



HALLESCHER BEITRÄGE ZU DEN
GESUNDHEITS- UND
PFLEGEWISSENSCHAFTEN



Sinn machen »quantitative« Untersuchungen nur als Teile »qualitativer« Studien

Zur Indikation von Interviews
zur Erzeugung externer Evidence – ein Überblick

von Prof. Dr. Johann Behrens

HERAUSGEBER: JOHANN BEHRENS
REDAKTION & GESTALTUNG: GERO LANGER

I. JAHRGANG
ISSN 1610-7268

I

Vor der Veröffentlichung werden Beiträge im üblichen »peer review«-Verfahren auf ihre Publikationswürdigkeit hin begutachtet. Außer der anonymen Beurteilung der Publikationswürdigkeit geben die Gutachtenden in der Regel Anregungen für Verbesserungen an die Autorinnen und Autoren. Die Aufnahme der Anregungen wird nicht in einer zweiten Begutachtungsrunde geprüft. Daher kann nicht notwendigerweise davon ausgegangen werden, daß die publizierten Fassungen allen Anregungen der Gutachtenden entsprechen. Die Verantwortung für die publizierte Fassung liegt bei den Autorinnen und Autoren.

Die Zeitschrift und alle in ihr enthaltenen Beiträge sind urheberrechtlich geschützt. Die Nutzung der Zeitschrift und der in ihr enthaltenen Beiträge ist insoweit frei, als nichtkommerziell handelnden Personen, Firmen, Einrichtungen etc. ein begrenztes Recht auf nichtkommerzielle Nutzung und Vervielfältigung in analoger und digitaler Form eingeräumt wird. Das betrifft das Laden und Speichern auf binäre Datenträger sowie das Ausdrucken und Kopieren auf Papier. Dabei obliegt dem Nutzer stets die vollständige Angabe der Herkunft, bei elektronischer Nutzung auch die Sicherung dieser Bestimmungen.

Es besteht – außer im Rahmen wissenschaftlicher und schulischer Veranstaltungen öffentlicher Träger – kein Recht auf Verbreitung. Es besteht kein Recht zur öffentlichen Wiedergabe. Das Verbot schließt das Bereithalten zum Abruf im Internet, die Verbreitung über Newsgroups und per Mailinglisten ein, soweit dies durch die Redaktion – oder durch den/die Urheber des betreffenden Beitrags – nicht ausdrücklich genehmigt wurde. Darüber hinausgehende Nutzungen und Verwertungen sind ohne Zustimmung des Urhebers unzulässig und strafbar.

Eine Markenbezeichnung kann warenzeichenrechtlich geschützt sein, auch wenn bei ihrer Verwendung das Zeichen ® oder ein anderer Hinweis fehlen sollte. Die angegebenen Dosierungen sollten mit den Angaben der Produkthersteller verglichen werden. Für Angaben über Dosierungen und Applikationsformen kann keine Gewähr übernommen werden.

Gesetzt mit L^AT_EX 2_ε in der Stempel Garamond

Redaktionsschluß: 16. November 2005

IMPRESSUM

Die »Halleschen Beiträge zur Gesundheits- und Pflegewissenschaft« werden herausgegeben von Prof. Dr. phil. habil. Johann Behrens

Redaktion & Gestaltung: Dipl. Pflege- u. Gesundheitswiss. Gero Langer

Kontakt: Martin-Luther-Universität Halle-Wittenberg · Medizinische Fakultät · Institut für Gesundheits- und Pflegewissenschaft ·

German Center for Evidence-based Nursing · Magdeburger Straße 27 · 06112 Halle/Saale · Deutschland

Telefon 0345 – 557 4450 · Fax 0345 – 557 4471 · E-Mail gero.langer@medizin.uni-halle.de

Website <http://www.medizin.uni-halle.de/pflegewissenschaft/journal/>

ISSN 1610-7268

Alle Rechte vorbehalten.

© 2002 Prof. Dr. Johann Behrens, Halle/Saale, Deutschland

Inhaltsverzeichnis

1	Einleitung	1
2	Interviews und ihre Indikation: Wann sind sie der Königsweg für die Produktion externer Evidence?	4
3	Sinn machen quantitative Untersuchungen nur als Teil qualitativer Studien – auf dem Wege zu einem Überblick	8
3.1	Induktive und deduktive Verfahren als alternative, nicht als unterschiedliche Abschnitte derselben Verfahrenskette?	9
3.2	Entgegensetzung der Gütekriterien qualitativer und quantitativer Verfahren?	12
3.3	Gleichsetzung von Gütekriterien mit Techniken	15
4	Schlussbemerkung	18
	Literatur	19

1 Einleitung

Wann sind bestimmte Formen von Interviews, wann sind Interviews überhaupt angesagt? Interviews in all ihren verschiedenen Formen erscheinen – zumindest was ihre Häufigkeit angeht – als der dominierende Königsweg der qualitativen Sozial-, Gesundheits- und Pflegeforschung. Warum wird so häufig den Interaktionsprotokollen, die durch Interviews entstehen, der Vorzug vor Interaktionsprotokollen gegeben, die in Arbeits- und Lebensprozessen aller Art fortwährend entstehen (zu solchen sogenannten »prozessproduzierten Daten« vgl. v. Ferber/Behrens 1996)? Gerade im Gesundheits- und Pflegebereich werden nicht nur in großer Zahl Verlaufsdokumentationen, Telefonmitschnitte, Entscheidungsprotokolle, Messungen aller Art, Videodokumentationen, Absprachen, Anweisungen, Vorschläge und Memoranden erstellt. Es ergibt sich auch etwas häufiger als in anderen Lebensbereichen die Gelegenheit für Forschende, ein Aufnahmegeräte mitlaufen zu lassen.

