



# Evidenztriangulation in der Gesundheitsforschung

## Kombination von experimentellen, deskriptiven und inhaltsanalytischen Ansätzen

Philipp Mayring

Online publiziert: 10. Juli 2017

© Der/die Autor(en) 2017. Dieser Artikel ist eine Open-Access-Publikation.

**Zusammenfassung** Der Beitrag schildert zunächst Probleme eines rein quantitativen Methodenverständnisses. Sodann werden die Besonderheiten der Gesundheitsforschung herausgestellt, die ein komplexes methodisches Vorgehen verlangen. Die heute gängige Vorstellung vom quantitativ-experimentellen Design (randomisierte kontrollierte Studien) als einzig fundierte Vorgehensweise oder als Spitze der Evidenzhierarchie wird kritisiert und das Konzept einer Evidenztriangulation im Sinne eines Mixed-Methods-Ansatzes dagegengestellt. Das Vorgehen wird an einer Beispielstudie verdeutlicht. Zur Evaluation einer Einrichtung zur Beratung und Betreuung Drogenabhängiger werden nacheinander vier methodisch unterschiedliche Bausteine vorgestellt. Ein experimentell orientierter Vorher-nachher-Vergleich mit standardisierter Methodik, qualitative Interviews mit den Klienten und qualitativ-inhaltsanalytischer Auswertung, qualitative und quantitative Dokumentenanalyse sowie ein Selbstevaluationsbaustein wurden hier eingesetzt. Um die unterschiedlichen Ergebnisse zusammenzuführen, wird schließlich ein dialektisches wissenschaftstheoretisches Argumentieren vorgeschlagen.

**Schlüsselwörter** Methoden der empirischen Sozialforschung · Mixed Methods · Qualitative Methoden · Quantitative Methoden · Evaluationsforschung · Evidenzhierarchie · Triangulation

---

Aus Gründen der besseren Lesbarkeit wird auf die gleichzeitige Verwendung weiblicher und männlicher Sprachformen verzichtet. Sämtliche Personenbezeichnungen gelten für beiderlei Geschlecht.

---

P. Mayring (✉)

Institut für Psychologie und Zentrum für Evaluation und Forschungsberatung ZEF, Alpen-Adria  
Universität Klagenfurt

Universitätsstr. 65–67, 9020 Klagenfurt, Österreich

E-Mail: philipp.mayring@aau.at

## Evidence triangulation in health research

The combination of experimental, descriptive and content-analytical approaches

**Abstract** The contribution starts with weaknesses of a strict quantitative methodology in social research. Especially in health research variable connections are so complex that a broader understanding of research methods seems to be necessary. The dominance of quantitative-experimental ongoing in the sense of randomized controlled trials (RCT) at the top of the evidence hierarchy is criticized and a concept of evidence triangulation in the sense of mixed methods is proposed as alternative. An example study shows the logic of this research design. An institution of drug addicts counseling and therapy is analyzed in its effects through an evaluation study, using four different methodological approaches, an experimental pre-post design, qualitative interviews with the clients, and qualitative and document analysis using qualitative content analysis are combined. Each of them is developed in criticizing and supplementing the other approaches. The mixed methods procedure therefor follows a dialectical theory of science.

**Keywords** Methods of social research · Mixed methods · Qualitative methods · Quantitative methods · Evaluation · Hierarchy of evidence · Triangulation

### 1 Das Methodenproblem in den Human- und Sozialwissenschaften

Angemessene empirische Forschungsmethoden sind in den Human- und Sozialwissenschaften Gegenstand heftiger Diskussionen geworden, das zeigt auch der Themenschwerpunkt dieses Heftes. Am drastischsten war hier das Behaupten von Wissenschaftskriegen (science wars; Ross 1996). Die Debatte entspannt sich am Gegensatzpaar quantitativ versus qualitativer Methodologie. An einem rein quantitativ-naturwissenschaftlichen Methodenverständnis werden besonders kritisiert (Flick et al. 2004, Kap. 1):

- Mangelnde Offenheit für neue Aspekte, die im Verlauf einer Studie auftauchen, rigides Festhalten am anfangs bestimmten Methodenkonzept,
- Mangelnde Fähigkeit, komplexe Zusammenhänge zu erfassen, Reduktion der Forschungsfrage auf lineare Zusammenhänge weniger Variablen,
- Mangelnder Problembezug und Anwendungsbezug, Beschränkung auf abstrakte Grundlagenforschung.

In der Sozialpsychologie hat die Kontroverse durch eine prominent (Zeitschrift „Science“) veröffentlichte Studie (Open Science Collaboration 2015) neuen Brennstoff erhalten, die nachgewiesen hat, dass nur ein geringer Teil experimentell-quantitativ ausgerichteter psychologischer Studien durch eine mit den Originalautorinnen und -autoren abgesprochene identische Replizierung zu denselben Ergebnissen führen konnte. Als Grund wird die einseitige, wenig gegenstandsangemessene Methodik angeführt. So hat als Reaktion darauf die Kommission „Qualität psychologischer Forschung“ der Deutsche Gesellschaft für Psychologie DGPs (Fiedler 2016) „ein klares Bekenntnis zum Pluralismus“ (Fiedler 2016, S. 5) in der Forschungsmetho-

dologie formuliert, das Festschreiben von Kriterien guter Forschung in Frage gestellt und den Wettbewerb von Methoden gefordert, ein Plädoyer also für Mixed Methods in einem Forschungsbereich, in dem quantitative Methodik überwiegt. Gerade an dem in der Psychologie so stark herausgestellten quantitativen Experiment ist immer wieder Kritik geäußert worden (Mertens 1975). Dazu gehören ethische Bedenken (Manipulation) und praktische Schwierigkeiten der Umsetzung (Zufallszuweisung in Versuchs- und Kontrollgruppe oft nicht möglich; vgl. nächster Abschnitt).

Wenn aber von prinzipiellen Schwächen des experimentellen Designs auszugehen ist, so muss zu einer angemesseneren Forschungsstrategie gegriffen werden. Hier kommen Mixed-Methods-Ansätze ins Spiel. Der Grundgedanke, der diesen Ansätzen zugrunde liegt, ist die Triangulation (Denzin 1970; Flick 2008). Die Begrifflichkeiten sind hier allerdings in den letzten 20 Jahren recht unterschiedlich gebraucht worden. Bei Triangulation geht es darum, mehrere Ansätze in der Forschungsmethodik (theoretische Rahmung, Forscherpersonen, Design, Erhebungsmethoden, Daten, Auswertungsmethoden) einzusetzen, dabei qualitative und quantitative Ansätze zu verfolgen und dann das Gesamtergebnis als Zusammenschau der verschiedenen Bausteine zu verstehen. Es ergeben sich dabei Möglichkeiten der Kombination (mehrere relativ unabhängige Untersuchungsbausteine) und/oder Integration (unterschiedliche Analyseschritte innerhalb eines übergeordneten Designs) (Mayring 2001). Mit Mixed Methods wird heute der Einsatz qualitativer und quantitativer Analyseschritte verstanden. In den letzten Jahren sind die verschiedenen Möglichkeiten von Mixed Methods ausdifferenziert, dimensioniert und inventarisiert worden (Tashakkorie und Teddlie 1998; Hesse-Biber 2010; Cresswell und Plano Clark 2010), sodass man durchaus davon ausgehen kann, dass ein Methodenmix heute als etabliert und voll anerkannt in der sozialwissenschaftlichen Methodologie etabliert gelten kann. Der Begriff der Triangulation hat gegenüber dem Mixed-Methods-Begriff den Vorteil, dass nicht nur ein Aneinanderreihen unterschiedlicher Teile gefordert wird, sondern ein Zusammenführen der Bausteine zu einem Gesamtergebnis (Flick 2008). Wie dies geschehen kann, wird im letzten Abschnitt dieser Arbeit diskutiert.

Ich möchte im Folgenden auf einen Bereich eingehen, in dem die Angemessenheit der Forschungsmethodik besonders heftig diskutiert wird, der Gesundheitsforschung.

