinden“ ebenfalls vorschriftsmäßig berechnet und auf eine 100%-Skala transformiert worden, wobei nur Fälle berücksichtigt wurden, in denen mindestens 50 % der Items gültig sind (vgl. Bullinger/Kirchberger, 1998, S. 17).

In einem weiteren Auswertungsschritt wurden neue Variablen, wie z. B. der BMI, generiert und mehrere Variablen in neue umkodiert – so wurden u. a. das Alter in Altersklassen und der BMI in BMI-Stufen eingeteilt.

Daran hat sich die deskriptive Auswertung des Datenmaterials angeschlossen, indem die Stichprobe ausführlich anhand von Häufigkeiten, Diagrammen und Mittelwerten beschrieben wird im Hinblick auf soziodemographische Daten, Gesundheitsverhalten, Gewichtsverteilung usw.

Zur längsschnittlichen Untersuchung von Effekten sind die Daten des Messzeitpunktes I-1 und I-2 in einem Gesamtdatensatz zusammengeführt worden, wobei aufgrund der speziellen Fragen in den HS-Fragebögen für jedes Handlungsfeld gesonderte Datensätze erstellt wurden. Der Prä-Post-Vergleich erfolgt mittels parametrischer oder nichtparametrischer Signifikanztests (T-Test für abhängige Stichproben und Wilcoxon) sowie anhand der einfaktoriellen Varianzanalyse zur Ermittlung des Anteils der aufgeklärten Varianz (η^2) und der Effektstärken (f-Index), wobei der Faktor Zeit und alle relevanten abhängigen Variablen (Skalenwerte zum Gesundheitsverhalten und zur Lebensqualität; Gewicht usw.) einbezogen werden.