Offenbar muss es Fragestellungen geben, für die ein retrospektives Interview mit einem Entscheidungsträger oder Experten einige Zeit nach einer Entwicklung geeigneter und aussagekräftiger ist als prozessproduzierte Interaktions-Daten aus der Zeit, über die dieser berichtet. Offenbar auch muss es Fragestellungen geben, für die offene Interviews geeigneter sind als geschlossene, und solche, für die generell quantitative Untersuchungs- und Auswertungsstrategien geeigneter als qualitative.

Ein solches Urteil der Eignung einer Methode für eine spezifische Fragestellung bezeichnete ich – mit einem aus diagnostischen und therapeutischen Entscheidungen vertrauten Begriff – als »Indikation«. Wie eine bestimmte Diagnose in Verbindung mit einem bestimmten Allgemeinzustand einer Patientin eine bestimmte Behandlung angeraten = »indiziert« sein lässt und andere Methoden nicht zur »Methode der Wahl« macht, so gilt auch für Forschungsmethoden, dass nicht eine Methode für alle Fragestellungen gleichermaßen indiziert ist. Alle therapeutischen Berufe, ob Ärzte oder andere Gesundheitsberufe, legen bekanntlich und mit Recht großen Wert darauf, dass sie Behandlungsmethoden nicht danach wählen, was sie selber könnten, sondern danach, was durch die Bedürfnisse ihrer Klienten am meisten indiziert sei. Da niemand alles gleich gut könnte, gäbe es die »Überweisung«. Überweisung setzt allerdings voraus, dass Therapeuten einen hinreichenden Überblick über verschiedene Methoden und ihre wichtigsten Stärken haben, um rational begründet überweisen und Methoden kombinieren zu können. Diese Voraussetzung ist keineswegs trivial und auch nicht einfach zu erfüllen. Man mag bezweifeln, ob faktisch im deutschsprachigen Gesundheitswesen immer und rechtzeitig Anhaltszeichen für oder gegen eine Methode rational erkannt und berücksichtigt werden. Unbezweifelbar aber ist im deutschsprachigen Gesundheitswesen trotz allem Schulstreit die normative Geltung des Ideals der rationalen Wahl zwischen Methoden: Es ist kein Angehöriger eines Gesundheitsberufes in der Schweiz, Österreich oder Deutschland bekannt geworden, der alle möglichen Fälle nach derselben Methode behandeln wollte, noch nicht einmal einer, der sich prinzipiell außerstande erklärte zu überweisen.

An dieser Stelle ist aber darauf einzugehen, dass nicht nur die Methoden der narrativen, familiengeschichtlichen und Experten-Interviews versus der Methoden der Auswertung prozessproduzierter Daten zur rationalen Wahl stehen, nicht nur »qualitative« und »quantitative« Methoden, sondern dass auch innerhalb der qualitativen Interviewverfahren eine gar nicht dokumentierbare Fülle von Methoden die Frage nach ihrer jeweils spezifischen Indikation aufwirft. Christel Hopf (1994) hat schon vor nahezu 10 Jahre sechs weitere Formen unterschieden, nämlich das offene, das thematisch fokussierte, das problemzentrierte, das episodale, das biographische, das Dilemma-Interview. Dem ist noch das am Frankfurter Sigmund-Freud-Institut von Michael Wolf und anderen weiter entwickelte, die Dynamik der Gegenübertragung nutzende »szenische Interview«, das kontrastive Interview und die Wahlhandlung Meinungsabgabe (vgl. zu den drei letztgenannten Behrens 1983, zur damaligen Diskussion Gerhardt 1985, Flick 1994, Lincoln/Guba 1985) zu zählen. Insgesamt schätze ich, dass es eher 30 als 20 verschiedene Schulen qualitativer Interviews gibt. Sie lassen daran zweifeln, ob ein Kollektiv-Singular »Qualitatives Interview« überhaupt existiert. Sie hier alle der Reihe nach in je sieben Zeilen charakterisieren zu wollen, würde ihnen allen Unrecht tun (vgl. Behrens/Langer 2002). Vielen dieser Interviewverfahren und Interviewschulen im qualitativen Bereich ist gemeinsam, dass sie nach einem Erzählanreiz (z. B. Focus-Interview, narratives Interview und viele andere) die Strukturierung des Gesprächsverlaufs weit mehr als beim standardisierten Interview dem Gesprächspartner überlassen – um an den so entstehenden und in der Regel per Tonbandgerät aufgezeichneten Texten nachverfolgen zu können, wie der Interviewte seine Darstellung strukturiert und auf die Anwesenheit des Interviewers reagiert. Dann hören die Gemeinsamkeiten schon auf. Es bestehen sehr unterschiedliche Auffassungen