## 2 Methodologische Herausforderungen in der Gesundheitsforschung

Die Erörterung forschungsmethodischer Fragen, so lehrt uns die Methodologie, sollte immer auch gegenstandsbezogen geführt werden. In diesem Abschnitt soll eine Beispielstudie aus dem Bereich der Gesundheitsforschung diskutiert werden; also müssen wir uns zunächst den Spezifikationen des Gesundheitsbereiches zuwenden.

Gesundheitsforschung (Hurrelmann und Razum 2012; Faltermaier 2005) bewegt sich in einem komplexen Feld. Unterschiedliche Professionen wie Ärzteschaft, Krankenpflege, Psychotherapie, Sozialarbeit, Gesundheitspolitik, Gesundheitsökonomie, Krankenhausverwaltung wirken hier zusammen. Die einzelnen Disziplinen der Gesundheitsforschung (Medizin, Psychologie, Soziologie, Politik, Wirtschaftswissenschaft) verfügen über unterschiedliche Forschungsansätze und Wissenschaftsstandards, die nicht immer leicht miteinander vereinbar sind.

Auch ist der Begriff der Gesundheit in der aktuellen Forschungsliteratur ein äußerst komplexer geworden. Unter Gesundheit versteht man heute nicht nur die Abwesenheit von Krankheit, sondern ein allgemeines physisches, psychisches und soziales Wohlbefinden der Person. Gesundheit und Krankheit werden als Kontinuum verstanden, die gesundheitsförderlichen oder -behindernden gesellschaftlichen Rahmenbedingungen werden mit einbezogen (Faltermaier 2005).

Andererseits werden an Gesundheitsforschung höchste Ansprüche der Wissenschaftlichkeit gestellt, da die Folgen von Fehlentscheidungen in der Regel drastisch sind, mit menschlichem Leiden verbunden sind. Deshalb steht dieser Bereich zum großen Teil unter staatlicher Aufsicht und wird vorwiegend aus öffentlichen Geldern finanziert. Medikamente sind einem äußerst aufwendigen Zulassungsverfahren unterworfen, das nach genauen methodischen, im Sozialgesetzbuch verankerten Vorgaben abzulaufen hat. Heilverfahren werden immer mehr bindenden Leitlinien unterworfen. So steht das Gesundheitssystem laufend in einem gesellschaftlichen Diskussionsprozess, in dem die Wissenschaftlichkeit, d. h. auch die forschungsmethodische Fundiertheit eine zentrale Rolle spielt.

Dieser immer stärker werdenden Verwissenschaftlichung und auch Technisierung wird jedoch auch kritisiert und durch ein stärker am Laienwissen orientiertes und weniger systematisch evaluiertes System der Alternativmedizin kontrastiert. Wie soll sich gesundheitspsychologische Forschung in diesem komplexen Feld verhalten? Es ist das Ziel dieser Arbeit zu zeigen, dass dieser Komplexität nur durch Mixed-Methods-Ansätze gerecht zu werden ist.

Bereits in den 1970er-Jahren wurde von dem schottischen Epidemiologen Archibald Cochrane eine strenge wissenschaftliche Fundierung gesundheitsbezogener Forschung gefordert, vor allem was den Wirkungsnachweis von Medikamenten und medizinischen, auch psychotherapeutischen Interventionen anbelangt (zum Folgenden s. z. B. Lauterbach und Schrappe 2004). Seit 1992 werden solche Wirkungsnachweise systematisch gesammelt, nach ihrer methodischen Fundiertheit ausgewertet und öffentlich zur Verfügung gestellt ([www.thecochranelibrary.com](http://www.thecochranelibrary.com)). Als Ideal (Gold-Standard) gilt bis heute die randomisierte kontrollierte experimentelle Studie („Randomized Controlled Trial“, RCT), da sie als einzige vorgeblich Kausalaussagen zulasse. Dieser Grundgedanke experimenteller Evidenzbasierung hat in die modernen Gesundheitssysteme breit Eingang gefunden. Gerade nach dem Contergan-Skandal (ein Beruhigungsmittel, das, in der Schwangerschaft eingenommen, reihenweise zu Fehlbildungen beim Neugeborenen geführt hatte) in der Bundesrepublik Deutschland wurde der RCT-Ansatz zur Zulassung von Medikamenten gesetzlich verankert.

Im Folgenden hat diese Auffassung vom Experiment als Goldstandard zu einer Bewertung von methodischen Ansätzen in der Gesundheitsforschung geführt, die als Evidenzhierarchie konzipiert ist (Lauterbach und Schrappe 2004):

- Einzelfallberichte, deskriptive Darstellungen, Experteneinschätzungen als niedrigste, das heißt am schlechtesten wissenschaftlich abgesicherte Evidenzstufe (V)
- Systematische Fallserien (IV)
- Retrospektive vergleichende Studien (III)
- Prospektive vergleichende Kohortenstudien (IIb)

- Systematische Übersichtsarbeiten zu vergleichenden Kohortenstudien (IIa)
- Randomisierte klinische Studien (Ib)
- Systematische Übersichtsarbeiten zu randomisierten klinischen Studien (Ia)

Das Konzept der Evidenzhierarchie bewertet also Metaanalysen mit mehreren randomisierten quantitativ-experimentellen Studien am höchsten. Nur wenn solche nicht vorliegen, kann auf niedrigere Studien ausgewichen werden, allerdings mit nur eingeschränktem Gültigkeitsanspruch.

Dabei werden verschiedene Schwachpunkte des experimentellen Designs übersehen:

- Experimentelle Untersuchungsanlagen implizieren immer eine Manipulation, einen Eingriff in die soziale Realität, nämlich die Herstellung der experimentellen Bedingung (unabhängige Variable) im Unterschied zur Kontrollgruppe. Placebo-Gruppen, gerade im Gesundheitsbereich dringend gefordert, stellen ebenfalls eine Manipulation dar (hier wird eine medizinische Behandlung vorgegaukelt). Das kann zu Unsicherheit oder Misstrauen seitens der Untersuchungsteilnehmerinnen und -teilnehmern führen.
- Darüber hinaus fordert das experimentelle Design in aller Regel, dass das Untersuchungskonzept den Untersuchungsteilnehmern verheimlicht wird (die Information könnte Erwartungseffekte auslösen und so die interne Validität schwächen). Solche Eingriffe sind oft ethisch bedenklich. Die Intervention erfolgt mit noch nicht gesicherten Mitteln (sie sollen ja erst überprüft werden). Oft stellt für die Betroffenen im Gesundheitsbereich die Zuordnung zur Kontroll- oder Placebo-Gruppe ein Vorenthalten von vielleicht wirksamen Behandlungen dar und wird deshalb von möglichen Untersuchungsteilnehmern abgelehnt. Dadurch kann die Stichprobe verzerrt werden, da nun nur noch weniger kritische Fälle in dieser Gruppe verbleiben.
- Die Zuordnung zu Experimental-, Placebo- oder Kontrollgruppe sollte per Zufall erfolgen (Randomisierung). In vielen Fällen (wie in dem unten angeführten Beispiel) ist dies jedoch praktisch gar nicht möglich, weshalb dann oft Kontrollgruppen konstruiert werden (Quasi-Experiment), was aber die Aussagekraft erheblich einschränkt.
- Das Experiment liefert nur eine einfache Ja-Nein-Aussage: Ist ein Effekt der Intervention gegenüber der Kontrollgruppe nachweisbar oder nicht? Pawson und Tilley (1997) bezeichnen dies als sukzessionistische Logik, da der Effekt nur als Folge der Intervention nachgewiesen wird. Den eigentlichen Mechanismus, wie es zum gemessenen Effekt kommt, auch die Frage, wie stark der Effekt ist oder von welchen Rahmenbedingungen der Effekt abhängt, erfahren wir in der Regel nicht.
- Auch bleiben die Aussagen in der Regel auf ein zentrales Effektkriterium (abhängige Variable) beschränkt. Komplexere Wirkzusammenhänge, vermittelnde Einflussgrößen, Gruppenunterschiede werden nicht untersucht.

Trotz dieser doch erheblichen Einschränkungen genießt das Experiment, oft als einziger Forschungsansatz zur Kausalanalyse bezeichnet, in der Gesundheitsforschung diesen hohen Rang. Mixed Methods könnte einen Weg darstellen, diese Einseitigkeiten und Schwächen zu überwinden.

Warum also nicht diesen Ansatz auch in der Evaluation von Maßnahmen im Gesundheitsbereich anwenden? Gerade dort, wo eine klare Evidenz, verlässliche Ergebnisse so wichtig sind (da die Folgen fehlerhafter Forschungsergebnisse zu Krankheit und Leiden führen können), erscheint eine Argumentation auf breiterer Grundlage vorteilhaft. Auch das Konzept der Evidenzhierarchie (s. oben) ist im Angesicht der fundamentalen Kritik des Experiments als „Goldstandard“ unzureichend. Wir möchten deshalb als Alternative das Konzept der Evidenztriangulation vorschlagen (Mayring 2009). Es geht darum, zur Beantwortung einer Forschungsfrage verschiedene methodische Ansätze zu verfolgen und die Ergebnisse zu einer Gesamtevidenz zusammenzubringen. Besonders auf der Ebene des Forschungsdesigns (z. B. quasi-experimentelle zusammen mit einzelfallanalytischen Vorgehensweisen) erscheint dies vielversprechend.

Ein solches Konzept der Evidenztriangulation erscheint uns natürlich auch außerhalb der Evaluation im Gesundheitsbereich, auch in anderen sozialwissenschaftlichen Forschungsbereichen sinnvoll.

Dass ein Mixed-Methods-Ansatz hier eine wesentliche methodische Bereicherung darstellen kann, soll nun an einem Beispiel aus der Gesundheitsforschung gezeigt werden.

### 3 Beispielstudie aus der Gesundheitsforschung: Evaluation einer Drogenambulanz

Die Substitutionsbehandlung chronisch opiatabhängiger Personen, also die kontrollierte Abgabe von synthetischen Opioiden, gilt heute als anerkannte Therapie (Fischer und Kayer 2006). Laut Weltgesundheitsorganisation (WHO 2004) reduziert sie den illegalen Drogenkonsum, senkt Morbiditäts- und Mortalitätsraten, stabilisiert das körperliche und psychische Wohlbefinden, minimiert delinquentes Verhalten und unterstützt so die Reintegration der betroffenen Personen in die Gesellschaft.

Dennoch ist die Substitutvergabe an Drogenabhängige in der Gesellschaft immer wieder umstritten. Nicht zuletzt deshalb wurden Behandlungskonzepte entwickelt, die die Substitutionsbehandlung in ein umfassendes Konzept medizinischer Versorgung, sozialarbeiterischer Beratung und psychotherapeutischer Behandlung einbetten. Eine solche multiprofessionelle Drogenambulanz in ihrer Wirkung zu evaluieren, war das Ziel der hier vorgestellten Studie (Brunner et al. 2009). Die untersuchte Einrichtung bestand aus einem Team von zwei Psychiatriefachärzten, zwei Allgemeinärzten, einem Psychotherapeuten, zwei Krankenschwestern und zwei Sozialarbeiterinnen, angesiedelt in einem niederschwellig zugänglichen Bereich.

#### 3.1 Baustein 1: Quantitativ-experimentell orientierte Untersuchungsanlage

Der klassische Absatz der Evaluationsforschung, auch im Sinne des oben angeführten Konzepts der Evidenzbasiertheit, ist das experimentelle Design. Nun ist, wie so oft im Gesundheitsbereich, der Einsatz einer Kontrollgruppe oder gar Placebo-Gruppe, wie es im Sinne des RCT-Ansatzes wäre, hier nicht möglich. Wir müssten einer Gruppe opiatabhängiger Personen, die in der Drogenambulanz Hilfe suchen,

die Behandlung verweigern, was ethisch nicht vertretbar wäre. Auch eine Placebo-Bedingung würde hier sofort durchschaubar sein. Nicht einmal eine Wartekontrollgruppe wäre möglich, also die Vertröstung einer Gruppe von Klienten um mehrere Monate, während sie als Kontrollgruppe für die Behandlungsgruppe fungieren.

### 3.1.1 *Methodischer Ansatz: Gesundheits-Test im Vorher-Nachher-Design*

Deshalb kommt hier zunächst nur ein einfaches Vorher-Nachher-Design in Frage. Erhoben werden sollen Daten von Neuzugängen zur Drogenambulanz und mit einer zweiten Erhebung ca. vier Monaten später verglichen werden. Eine reine Zufallsauswahl ist ebenfalls nicht möglich. Um aber eine breite Verteilung von Personenmerkmalen zu erreichen, wird eine konsekutive Stichprobengewinnung gewählt. Über einen Zeitraum von sechs Monaten werden alle Neuzugänge in die Stichprobe einbezogen. Wir konnten erreichen, dass 23 der 25 Neuzugänge in dieser Zeit zur Studienteilnahme zu beiden Erhebungszeitpunkten bereit waren. Das Vorher-Nachher-Design erfordert, dass ein klarer Indikator aus den Zielen des Programms abgeleitet und operationalisiert wird. Die Ziele der Drogenambulanz, so die Vorgaben, sind eine Stabilisierung der Personen, eine breite Gesundheitsförderung, die körperliche, psychische und soziale Aspekte einschließen soll. Wir haben als Instrumente bewährte standardisierte Skalen ausgewählt, die im Gesundheitsbereich weit verbreitet sind: den „World Health Organisation Quality of Life Fragebogen“ in seiner Kurzversion (WHOQOL-BREF; Angermeyer et al. 2000), der die Subskalen physisches Wohlbefinden (z. B. Schmerz, Unbehagen, Mobilität), psychisches Wohlbefinden (z. B. Selbstachtung, Körperwahrnehmung), Umwelt (z. B. finanzielle Lage, häusliche Umwelt), soziale Beziehungen (z. B. Sexualität, soziale Unterstützung) und eine globale Einschätzung der subjektiven Lebensqualität umfasst, sowie die Beschwerdenliste B-L (von Zerssen 1976). Dies ist ein Instrument, das als Testbogen den Personen zum Ankreuzen vorgegebener Antwortalternativen vorgelegt wird.

### 3.1.2 *Auswertung und Ergebnisse*

Die Ergebnisse wurden nach den Anweisungen des Testbogens addiert, Mittelwerte gebildet und sodann Vorher-Nachher-Mittelwertsvergleiche statistisch mittels t-Test auf Signifikanz überprüft. Es zeigten sich folgende Resultate (s. Tab. 1):

Leichte Verbesserungen in der subjektiven Lebensqualität und eine Abnahme der Beschwerden der Klienten der Drogenambulanz sind also bereits nach wenigen Monaten feststellbar. Die Ergebnisse sind umso bemerkenswerter, als dass es sich hier um eine kleine Stichprobe handelt, in der Mittelwertsunterschiede sehr groß sein müssen, um signifikant zu werden. Deshalb erscheint uns der Trend im physischen Befinden auch interpretierbar. Es verwundert nicht, dass in den Bereichen der Umweltbedingungen (z. B. Wohnen) und der sozialen Beziehungen sich in der kurzen Zeit noch keine Veränderungen feststellen lassen. Auf lange Sicht sind aber gerade auch hier Verbesserungen zu erwarten, zumal das Interventionskonzept eine sozialarbeiterische Komponente beinhaltet, die sich gerade auf diese Dimensionen beziehen

**Tab. 1** Vorher-Nachher-Vergleich von Lebensqualität (WHOQOL) und Beschwerden (B-L),  $n = 23$ 

Befindensdimension	Vorher-nachher-Vergleich	Statistik
Physisch (WHOQOL)	Tendenzielle Verbesserung auf Körperebene	$t = -2057, p = 0,053$
Psychisch (WHOQOL)	Signifikante Verbesserung im psychischen Wohlbefinden	$t = -2288, p = 0,032$
Umwelt (WHOQOL)	Keine signifikante Veränderung	$t = -1005, p = 0,326$
Sozial (WHOQOL)	Keine signifikante Veränderung	$t = +1134, p = 0,269$
Global (WHOQOL)	Signifikante Verbesserung der subjektiven Lebensqualität	$t = -2336, p = 0,029$
Befinden (B-L)	Signifikante Verminderung der Beschwerden	$t = +2319, p = 0,030$

soll. Die Ergebnisse lassen also insgesamt zunächst eine eher positive Evaluation zu.