davon, wie die Interviewer sich verhalten sollten, ob Unbefangenheit und Voreingenommenheit im Interview selber zu demonstrieren sind, oder sich erst durch falsifikationistische Auswertungsverfahren einstellen, wie die Verfahren kombinierbar sind usw. Eine – unbeabsichtigte – Gemeinsamkeit scheint meiner Meinung nach wieder darin zu bestehen, dass qualitative Interviews welcher Schule auch immer die Interpreten häufig in Zeitprobleme führen oder, besser gesagt, Mängel bei der zeitlichen Planung und Konzipierung von Projekten im Stadium der Auswertung ans Licht bringen. Seit Jahrzehnten gibt es Klagen darüber, wie die Vorteile offener Interviews, nämlich verschiedene Lesarten der Strukturierungen des Interviewers selber am Text zu überprüfen, unter Zeitdruck aufgegeben werden (Merton und Kendall 1979, S. 185). Schon die Interviews selber verletzen häufig unter Zeitdruck und Ungeduld des Interviewers alle von der jeweiligen Schule aufgestellten Regeln des Interviewens (vgl. auch dafür schon Merton und Kendall 1979).

Qualitative Interviewverfahren berufen sich häufig auf die Grundannahme der verstehenden Soziologie, aber der Umkehrschluss ist nicht möglich. Auch quantitative Verfahren teilen die Annahme der verstehenden Soziologie, dass aus der beständigen alltäglichen Herstellung einer gemeinsamen Welt in wechselseitig deutenden Interaktionen der Prozesscharakter, die Rekursivität und Reflexivität sozialer Wirklichkeit resultiere. Daher kann diese Annahme über die Entstehung sozialer Wirklichkeit nicht qualitative von quantitativen Verfahren unterscheiden. Auch unterscheidet es qualitative und quantitative Verfahren sicher nicht, dass qualitative Verfahren davon ausgehen, eine Handlung zu verstehen sei gleichbedeutend damit, die Absichten des Handelnden bei dieser Handlung zur Kenntnis zu nehmen (ausführlicher Behrens 1980). Nahezu allen qualitativ arbeitenden Forschern ist bekannt und Ausgangspunkt ihrer Untersuchungen, dass Handlungen unbeabsichtigte Folgen haben können und das Verstehen von Handlungen oder sogar Personen sich keineswegs auf die Identifikation von Handlungsabsichten beschränkt. Schon an diesen wenigen Bemerkungen wird deutlich, dass eine Kurzcharakteristik aller 20 bis 30 Schulen qualitativer Interviews keiner einzigen gerecht würde.

Vielmehr ist auf einen Umstand zurück zu kommen, den Christel Hopf (1994, überarbeitet 2000) schon vor fast 10 Jahren beklagte und der erstaunlicherweise unverändert aktuell geblieben ist. Die Methoden werden häufig ohne Bezug aufeinander verwandt, mal gemischt, mal wieder emphatisch gegeneinander gestellt, ohne dass sich eine Entwicklung in der Diskussion ergäbe. Von der rationalen Auswahl der »Methode der Wahl«, von der allseits bekannten Indikation der jeweiligen Methode, wie sie unbestrittene Norm bei der Auswahl von Therapien ist und dort jeder ausgebildete Bademeister beansprucht, scheinen wir noch etwas entfernt.

Das hat – auflösbare – Gründe, die ich in diesem einleitenden Kapitel kurz anspreche. Dabei gehe ich nicht auf die Frage der Prozess-Qualität qualitativer Gesundheits- und Pflegeforschung ein, die z. B. auch den Datenschutz, die Rechte der Mitarbeiterinnen als Co-Autorinnen usw. umfasst und die Uwe Flick in diesem Band zum Thema macht. Ich beschränke mich vielmehr auf den Weg zu einer rationalen Wahl zwischen Methoden auf wenige Haupthindernisse, die bereits der Frage nach der Indikation und damit einem Überblick über Methoden im Wege stehen.

Um die zu begründenden Ergebnisse vorweg zu nehmen: Ein Hindernis kann die Site sein, eindimensionale Studienhierarchien mit einem Goldstandard oben zu errichten. Eindimensionale Hierarchien kann es nur geben, wenn es nur eine Täuschungsmöglichkeit gibt oder in der alltäglichen Praxis mit einer Methode alle Selbsttäuschungsmöglichkeiten zugleich bewältigt werden können.

Ein Hindernis, das mit diesem zusammen hängt, ist, nicht zwischen Zielen und Mitteln zu unterscheiden. Wenn die Ziele mit ihren Mitteln identisch sind, lassen sich Methoden weder in ihren Stärken bewerten noch kombinieren. Z. B. versuche ich zu zeigen, dass auch unterschiedliche Methoden wie Grounded Theory, Objektive Hermeneutik und prognostische Verlaufsanalysen im Falsifikationismus, in der Sequenzanalyse und – mit Vorsicht – im fallibilistischen Realismus hinreichende Ähnlichkeiten aufweisen (Abschnitt 3), um einmal mit Blick auf ihre Indikation verglichen zu werden.