### 3.1.3 Limitationen

So positiv das zunächst klingt, sind jedoch prinzipielle methodische Einschränkungen anzumerken. Das Fehlen von Kontrollgruppen wiegt am stärksten. Bortz und Döring (2006) weisen darauf deutlich hin. Zwischen der ersten und zweiten Messung können Faktoren gewirkt haben, die mit der zu evaluierenden Intervention nichts oder nur indirekt zu tun haben. Allein die Tatsache, dass ein Interventionsangebot existiert, könnte unabhängig von den Konzepten der Maßnahme Verbesserungen bedingen. Bortz und Döring (2006) verdeutlichen dies mit ihrem „Kurschattenbeispiel“: Ein positiver Gesundheitseffekt im Vorher-Nachher-Vergleich einer speziellen Kurmaßnahme (radonhaltiger Heilstollen) kann erklärbar sein als allgemeiner Erholungseffekt (keine Arbeit) oder auch durch angenehme soziale Begegnungen während des Kuraufenthaltes. Es könnten auch andere Ereignisse innerhalb der Vorher-Nachher-Zeitspanne aufgetreten sein (z. B. positiver Wetterumschwung), die den Effekt bedingt haben. Im experimentellen Design (Huber 2005) wird all dies unter dem Begriff der „internen Validität“ diskutiert, also der Frage, ob der beobachtete Effekt ausschließlich durch die untersuchte Intervention zustande gekommen ist. Und die methodische Konsequenz muss klar sein: Wenn eine Alternativerklärung für das Zustandekommen eines Effektes nicht auszuschließen ist, so ist der Kausalzusammenhang „Intervention  $\rightarrow$  Effekt“ nicht belegbar, eine positive Evaluation allein aufgrund der experimentellen Ergebnisse nicht ableitbar.

Eine weitere Einschränkung der Schlussfolgerungen aus diesem Evaluationsbaustein besteht darin, dass die abhängige Variable, also der untersuchte Effekt in Lebensqualität und Beschwerden, quantitativ mit einem standardisierten Instrument erhoben wurde. Die Spezifika der Lebenssituation der einzelnen Klienten können nicht berücksichtigt werden. Es kann auch nicht tiefgehend nach Begründungen für die Befindensveränderungen gefragt werden, wie dies nur mit qualitativen Erhebungsinstrumenten möglich ist. Auch dies ist eine Frage der Validität, hier der Problematik, ob die verwendeten Instrumente (WHOQOL und B-L) wirklich das messen können, was mit der Intervention erreicht werden soll, ob nicht noch andere Dimensionen von Lebensqualität und Beschwerden wichtig wären.

## 3.2 Baustein 2: Qualitative Interviews zur Zufriedenheits-Selbsteinschätzung

Eine Konsequenz aus der Problematik mangelnder interner Validität experimenteller Untersuchungsanordnungen ist, auf die subjektive Ebene auszuweichen und nach wahrgenommener Kausalität zu fragen. In der Gesundheitsforschung wird heute immer mehr gefordert, die Patientenzufriedenheit und nicht nur den objektiv beobachtbaren Gesundheitseffekt in die Evidenzerhebung einzubeziehen (Applebaum et al. 2004). In vielen Untersuchungen hat sich herausgestellt, dass die Gesundheits-selbsteinschätzung sehr hoch mit einem objektiveren, kriteriengeleiteteren Arzturteil übereinstimmt, Personen also sehr gut selbst einschätzen können, wie gut es ihnen geht und welche Effekte bestimmte Maßnahmen auf ihr gesundheitliches Befinden haben. Neben einfachen Zufriedenheitsskalen sind hier auch qualitative Erhebungsinstrumente erprobt.

### 3.2.1 *Methodischer Ansatz*

Es soll also eine Zufriedenheitserhebung bei den Klienten vorgenommen werden. In der Regel wird per standardisierter Befragung erhoben, ob die Zielgruppe der Maßnahme, die zu evaluieren Interventionen, unterteilt in verschiedene Dimensionen, positiv oder negativ einschätzt. Da geschlossene Fragebögen diversen Fehlerquellen unterliegen (zum Beispiel der sozialen Erwünschtheit), haben wir uns für teilstrukturierte qualitative Interviews (Witzel 2000) entschieden. Gleichzeitig kann in einer Interviewsituation auch sehr gut nach den für eine Evaluation wichtigen Verbesserungsvorschlägen bezüglich der Konzeption der Drogenambulanz offen gefragt werden. Für das halb-strukturierte Interview wurde ein Leitfaden ausgearbeitet und pilotgetestet, der folgende Fragenkomplexe umfasste:

- Subjektive Zufriedenheitseinschätzung der Drogenambulanz;
- Beanspruchte Maßnahmen in der Drogenambulanz und auch außerhalb;
- Subjektiv erlebte Veränderungen in einzelnen Lebensbereichen seit Behandlungsbeginn; und
- Verbesserungsvorschläge in Bezug auf Behandlungsmaßnahmen und Rahmenbedingungen.

Einschlusskriterium war in diesem Baustein, dass die zu befragenden Personen seit mindestens einem Jahr an einer Substitutionsbehandlung durch die Drogenambulanz teilgenommen haben sollten. Auch hier wurde eine konsekutive Stichprobengewinnung praktiziert; es wurden also über einen festen Erhebungszeitpunkt alle Klienten mit Einschlusskriterium um ein Interview gebeten. Hier gab es allerdings eine hohe Quote von Ablehnungen. Aufgrund des hohen Aufwands (in Erhebung und Auswertung) konnte nur eine kleine Stichprobe ( $n = 14$ ) realisiert werden. Die Gespräche wurden auf Tonband aufgenommen und komplett wörtlich transkribiert.

### 3.2.2 *Auswertung und Ergebnisse*

Die Transkripte wurden von den Interviewern selbst qualitativ-inhaltsanalytisch ausgewertet (induktive Kategorienentwicklung und deduktive Kategorienanwendung

**Tab. 2** Häufigkeit deduktiver inhaltsanalytischer Kategorien zur Behandlungsart (14 Interviewpartner)

Kategorie	<i>n</i>	Zufrieden	Teils/teils
K1: Substitutionsbehandlung	14	14	0
K2: Zusätzliche medizinische Versorgung	6	6	0
K3: Sozialarbeiterische Betreuung	12	11	1
K4: Psychotherapeutische Angebote	6	6	0
K5: Gesamtzufriedenheit mit dem Angebot	14	12	2

**Tab. 3** Häufigkeit induktiver Kategorien zur Zufriedenheit und Unzufriedenheit (14 Personen)

<i>Zufriedenheitskategorien</i>	<i>n</i>
Z1: Gute Betreuung vor Ort	7
Z2: Diskretion aller Mitarbeiter, Vertrauen	7
Z3: Offener Austausch/Rücksichtnahme	5
Z4: Ausgewogenes Programm	5
Z5: Klare Einhaltung der Vorschriften	2
<i>Unzufriedenheitskategorien</i>	–
U1: Zu lange Wartezeiten	8
U2: Gestaltung des Warteraumes	8
U3: Zu kurze Öffnungszeiten	7
U4: Zu rigide, wenig individuelle Substitutbehandlung	5
U5: Zu schnelle Aufnahme ins Substitutionsprogramm	5
U6: Schwierige Erreichbarkeit von Mitarbeiter	4
U6: Szenekontakt im Warteraum	2

nach Mayring 2015). Zunächst wurde ausgewertet, welche Maßnahmen wahrgenommen wurden und wie zufrieden die Klienten damit über die gesamte Behandlungszeit waren. Hier wurden deduktive Kategorienanwendungen eingesetzt, für die Maßnahmen der Katalog der Angebote (vier Kategorien) und für die Zufriedenheit eine einfache Skalierung in zufrieden, teils/teils und unzufrieden (drei Kategorien). Für die Kategorien, besonders das zweite Kategoriensystem, wurde ein Kodierleitfaden theoriegeleitet festgelegt, der für jede einzelne Kategorie Definitionen, typische Textbeispiele und Abgrenzungsregeln umfasst. So wurde für hohe Zufriedenheit beispielsweise festgelegt, dass subjektiv positive Punkte der Behandlung von der Person genannt werden, keine negativen Punkte oder Kritik auftauchen und dass das Gesamturteil positiv ausfällt (Ankerbeispiel: „Von meiner Seite gibt es nichts, was nicht gut läuft.“ Pbn 12). In Tab. 2 sind die deduktiven Kategorien aufgelistet, zusammen mit den Häufigkeiten der Personen (von 14), bei denen sie zugeordnet werden konnten sowie der Anzahl von Personen, die mit der Maßnahme (sehr) zufrieden waren (Unzufriedenheit konnte nirgends gefunden werden).

Zusätzlich wurde mittels induktiver Kategorienbildung ausgewertet, womit konkret die Interviewpartner zufrieden oder unzufrieden waren. Tabelle 3 gibt einen Überblick.

Es wird klar, dass durch die subjektbezogene, qualitativ orientierte Vorgehensweise viel spezifischere, zum Teil auch überraschende Ergebnisse zutage gefördert werden. So war es für uns erstaunlich, wie differenziert sich die Klienten zu den

einzelnen Behandlungen äußern. Sie verstehen sich nicht nur als zu versorgende Klienten, sondern versetzen sich in die Lage des Arztes/der Ärztin. Auch wird deutlich, dass diese offene Erhebung weniger dem in der Umfrageforschung so oft gefundenen positiven Bias (Zustimmungstendenz, vgl. Kallus 2016) unterliegt, denn beim Erfragen konkreter Faktoren überwiegen plötzlich die Unzufriedenheitsnennungen (bei aller Vorsicht vor einem einfachen Addieren der zum Teil ungleichgewichtigen Faktoren). Das Gesamtbild der Evaluation bleibt jedoch im positiven Bereich.

### 3.2.3 Limitationen

Klar ist aber auch, dass dieser Baustein nur die subjektive Sichtweise der Klienten darstellt. Das Team der Drogenambulanz hat sich beispielsweise dem Wunsch nach schönerer Gestaltung des Wartebereiches (bequemere Sitzgelegenheiten, Zeitschriften) sofort widersetzt mit dem Hinweis auf den engen finanziellen Rahmen der Maßnahme. Auch ist es die Frage, ob die Klienten die ärztlichen Entscheidungen (Aufnahme ins Substitutionsprogramm, konkrete Behandlung) selbst umfassend beurteilen können. Deshalb schien es uns wichtig, in der Evaluation objektivere Kriterien und auch die fachärztliche Einschätzung des Behandlungserfolges mit einzu beziehen.

## 3.3 Baustein 3: Quantitativ-qualitativ-inhaltsanalytische Dokumentenanalyse mit Arzturteil

Mit einer Dokumentenanalyse setzen wir hier daran, dass im Gesundheitsbereich unabhängig von besonderen Evaluationen laufend Behandlungsverläufe dokumentiert werden. Seit Eröffnung der Ambulanz wurde hier ein computerunterstütztes Dokumentationsprogramm zum Festhalten der Personenmerkmale als Ausgangspunkt des medizinischen, psychischen und sozialen Verlaufs der Klienten eingesetzt. Das Evaluationsteam hat in diese Datei (146 Klienten) unter Aufsicht Einsicht bekommen.

### 3.3.1 Methodischer Ansatz

Die Daten der 146 Personen wurden auf zweierlei Arten analysiert: zum einen wurden quantitativ Personenmerkmale zur Deskription ausgewertet. Dies sollte einen Überblick über die objektive Lage der Klientel, was den medizinischen Status und die soziale Situation betrifft, geben.

Zum anderen sollten die Behandlungsverläufe der Personen analysiert werden und zusammen mit dem Leitungsteam der Drogenambulanz daraufhin eingeschätzt werden, ob ein Behandlungserfolg festzustellen sei. Dieser zweite, aufwendigere Schritt konnte nur an einer Teilmenge der Klienten vorgenommen werden. Dafür haben wir aus den 146 Personen eine Zufallsstichprobe von 50 Personen gezogen.

Beide Auswertungsschritte sind inhaltsanalytischer Art, da sie mit Kategoriensystemen arbeiten; der erste Ansatz ist ein quantitativ-inhaltsanalytisches Vorgehen, der zweite Ansatz ein qualitativ-inhaltsanalytisches (deduktive Kategorienanwendung

**Tab. 4** Quantitativ-inhaltsanalytische Dokumentenanalyse  
( $n = 146$ )

Variable	Ausprägung	Prozent (von 146 Personen)
Familienstand	F1: ledig	85
	F2: verheiratet	8
	F3: geschieden	5
	F4: unbekannt	2
Kinder	K1: keine	66
	K2: eins oder mehrere	31
	K3: unbekannt	3
Wohnsituation	W1: selbständig gesichert	94
	W2: Institution	1
	W3: obdachlos	5
Schulbildung	S1: nur Pflichtschule	65
	S2: Lehre	24
	S3: Fachschule	3
	S4: Höhere Schule	4
	S5: Hochschulabschluss	1
	S6: Unbekannt	3
Berufstätigkeit	B1: voll erwerbstätig	24
	B2: Teilzeit	1
	B3: geringfügig erwerbstätig	1
	B4: arbeitslos/ arbeitsunfähig	71
	B5: unbekannt	3

mittels eines Kodierleitfadens), da hier interpretative Einschätzungen vorgenommen werden mussten.

### 3.3.2 Auswertung und Ergebnisse

Zunächst werden ausgewählte quantitative Ergebnisse dargestellt. Tab. 4 gibt einen Überblick über die Häufigkeiten der inhaltsanalytischen Kategorien in Prozent.

Aus dieser Aufstellung wird schnell ersichtlich, wo die zentralen Probleme der Klientel liegen: Der überwiegende Teil ist nicht erwerbstätig und verfügt über eine nur niedrige Schulbildung. Dies war dem Team der Drogenambulanz in dieser Deutlichkeit nicht klar. Konsequenzen, vor allem für den sozialarbeiterischen Teil der Betreuung sollten gezogen werden. Denn danach darf die rein medizinische Behandlung nicht alleine stehen, um Gesundheit und Wohlbefinden nachhaltig zu stabilisieren.

Für die qualitativ-inhaltsanalytische Einschätzung des Behandlungserfolges (Evaluationsteam plus Leitung der Drogenambulanz) wurden zur differenzierteren Analyse Veränderungsvariablen gebildet und als Kategorien mittels Kodierleitfaden genau definiert sowie pilotgetestet. Die Einschätzungen wurden für verschiedene Lebens-

**Tab. 5** Häufigkeit deduktiv-inhaltsanalytischer Verlaufskategorien in unterschiedlichen Lebensbereichen ( $n = 50$ )

<i>Veränderungskategorie Berufssituation</i>	<i>Prozent (von 50 Personen)</i>
E1: Konstant positiver Verlauf	12
E2: Verbesserung im Verlauf	34
E3: Konstanz mittlerer Status/keine Veränderung	4
E4: Schwankend chaotischer Verlauf	6
E5: Verschlechterung im Verlauf	6
E6: Konstant negativer Verlauf	38
E7: Unentscheidbar	0
<i>Veränderungskategorie psycho-sozial</i>	<i>Prozent (von 50 Personen)</i>
E1: Konstant positiver Verlauf	42
E2: Verbesserung im Verlauf	12
E3: Konstanz mittlerer Status/keine Veränderung	20
E4: Schwankend chaotischer Verlauf	0
E5: Verschlechterung im Verlauf	0
E6: Konstant negativer Verlauf	14
E7: Unentscheidbar	12
<i>Veränderungskategorie Wohlbefinden</i>	<i>Prozent (von 50 Personen)</i>
E1: Konstant positiver Verlauf	34
E2: Verbesserung im Verlauf	18
E3: Konstanz mittlerer Status/keine Veränderung	26
E4: Schwankend chaotischer Verlauf	6
E5: Verschlechterung im Verlauf	2
E6: Konstant negativer Verlauf	12
E7: Unentscheidbar	2
<i>Veränderungskategorie Drogenkonsum</i>	<i>Prozent (von 50 Personen)</i>
E1: Konstant positiver Verlauf	22
E2: Verbesserung im Verlauf	46
E3: Konstanz mittlerer Status/keine Veränderung	6
E4: Schwankend chaotischer Verlauf	10
E5: Verschlechterung im Verlauf	2
E6: Konstant negativer Verlauf	14
E7: Unentscheidbar	0

bereiche vorgenommen (Wohnsituation, Berufssituation, psycho-soziale Situation, psychisches Wohlbefinden, körperliche Befindlichkeit, Drogenkonsum). Tabelle 5 zeigt ausgewählte Ergebnisse.