Ein Hindernis ist, sich über den Zweck keine Rechenschaft abzulegen. Ohne Zweck lässt sich über eine Methode nichts sagen. Viele Gespräche ähneln einem – symmetrischen oder asymmetrischen – Interview. Aber Interviews, die eine Verständigung über gemeinsam zu verfolgende Ziele erbringen sollen, haben anderen Kriterien zu genügen als Interviews, die Erfahrungen für andere zugänglich und prüfbar machen sollen, die diese Erfahrungen selber nicht machten (Abschnitt 2). Auf diese Hindernisse gehe ich im Folgenden kurz ein.

2 Interviews und ihre Indikation: Wann sind sie der Königsweg für die Produktion externer Evidence?

Das erste ganz triviale Hindernis auf dem Wege zu einem Überblick über Interviewverfahren und alternativer Verfahren besteht darin, dass wir uns über den Zweck keine Klarheit verschaffen (vgl. v. Kardorff 2000). Ohne Ziel kann es offenbar keine »Indikation« geben. In den Gesundheits- wie Pflegewissenschaften ist die Unterscheidung elementar, ob ein Interview der Verständigung zwischen einem Klienten und einem Professionsangehörigen im Arbeitsbündnis dienen soll (»interne Evidenz«) oder ob es (darüber hinaus) auch Wissen erzeugen können soll, auf das in weiteren Klientenbeziehungen zurück gegriffen werden kann. Dieses Wissen ist außerhalb der ersten Klientenbeziehung erzeugt und insofern »extern«, trotzdem soll es für diese brauchbar sein und zu dem beitragen, was der in Gesundheitswesen enorm verbreitete unbestimmte Rechtsbegriff des »gesicherten Wissens« oder des Erkenntnisstandes bezeichnet.

Jede – machtasymmetrische oder symmetrische – Verständigung zum Aufbau interner Evidenz ist körpersprachliche und sprachliche Kommunikation, und das Gespräch erinnert häufig an die Form eines Interviews. Aber das Interview, in dem sich Klienten und Professionen über Ziele der Pflege und Behandlungen verständigen und verhandeln, muss anderen Ansprüchen genügen als das Interview, das ihnen für ihre Entscheidungen »externe evidence« vermitteln soll. Unter welchen Bedingungen erzeugen Interviews externe evidence, also Erfahrungen, die wir nicht selber gemacht haben und aus denen wir als Professionen und Klienten trotzdem »gesichertes Wissen« für unsere Entscheidungen

beziehen können? Auch wenn sich der Aufbau interner Evidenz und externer evidence in der Klientenbeziehung in der Regel verschränkt und etwa im action research auch in der Forschung beide Bereiche verschränkt wurden, so ist doch immer eindeutig, ob ein Wissen als interne Evidenz oder externe Evidence in Anspruch genommen wird.

Für alle Professionen, die nur im Auftrag ihres Klienten und in Zusammenarbeit mit ihm ihre Aufgabe erfüllen können, ist die Unterscheidung von interner Evidenz und externer Evidence grundlegend und unhintergebar: Interne Evidenz entsteht innerhalb der Klientenbeziehung. Externe Evidence baut sich aus kontrollierten Erfahrungen außerhalb der Klientenbeziehung auf, wie sie in »Studien« zusammengefasst heranziehbar sind.

Externe Evidence ziehen wir in Klientenbeziehungen heran, weil es zu schmerzhaft, leidvoll und verschwenderisch wäre, sämtliche Irrwege ausschließlich innerhalb der Klientenbeziehung an einzelnen Klienten durchzuprobieren. In der Regel würden die Klienten ebenso wenig wie die Professionen zu Lebzeiten noch erleben, dass sich am Ende von Versuch und Irrtum das Richtige herausstellte.

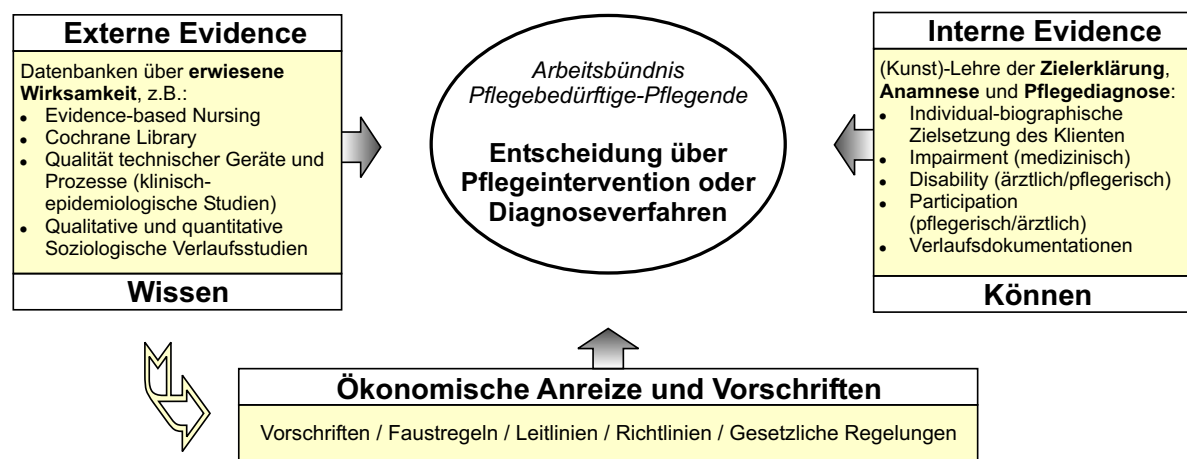


Abbildung 1: Evidencebasierte pflegerische professionelle Praxis: interne und externe evidence, moralische und ökonomische Anreize

Aber die Nutzung externer Evidence innerhalb der einzigartigen Klientenbeziehung, die Verschränkung externer Evidence und interner Evidenz ist keineswegs einfach und trivial. Externe Evidence ersetzt nie interne Evidenz. Aus externer Evidence ergibt sich nie umstandslos die angemessene Handlung. Das wird sogar an scheinbar einfachen Beispielen gesicherten Wissens klar. Wenn wir bei bestimmten Krankheiten aus externer Evidence wissen, dass 95% der Kranken das nächste viertel Jahr nur überleben, wenn ihnen ein Bein amputiert wird, folgt daraus selbstverständlich nicht die Amputation des Beins. Für diese Entscheidung des Erkrankten bedarf es der Verständigung über seine Ziele und Bedürfnisse, seine Schmerzen und Perspektiven, bedarf es des Aufbaus interner Evidenz als Fallverstehen. Bei diesen Entscheidungsprozessen kann nicht einmal die evident ermittelte Häufigkeit (95% der Erkrankten überleben ohne Amputation nicht das nächste viertel Jahr) in eine individuelle Prognose automatisch überführt werden. Der einzelne

kann ja zu den 5% gehören. So gut Häufigkeitsaussagen in Prognosen für Gruppen übersetzt werden können, sowenig erlauben sie Prognosen für den Einzelfall (sonst gäbe es keine Versicherungen).

In den Gesundheitsberufen kommt es aus verständlicher Entscheidungsfurcht und zur Vermeidung schwieriger körpersprachlicher und sprachlicher Kommunikationsprozesse bekanntlich immer wieder vor, dass so getan wird, als ergäben sich Entscheidungen aus externer evidence von selbst. Schlaganfallpatienten werden so gelagert, wie es sich für die meisten Schlaganfallpatienten in externen Studien als günstig für den weiteren Verlauf erwiesen hat – ohne nach ihren individuellen Gewohnheiten zu fragen. Das ist zweifellos ein Kunstfehler.

Die Vermittlung und der Aufbau von externer Evidence und interner Evidenz sind das eigentliche Thema insbesondere von Evidence based Nursing und auch von Evidence based Medicine. Die Frage hingegen, wie eine Studie durchgeführt sein muss, damit sie uns externe evidence vermitteln kann, ist nur ein kleiner und eher einfacher Teil dieses Themas. Der Platz in diesem Band reicht nicht, um auf das ganze Thema der Vermittlung von externer Evidence und interner Evidenz in der Klientenbeziehung einzugehen, daher muss auf Raspe 2000, Behrens 1980 und 1983, Behrens und Langer 2002 verwiesen werden.

Zweifellos beanspruchen die meisten qualitativen Interviewverfahren (für einen Überblick der ausdifferenzierten Schulbildung siehe schon Hopf 1994, 2000) nicht nur, »gesicherte« externe evidence zu erzeugen. Dieser Anspruch wird auch anerkannt. Eine Durchsicht der ersten drei Jahrgänge der Zeitschrift Evidence based Nursing des BMJ-Verlags von 1999 - 2001 ergab, dass ein Viertel der als externe evidence erzeugende aufgenommenen Studien »qualitative« Verfahren, überwiegend Interviews, verwendet hatte. Seit René Königs Handbuch wird vom Interview als vom Königsweg der Sozialforschung gesprochen.

Die Verbreitung, ja Vorliebe für Interviews in der qualitativen Sozial-, Pflege- und Gesundheitsforschung ist trotzdem nicht ganz nachvollziehbar, insbesondere wenn Interviews in einem erkenntnistheoretisch eher naiven Realismus genutzt werden. (Im 3. Abschnitt dieses Aufsatzes versuche ich zu zeigen, dass dieser naive Realismus prominent im Mantel des radikalen Konstruktivismus auftritt.) Für viele Fragen sind andere Handlungsprotokolle nicht schlechter, sondern eher besser geeignet als das Handlungsprotokoll Interview. Nicht selten werden Interviews zum Ersatz für die Mühe, prozessproduzierte Daten auszuwerten. Das mag im Tagesjournalismus noch verständlich sein, insbesondere wenn Volontäre eine Vielzahl von Themen ohne Zeit zu eigenem Quellenstudium abarbeiten müssen. Das Interview erspart scheinbar eigene Vorarbeiten und zeitraubendes Graben in Archiven. Da das Interview selber schon eine öffentliche Verlautbarung und damit eine Nachricht ist, die über den Sender gehen kann, funktioniert es auch im Tagesjournalismus (Bezeichnenderweise verlassen sich aber erfahrene Journalisten nie auf Interviews, sondern werten immer parallel prozessproduzierte Daten aus).

Aber was macht das Interview zu einem Königsweg in den Sozialwissenschaften? Insbesondere alle Interviews, die nach den Hergang von Ereignissen, nach den damaligen Meinungen und Einstellungen beteiligter Akteure, nach damaligen Netzwerken, nach damaligen Entscheidungsprozeduren fragen, müssen zeigen, dass sie einen privilegierten

teren Zugang zum fremden Innersten, insbesondere im Rückblick, verschaffen als das Studium zeitgenössischer Dokumente. Verläufe sind Resultate von Handlungen, die nur zum Teil auf Intentionen und Einschätzungen zurückgehen. Hinterher, zum Zeitpunkt des Interviews, ist jeder klüger, zumindest anders.