Hier wird ein sehr differenziertes Bild vom Behandlungsverlauf von 50 zufällig ausgewählten Klienten deutlich. Bei der Hälfte der Personen hat sich die berufliche Situation während der Kontaktzeit mit der Drogenambulanz sogar verschlechtert oder ist konstant schlecht oder chaotisch schwankend geblieben. Im psychosozialen Bereich ist dieser Teil mit 14 % erheblich kleiner; bei vielen Personen kam es hier zu Verbesserungen. Auch das psychische Wohlbefinden hat sich bei einem Großteil verbessert. Allerdings konnte der Drogenkonsum, entscheidendes Merkmal des Behandlungskonzeptes, nur bei 46 % der Personen vermindert werden; andererseits

gibt es hier direkte Verschlechterungen oder einen konstant negativen Verlauf nur bei 16 %. Ob es insgesamt als positiv zu werten ist, dass Interventionserfolge bei etwa der Hälfte der Personen zu konstatieren sind, bleibt der Interpretation vorbehalten.

### 3.3.3 Limitationen

Die Ergebnisse dieses Bausteins fußen auf einer stärker objektiven Sichtweise im Gegensatz zur subjektiven Befindlichkeit der Klienten und beziehen das Expertenurteil in der Einschätzung des Interventionserfolges mit heran. Allerdings liegen hier auch die kritischen Punkte. Zum einen ist bei Dokumentenanalysen immer zu reflektieren, unter welchen Bedingungen und mit welchem Zweck die Dokumentation angelegt wurde. Vielfältige Verzerrungen sind hier denkbar. Baur (2009) hat hier die Probleme der Datenqualität von Dokumenten herausgearbeitet.

Zum anderen ist auch das hinzugezogene Expertenurteil des Behandlungserfolges Fehlerquellen unterworfen, da hier die Perspektive des behandelnden Arztes überwiegt. Die Tatsache, dass das Urteil insgesamt eher negativer ausgefallen ist als die Selbsteinschätzung der Klienten (Baustein 2) zeigt zwar, dass die Experteneinschätzung keinem prinzipiellen positiven Bias unterworfen scheint. Es könnte sich aber auch um eine negative Urteilstendenz handeln, wenn die Drogenambulanzleitung den eigenen Erfolgen besonders kritisch gegenüber gewesen sein könnte. Allerdings wurden hier inhaltsanalytische Regeln formuliert, um diesen Bias einschränken zu können.

Eine prinzipielle Beschränkung aller bisher erarbeiteten Evaluationsergebnisse besteht darin, dass das Team der zu evaluierenden Institution nicht einbezogen wurde, die Daten sozusagen über ihre Köpfe hinweg erhoben wurden. Deshalb erscheint hier ein stärker partizipatives Evaluationselement vonnöten.

## 3.4 Baustein 4: Qualitätszirkel

Qualitätszirkel stellen Ansätze eines partizipativen Evaluationskonzeptes (Patton 2015; Beywl 1988) insofern dar, als dass hier im Zusammenhang einer Institution Teambesprechungen organisiert werden, in denen die Beteiligten selbst die Arbeitsabläufe und Arbeitsergebnisse evaluieren und selbst qualitätsverbessernde Schlussfolgerungen daraus ziehen. Gerade im Gesundheitsbereich wird dieses Instrument immer wieder empfohlen (Bundeszentrale für gesundheitliche Aufklärung BZgA 2005).

### 3.4.1 Methodischer Ansatz

Das Vorgehen ist an den Ansätzen von Selbstevaluation (König 2007) orientiert. Da die Beteiligten die Träger der Umsetzung von Evaluationsempfehlungen sind, sollen sie selbst auch Träger der Evaluation sein und im Sinne einer Selbstreflexion eigener beruflicher Tätigkeit Stärken und Schwächen finden und auch die Bewertungsmaßstäbe selbst setzen. Dazu sollte das Team in die Methoden der Evaluation eingearbeitet werden. Das Gesundheitszirkelkonzept nimmt das auf, indem regelmäßige Treffen (in unserem Falle sechs halbtägige Sitzungen) eines Qualitätsarbeits-

kreises (in unserem Falle des gesamten Teams aus zwei Psychiatriefachärzten, zwei Allgemeinärzten, einem Psychotherapeuten, zwei Krankenschwestern und zwei Sozialarbeiterinnen) veranstaltet wurden. Zwei Evaluatoren haben den Prozess begleitet und moderiert. Die selbstgesetzten Ziele des Arbeitskreises waren

- die methodische Kompetenzerweiterung im Sinne regelmäßiger Qualitätsüberprüfungen;
- das Sammeln und Diskutieren von Stärken und Schwächen in der eigenen Berufsarbeit;
- die Diskussion der Ziele der Institution, die Entwicklung eines Leitbildes; und
- das Aufstellen eines Handlungsplanes mit Maßnahmen der Qualitätsverbesserung.

Zu Beginn wurden die zentralen Begriffe Evaluation und Qualität im Gesundheitsbereich im Gespräch geklärt und das Konzept der Leitbildarbeit vorgestellt: Identität der Beteiligten (Wer sind wir?), Ziele (Was wollen wir?) und Visionen (Wohin soll es gehen?) wurden diskutiert. Als Input wurden der gesetzliche Auftrag, die bisherige schriftlich fixierte Selbstdefinition der Drogenambulanz sowie die vorgegebenen Aufgaben der Institution analysiert. Reflexionen zu der Identität der Mitarbeiter der Drogenambulanz, den Arbeitsbedingungen, den Zielen und Visionen wurden angeregt. Dabei kamen unterschiedliche Moderationstechniken (z. B. Erwartungsabfrage, Themensammlungen mittels Brainstorming, Problem-Analyse-Schema; vgl. Seifert 1992) zum Einsatz. Dieses methodische Vorgehen bildete die Grundlage für die Entwicklung eines Leitbildes. Für die Vermittlung methodischer Kompetenzen wurde den Mitarbeitern eine Materialienmappe, basierend auf dem Konzept der Selbstevaluation (König 2007), zusammengestellt. Eine Sitzung widmete sich der Weitervermittlung der wichtigsten Eckpunkte mittels Vortrag und gemeinsamer Diskussion.

Die gemeinsamen Sitzungen wurden schriftlich festgehalten. Zusätzlich wurden die Ergebnisse zur Qualitätszirkelarbeit – zumeist festgehalten auf Moderationskarten – fotoprotokolliert und im Anschluss nachbereitet. Das Evaluationsteam hat die Diskussionen rund um die Leitbildentwicklung moderiert, die Formulierungen aufgenommen und dem Team in der nächsten Sitzung zurückgemeldet. Gemeinsam wurden die einzelnen Textbausteine überarbeitet. Bei unklaren Punkten und Formulierungsschwierigkeiten fungierte das Evaluationsteam als Hilfe; eine weitere Aufgabe stellte die endgültige Verschriftlichung des Leitbildtextes dar. Dabei war auch das Aufstellen eines Organigramms ein wichtiger Teilbereich und hat wertvolle Diskussionen über die interdisziplinäre Zusammenarbeit im Team, über die Leitungsfunktion sowie die institutionelle Einbettung der Drogenambulanz angeregt.

### 3.4.2 Ergebnisse

Das entwickelte Leitbild sowie die selbst eingeschätzten Stärken und Schwächen inhaltlich darzustellen würde den Rahmen dieser Arbeit sprengen und ist auch zunächst für die interne Diskussion gedacht. Auf Basis dieser Leitgedanken des beruflichen Selbstverständnisses des Teams wurden ein Handlungsplan, Forderungen an die übergeordnete Behörde und an die Politik aufgestellt. Im Ganzen war das

Team äußerst zufrieden mit den abgelaufenen Prozessen, wie sich im Feedback an das Evaluationsteam geäußert wurde (und hat zwei Jahre später gerade diesen Evaluationsbaustein beim gleichen Evaluationsteam beauftragt).