In der Quantenphysik nicht weniger als in den Sozialwissenschaften rührt eine strukturelle, unüberwindliche Zweideutigkeit und »Unschärfe« im Sinne Heisenbergs daher, dass der Beobachter das Beobachtete zwangsläufig beeinflusst. Er kann nie das Objekt direkt messen, sondern versucht – mit mehr oder weniger Glaubwürdigkeit – den eigenen Einfluss aus seinen Daten herauszurechnen – wenn er nicht besser gleich das Beobachtete als Resultat auch seines Eingreifens interpretieren mag. Was als Interview zu Buche schlägt, ist ein hochselektives Handlungsprotokoll der Kooperation von Interviewer und Interviewten – ein Experiment mit z.T. ungenau dokumentierten Experimentalbedingungen (vgl. ausführlicher Behrens 1983). Interviewprotokolle, die selber als Protokolle von Quasi- oder natürlichen Experimenten mit sehr spezifischer Indikation verstanden werden, wie das narrative (Schütze in diesem Band), das szenische (Behrens 1983), das Dilemma-Interview (Hopf 1994), das kontrastive und die Gelegenheit zur Abgabe einer Meinung (Behrens 1983) haben so eine größere Plausibilität als Interviews in einem erkenntnistheoretisch naivem Realismus. Man kann einen Interviewerbias nicht vermeiden durch Interviewverhalten. Auch wenn wir Interviews noch so freundlich-zurückhaltend durchführen, können wir nicht vermeiden, ein Alter zu haben, ein Geschlecht und ein Aussehen. Das reicht schon als Einflussnahme. Man kann einen Intervieweinfluss nur nutzen. Er schlägt sich in jedem Fall ansatzweise nachvollziehbar und prüfbar auf einem Tonband nieder (vgl. ausführlicher Behrens 1983).

Hieraus ergeben sich zwei Vorschläge, deren Begründung im nächsten Teil dieses Aufsatzes systematischer diskutiert wird. Wir sollten uns immer fragen, ob nicht prozessproduzierte Daten eher die erste Wahl sind. Und wir sollten darauf achten, Interviews so zu protokollieren, dass wir sie wie prozessproduzierte Interaktionsprotokolle – sequenzanalytisch und an Falsifizierbarkeit orientiert – auswerten können.

Quellen sind Historikern umso verdächtiger, je mehr sie als Erinnerungen abgefasst vorliegen und je weniger als zeitnah gefertigte Protokolle. Von der Quellenkritik können qualitative Forscher viel lernen. Prozessproduzierte Daten haben gegenüber Interviewdaten den Vorteil, dass es keinen durch den Forscher verursachten reporting bias gibt. Selbstverständlich sind sie geprägt durch den Zweck, zu dem sie erstellt sind. Sie stellen an Forscher eine Qualifikationsanforderung, deren Erfüllung nicht einfach vorausgesetzt werden kann: Forscher müssen mit dem Prozess der Entstehung solcher Daten vertraut sein (vgl. ausführlich den Sammelband v. Ferber/Behrens 1996). Gerade in der Gesundheits- und Pflegeforschung gibt es Gelegenheiten, ein Tonband mitlaufen zu lassen oder sogar eine Kamera, die es in anderen Arbeitssituationen nicht gibt. Situationen in der Klinik zeichnen sich nämlich vor anderen Situationen dadurch aus, dass die Klienten ohnehin mit einer Fülle von Beobachtungsinstrumenten umgeben sind. Ein von Tonbandgerät und Kamera aufgezeichnetes Kooperationsprotokoll sagt mehr über die Kooperation als ein nachträgliches narratives Interview, so unverzichtbar dieses ist, wenn es um die nachträgliche, auf den Interviewer bezogene Erzählung oder aber Meinungsdarstellung geht. Der Vorteil der Grounded Theory (Strauss/Corbin 1990, Cor-

bin/Hildenbrand 2000) ist in dieser Hinsicht darin zu sehen, dass alle verfügbaren Daten in die Auswertung einbezogen werden. Auch Verfahren der objektiven Hermeneutik konzentrieren sich nicht auf bestimmte Verfahren der Datenerhebung, sondern werten tendenziell jede Art von textförmigen Interaktionsprotokollen aus (vgl. z.B. Allert 1998, Oevermann/Simm 1985).