### 3.4.3 Limitationen

Die Schwächen von Selbstevaluationsansätzen liegen immer in zwei Bereichen (König 2007): Einerseits handelt es sich im Großteil um Selbsteinschätzungen, die dadurch natürlich Verzerrungen unterliegen können. Zum anderen sind das methodische Konzept und die methodische Durchführung selten so strikt und professionell wie von einem spezialisierten Evaluationsteam.

## 4 Verbindung der Bausteine im Sinne einer dialektischen Evidenztriangulation

Eine der zentralen methodologischen Fragen von Mixed-Methods-Ansätzen ist, wie man die unterschiedlichen Teile zusammenfügt (Erzberger und Kelle 1999). Wenn sich die Ergebnisse ergänzen, ist dies als Bestätigung der Gültigkeit (Kreuzvalidierung) zu werten. Kritisch wird es, wenn einzelne Bausteine zu sich widersprechenden Ergebnissen führen. Hier sind im Zusammenhang mit Mixed Methods verschiedene Argumentationsstränge versucht worden.

Die konstruktivistische Position (Lincoln und Guba 2013) würde die verschiedenen Ansätze als unterschiedliche Sichtweisen oder unterschiedliche Rekonstruktionen nebeneinander stehen lassen, da für sie ein richtiges Ergebnis Illusion ist. Ein solcher Ansatz ist von Guba und Lincoln auch in der Evaluation versucht worden, wonach als (vorläufiges, raum-zeitlichem Wandel unterliegendes) Evaluationsergebnis gilt, worauf sich die an dem zu Evaluierenden Beteiligten, unter Moderation der Evaluatoren, einigen. Eine solche Position erscheint mir aber, gerade im Gesundheitsbereich, wenig zielführend. Sie liefert auch keine Begründung für einen Mixed-Methods-Ansatz, da nicht unbedingt unterschiedliche methodische Herangehensweisen an die Rekonstruktion subjektiver Perspektiven für nötig gehalten werden.

Im Bereich von Mixed Methods wird heute meist die Position des Pragmatismus (als wissenschaftstheoretischer Ansatz) eingenommen (z. B. Creswell und Plano Clark 2010). Danach werden die verschiedenen Ansätze und deren Ergebnisse danach gewertet, ob sie einer Problemlösung am nächsten kommen. Handlungsergebnisse in der konkreten Lebenswelt, Nützlichkeitsabwägungen stehen hier im Vordergrund. Aber auch diese Position kann wenig befriedigen, da sie die Entscheidung, welche Evaluationsergebnisse am verlässlichsten sind, von äußeren Kriterien abhängig macht.

Aus diesen Überlegungen heraus soll hier mit dem dialektischen Ansatz eine Alternative aufgezeigt werden (Mayring et al. 2007). Die dialektische Position, wie sie auch von Greene (2007) zur Begründung von Mixed Methods vorgeschlagen wurde, geht davon aus, dass eine Zusammenschau verschiedener mentaler Modelle zu höherer Einsicht führen kann. Die Hegel'sche Theorie von These, Antithese und Synthese verdeutlicht dies. Ein erster Untersuchungsbaustein im Evaluationsprojekt

führt zu Ergebnissen, die als erste These fungieren. Die Kritik an der Methodik führt zur Konzeption eines alternativen Forschungsansatzes, der Ergebnisse nach sich zieht, die eine Antithese darstellen. Nun geht es darum, in kritischer Diskussion eine Synthese aus den beiden Ergebnissen zu ziehen. Gegebenenfalls ist diese Synthese aber wiederum mit methodischen Einschränkungen behaftet; sie dient nun als neue These, die eine neue Antithese auf den Plan ruft und wiederum synthetisiert werden muss. Ich möchte dies an dem hier dargestellten Beispiel verdeutlichen.

*These I* Der klassische Evaluationsansatz (quantitative Messung im Vorher-Nachher-Design) führt zum Nachweis leichter Befindensverbesserungen im subjektiven Bereich.

*Antithese I* Die Kritik (s. oben Limitationen) am einfachen Design (z. B. keine Kontrollgruppe) und standardisierter Messung (Test) führt zur Konzeption einer qualitativen Interviewstudie mit den Betroffenen. Die Ergebnisse sind sehr viel positiver, abgesehen von leicht veränderbaren Einschränkungen.

*Synthese I* Die Drogenambulanz trifft auf hohe Zufriedenheit der Klienten, was sich in einer leichten Befindensverbesserung bereits in den ersten Monaten der Behandlung abzuzeichnen scheint.

*Antithese II* Die Analyse der objektiven Rahmenbedingungen fördert zutage, dass mangelnde Schulbildung und in der Konsequenz ein extrem hoher Anteil Erwerbsloser einer Rehabilitation der Klienten im Wege steht. Auch eine objektivere Sicht auf den Behandlungserfolg durch Hinzuziehen des Arzturteils und der Behandlungsdokumentation zeigt, dass nur bei ca. der Hälfte der Personen von einer substanziellen Verbesserung zu reden ist.

*Synthese II* Trotz hoher Behandlungszufriedenheit zeigt nur ein Teil der Personen grundlegende Behandlungserfolge.

*Antithese III* Die Ergebnisse müssen gerahmt werden durch die Einschätzungen des Teams vor Ort. Hier liegt auch der Träger von Verbesserungsprozessen.

*Synthese III* Klare Verbesserungen bei der Hälfte der Klientel werden vom Team als großer Erfolg wahrgenommen, auch weil Veränderungsprozesse (schulische Bildung, Berufsstatus) hier nur sehr langfristig zu erreichen sind. Die hohe subjektive Zufriedenheit wird als der Schlüssel für selbstbestimmte Veränderungsprozesse gesehen. Solche Überlegungen werden im Leitbild der Institution festgehalten, auch für die zukünftige Arbeit.

Ein solches Verständnis von kritischer Methodendiskussion und sukzessiver Erweiterung der Evidenzbasis entspricht unserer Meinung nach am treffendsten Erkenntnisfortschritt durch Mixed-Methods-Ansätze und ist, in Absetzung von der Evidenzhierarchie, als Evidenztriangulation zu fassen. Nicht nur im Bereich der Gesundheitsforschung erscheint er vielversprechend.

## 5 Fazit und Ausblick

Ausgangspunkt der Überlegungen waren die Unzulänglichkeiten einer rein quantitativ ausgerichteten Forschungsmethodik, wie sie im Mangel an Replizierbarkeit experimenteller psychologischer Studien gezeigt werden konnte. Im Bereich von Grundlagenforschung hat dies zu großen Verunsicherungen geführt. Im Bereich von angewandter Forschung, und hier besonders auf dem Gebiet von Gesundheitsforschung kann aber durchaus Schaden (z. B. an der Gesundheit von Menschen) durch einseitige Forschungsergebnisse angerichtet werden. Mixed Methods kann hier sicher weder als Zauberformel noch als Allheilmittel dienen, und die Möglichkeiten, Befunde durch verschiedene kombinierte und integrierte Methodenbausteine abzusichern, klingt zunächst vielversprechend. Die mittlerweile zahllosen Vorschläge solcher Mixed-Methods-Designs lassen den Eindruck entstehen, jede Art von Methodenkombination sei nützlich. Die wissenschaftstheoretische Position des Pragmatismus, die hier gerne zur Fundierung herangezogen wird (Cresswell und Plano Clark 2010), unterstützt diesen Eindruck. Unserer Meinung nach ist aber hier ein vorsichtiges, schrittweises und kritisches Argumentieren der fragestellungs- und gegenstandsadäquaten Forschungsmethodik unbedingt notwendig. Und das ist der Kern des hier an einem Beispiel aufgezeigten dialektischen Herangehens. Es geht nicht nur um ein Aufsummieren verschiedener Evidenzen, ein additives Verbreitern der empirischen Basis, sondern um ein aus der immer notwendigen Methodenkritik schrittweise Entwickeln alternativer methodischer Herangehensweisen unter Einsatz qualitativer und quantitativer Forschungsstrategien. Der Gegenstand wird auf diese Weise aus unterschiedlichen Blickwinkeln beleuchtet, eingekreist, um das Verständnis zu vertiefen. Denn lange schon haben wir uns in der Methodologie davon verabschiedet, mit einer einzigen Studie, einer einzigen Methode eine sozialwissenschaftliche Fragestellung hinreichend beantworten zu können.

Open access funding provided by University of Klagenfurt.

**Open Access** Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

## Literatur

- Angermeyer, Matthias C., Reinhold Kilian und Herbert Matschinger. 2000. *WHOQOL-100 und WHOQOLBREF. Handbuch für die deutschsprachige Version der WHO Instrumente zur Erfassung von Lebensqualität*. Göttingen: Hogrefe.
- Applebaum, Robert A., Jane K., Straker und Scott M. Geron. 2004. *Patientenzufriedenheit: benennen, bestimmen, beurteilen*. Bern: Huber.
- Baur, Nina. 2009. Measurement and selection bias in longitudinal data. A framework for re-opening the discussion on data quality and generalizability of social bookkeeping data. *Historical Social Research (HSR)* 34:9–50.
- Beywl, Wolfgang. 1988. *Zur Weiterentwicklung der Evaluationsmethodologie. Grundlegung, Konzeption und Anwendung eines Modells der responsiven Evaluation*. Frankfurt a. M.: Lang
- Bortz, Jürgen, und Nikola Döring. 2006. *Forschungsmethoden und Evaluation: für Human- und Sozialwissenschaftler* (4. Aufl.). Berlin: Springer.
- Brunner, Eva, Claudia Brunner, Philipp Mayring, Claudia Öhner-Scheiber und Brigitte Prehslauer. 2009. Suchbehandlung auf dem Prüfstand – Evidenztriangulation statt Evidenzhierarchie. *Suchtmedizin* 11:59–62.
- Bundeszentrale für gesundheitliche Aufklärung BZgA. 2005. *Qualitätszirkel in der Gesundheitsförderung und Prävention*. Köln: BZgA.
- Creswell, John W., und Vicki L. Plano Clark. 2010. *Designing and conducting mixed methods research* (3<sup>rd</sup> edition). Thousand Oaks, CA: Sage.
- Denzin, Norman. 1970. *The research act. A theoretical introduction to sociological methods*. Chicago: Aldine.
- Erzberger, Christian, und Udo Kelle. 1999. The integration of qualitative and quantitative methods. Methodological models and their significance for practicing social research. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 51:509–531.
- Faltermaier, Toni. 2005. *Gesundheitspsychologie. Grundriss der Psychologie*, Bd. 21. Stuttgart: Kohlhammer.
- Fiedler, Klaus. 2016. Empfehlungen der DGPs-Kommission „Qualität der psychologischen Forschung“ <http://www.dgps.de/uploads/media/Empfehlungen-Qualitaet-der-Forschung-DGPs.pdf> (Zugegriffen: 15.10.2016).
- Fischer, Gabriele, und Beate Kayer. 2006. Substanzabhängigkeit von Morphinotyp. State-of-the-Art der Erhaltungstherapie mit synthetischen Opioiden. *Psychiatrie & Psychotherapie* 2:39–54.
- Flick, Uwe, Ernst von Kardorff und Ines Steinke. Hrsg. 2004. *Qualitative Forschung. Ein Handbuch* (3. Auflage). Reinbek: Rowohlt.
- Flick, Uwe 2008. *Triangulation – eine Einführung* (2. Auflage). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Greene, Jennifer C. 2007. *Mixed methods in social inquiry*. San Francisco, CL: Wiley.
- Hesse-Biber, Sharlene N. 2010. *Mixed methods research. Merging theory with practice*. New York: Guilford Press.
- Huber, Otto. 2005. *Das psychologische Experiment. Eine Einführung* (4. Aufl.). Bern: Huber.
- Hurrelmann, Klaus, und Oliver Razum. Hrsg. 2012. *Handbuch Gesundheitswissenschaften* (5. Aufl.). Weinheim: Beltz.
- Kallus, Wolfgang. 2016. *Erstellung von Fragebogen* (2. Aufl.). Wien: facultas-utb.
- König, Joachim. 2007. *Einführung in die Selbstevaluation. Ein Leitfaden zur Bewertung der Praxis sozialer Arbeit* (2. überarb. Aufl.). Freiburg: Lambertus.
- Lincoln, Yvonne S., und Egon G. Guba. 2013. *The constructivist credo*. Walnut Creek, CL: Left Coast Press.
- Mayring, Philipp. 2001. Kombination und Integration qualitativer und quantitativer Analyse. Forum Qualitative Sozialforschung. <http://www.qualitative-research.net/index.php/fqs/article/view/967> (Zugegriffen: 15.8.2015).
- Mayring, Philipp, Günther L. Huber, Leo Gürtler und Mechthild Kiegelmann. Hrsg. 2007. *Mixed methodology in psychological research*. Rotterdam: Sense Publishers.
- Mayring, Philipp. 2009. Designs in der Psychotherapieevaluation – von der Evidenzhierarchie zur Evidenztriangulation. In *Bei Jutta auf der Couch*, Hrsg. Axel Krefting, 147–152. Klagenfurt: Drava.
- Mayring, Philipp. 2015. *Qualitative Inhaltsanalyse. Grundlagen und Techniken* (12. Aufl.). Weinheim: Beltz.
- Mertens, Wolfgang. 1975. *Sozialpsychologie des Experiments. Das Experiment als soziale Interaktion*. Hamburg: Hoffmann und Campe.

- Lauterbach, Karl W., und Matthias Schrappe. Hrsg. 2004. *Gesundheitsökonomie. Qualitätsmanagement und Evidence-based Medizin. Eine systematische Einführung* (2. Aufl.). Stuttgart: Schattauer.
- Open Science Collaboration. 2015. Estimating the reproducibility of psychological science. *Science* 349(6251). doi:[10.1126/science.aac4716](https://doi.org/10.1126/science.aac4716) PMID:26315443.
- Patton, Michael Quinn. 2015. *Qualitative research and evaluation methods* (4. Aufl.). Thousand Oaks: Sage.
- Pawson, Ray, und Nicholas Tilley. 1997. *Realistic evaluation*. London: Sage.
- Ross, Andrew. Hrsg. 1996. *Science wars*. Durham: Duke University Press.
- Seifert, Josef W. 1992. *Visualisieren, Präsentieren, Moderieren*. Speyer: Gabal Verlag.
- Tashakkori, Abbas, und Charles. Teddlie. 1998. *Mixed methodology: Combining qualitative and quantitative approaches*. Thousand Oaks: Sage.
- v. Zerssen, Detlev. 1976. *Die Beschwerden-Liste. Klinische Selbstbeurteilungsskalen aus dem Münchener Psychiatrischen Informations-System*. Weinheim: Beltz Test Gesellschaft GmbH.
- Witzel, Andreas. 2000. Das problemzentrierte Interview. *Forum Qualitative Sozialforschung* <http://nbn-resolving.de/urn:nbn:de:0114-fqs0001228> (Zugegriffen: 15.10.2016).
- World Health Organisation, United Nations Office on Drugs and Crime, Joint United Nations Programme on HIV/AIDS. 2004. *Position Paper: Substitution maintenance therapy in the management of opioid dependence and HIV/AIDS prevention*. WHO Library Cataloguing-in-Publication Data.

**Philipp Mayring** 1952, Dr. phil., Professor für psychologische Methodenlehre, Leiter der Abteilung für Angewandte Psychologie und Methodenforschung sowie des Zentrums für Evaluation und Forschungsberatung ZEF der Alpen-Adria Universität Klagenfurt, Vorsitzender der Gesellschaft für Evaluation DeGeval. Forschungsgebiete: Textanalysemethoden, Evaluationsmethoden, Gesundheitspsychologie. Veröffentlichungen: *Einführung in die Qualitative Sozialforschung* (5. Aufl.). Weinheim 2002; *Qualitative Inhaltsanalyse* (12. Aufl.). Weinheim 2014; *Mixed methodology in psychological research*. Rotterdam 2007 (mit L.G. Huber, L. Gürtler und M. Kiegelmann); *Glück – aber worin liegt es? Zu einer kritischen Theorie des Wohlbefindens*. Göttingen 2013 (mit N. Rath).