Was heißt, Interviews wie prozessproduzierte Daten sequenzanalytisch auszuwerten? Zu offenen und wenig standardisierten Interviews haben Hoag und Allerbeck bereits 1981 geschrieben,

»Es wäre ein Irrtum, zu glauben, dass die [...] gezeigten Situationseffekte durch Verzicht auf Standardisierung und Quantifizierung verschwänden« (Hoag und Allerbeck 1981: 425). Unzutreffend ist aber ihr nächster Satz: »Der einzige Unterschied wäre, dass sie bei qualitativer Datenerhebung nicht mehr erfassbar wären«. Im Unterschied zu standardisierten Verfahren werden Situationseffekte bei all den 'qualitativen' Verfahren, die Tonbandaufnahmen verwenden (und das sind die meisten) wenigstens protokolliert. In diesen auswertbaren Handlungsprotokollen schlagen sich auch Äußerungen nieder, die rekonstruierbar machen, wie die Befragten Fragen verstanden und Strukturierungen sich entwickelten. Nicht das möglichst unvoreingenommene Interview, sondern die im nächsten Abschnitt zu diskutierende sequenzanalytische Auswertung des Interaktionsprotokolls erzeugt mit der Falsifikation von Vorverständnissen, Lesarten und Hypothesen Objektivität. Unvoreingenommen sind wir nicht immer schon, sondern Unvoreingenommenheit stellt sich immer nur näherungsweise her durch harte Arbeit an der Falsifikation unserer Vorverständnisse.

3 Sinn machen quantitative Untersuchungen nur als Teil qualitativer Studien – auf dem Wege zu einem Überblick

Eine solche Orientierung der qualitativen Auswertung offener Interviews an der Entdeckung von Lesarten und der Falsifikation von Vorverständnissen und Hypothesen lässt sich keineswegs in den programmatischen Erklärungen aller qualitativen Sozialforscher erkennen. Programmatisch wird die Möglichkeit von Falsifikation von radikalen Konstruktivisten geradezu bestritten. Dass qualitative und quantitative Forschungen sich nicht durch ihre Gütekriterien unterscheiden, sondern durch die Strategien, mit denen sie gemeinsame Gütekriterien erfüllen wollen (Kelle/Kluge 2001 nennen diese unterschiedlichen Strategien zum selben Ziel »Validierungsstrategien«), ist keineswegs Allgemeingut. Statt dessen herrschen drei verbreitete Entgegensetzungen vor und begrenzen die Frage nach der Indikation:

- die Entgegensetzung induktiver und deduktiver Verfahren als unterschiedliche Verfahren, nicht als unterschiedliche Abschnitte derselben Verfahrenskette;
- die Entgegensetzung völlig unterschiedlicher Gütekriterien quantitativer und qualitativer Verfahren;
- die Gleichsetzung bestimmter Validierungstechniken mit den Gütekriterien selber.

3.1 Induktive und deduktive Verfahren als alternative, nicht als unterschiedliche Abschnitte derselben Verfahrenskette?

Seit Jahrzehnten werden in Lehr- und Handbüchern ein quantitatives und ein qualitatives Konzept des Verhältnisses von Theorie und Gegenstand und der entsprechenden Verfahrenskette gegenübergestellt (vgl. z.B. die sich durch Klarheit und Gründlichkeit auszeichnende Darstellung bei Flick schon 1994: 172)

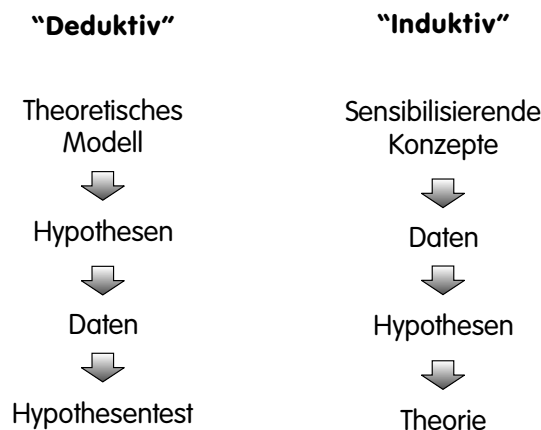


Abbildung 2: Induktiv versus deduktiv

Bei näherem Hinsehen zeigt sich allerdings, dass es sich bei diesen seit Jahrzehnten einander entgegengesetzten Verfahrensketten gar nicht um zwei unterschiedliche Ketten, sondern im Gegenteil lediglich um zwei unterschiedliche Abschnitte ein- und derselben Kette handelt. Woher kommen denn theoretischen Modelle und Hypothesen der quantitativen Kette, wenn nicht aus den sensibilisierenden Konzepten und in ihrem Licht durchgeführten früheren Studien der qualitativen Kette? Wie werden denn Konzepte und Hypothesen durch Daten erweitert, wenn nicht durch ihre Falsifikation? Die ganze Kette, von der die »quantitative Kette« nur ein Teil ist, sähe also etwa so aus:

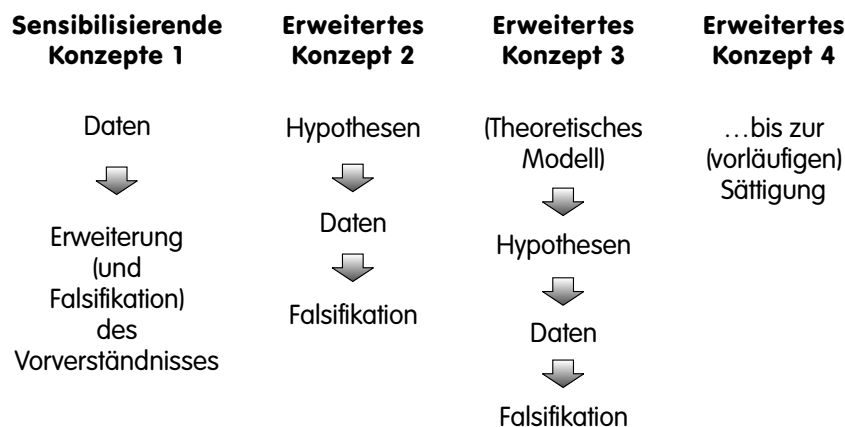


Abbildung 3: Verfahrenskette zum Verhältnis von Theorie und Gegenstand

Forschung – in den Naturwissenschaften wie in den Sozialwissenschaften – beginnt ursprünglich immer mit sensibilisierenden Konzepten, Erwartungen an Verläufe in der Welt, von denen aus Unterschiede zwischen Fällen zum Problem für Zusammenschauen (solche Zusammenschauen bezeichneten die Griechen als »Theorien«) werden. Solche Verläufe in der Zeit sind Strukturierungen. Auf den zentralen Begriff der Strukturierung ist gleich zurück zu kommen. Ohne sensibilisierende Konzepte, ohne Erwartung nähmen wir gar nichts wahr (insofern bestätigt die Hirnforschung den erkenntnistheoretischen Konstruktivismus, demzufolge unsere Sinneseindrücke erst durch unsere Modelle zu Häusern und Bäumen geordnet und wahrnehmbar werden).

Mit unerwarteten Unterschieden können wir aber auf zwei Weisen umgehen:

- Immunisierung:

Wir können unsere Erwartungen gegen unsere Erfahrungen immunisieren, etwa mit dem beschwörendem Spruch »Keine Regel ohne Ausnahme« oder gleich mit dem Aufstellen so allgemeiner »Theorien«, dass kein Ereignis sie widerlegen könnte (eine solche unwiderlegbare Schein-Theorie ist die in der Sozialwissenschaft Ökonomie weit verbreitete Nutzen-»Theorie«, demzufolge Menschen nach ihrem Nutzen streben: Solange ich mich nicht festlege, was als Nutzen gelten soll, kann buchstäblich jede beliebige Handlung nachträglich als Verwirklichungsversuch einer Nutzenvorstellung einsortiert werden). Eine dritte beliebte Immunisierungsstrategie besteht darin, mich nicht auf eine Beobachtungsstrategie festzulegen, sondern jedes Ergebnis, das mir nicht in die Theorie passt, als »falsch gemessen« hinzustellen. Vollkommene Immunität habe ich natürlich dann erreicht, wenn ich den Eintritt der Ereignisse, die aus meinen Erwartungen folgen müssten, als völlig unbeobachtbar behaupte.

- Falsifikationsversuch:

Wir können unsere Erwartungen und die Beobachtungsregeln so präzise formulieren, dass die Erwartungen durch Erfahrungen widerlegt werden können, ja, wir können geradezu nach widersprechenden Erfahrungen suchen. Nur dann können Erfahrungen Erwartungen widersprechen und dazu führen, dass eine Theorie oder ein theoretisches Konstrukt verändert werden müssen. Diesen falsifikationsorientierten Umgang finden wir ebenso in quantitativen wie in qualitativen, in Hypothesen testenden Verfahren wie in Verfahren der Grounded Theory.

Eine Hypothese ist eine Behauptung, die so präzise formuliert ist und deren beobachtbare Falsifikationsbedingungen soweit festgelegt sind, dass die Behauptung widerlegt werden kann (Die in der Ökonomie verbreitete Nutzen-Theorie z. B. ist nicht präzise genug formuliert, um eine Hypothese abzugeben, bestenfalls kann sie als vages Weltkonzept angesehen werden, das uns bei der Formulierung prüfbarer Hypothesen motiviert.) Wenn eine Hypothese widerlegt wird, die als untrennbarer Teil einer Theorie formuliert war, führt die Widerlegung zu einer Änderung, also einer Präzisierung oder Erweiterung der Theorie. Die veränderte Theorie wird wieder in testbaren Hypothesen prüfbar. Dieser Falsifikations- und Theoriebildungsprozess wiederholt sich viele Jahre und

Jahrzehnte – bis die beteiligten Forscher den Eindruck haben, es fänden sich nun keine Erscheinungen mehr, die mit der Theorie unvereinbar wären. Die Theorie erscheint der Gemeinschaft der Forschenden gesättigt. Von weiteren Experimenten und Untersuchungen erwarten sich die Forscher wenig Neues. Der Nachwuchs widmet sich der Anwendung der gesättigten Theorien und Konzepte – oder wechselt in andere Fächer, wo die Konzepte noch weit von ihrer empirischen Sättigung entfernt sind. Manchmal trägt dieser Eindruck der Sättigung. Am Ende des 19. Jahrhunderts erschienen vielen Physikern weite Teile der physikalischen Theorie als empirisch gesättigt. Die großen empirischen Falsifikationen waren scheinbar erfolgt. Das Fach selber wurde für junge Leute, die überraschende Entdeckungen und Ergebnisse publizieren wollten, uninteressant. Doch dann ging es noch einmal richtig los. Der Eindruck der Sättigung hatte getrogen.

Auch in der Grounded Theory und in der objektiven Hermeneutik wie in anderen »qualitativen« Verfahren, die sich nicht nur dem Archivieren widmen, ist die Suche nach widersprechenden Erfahrungen deutlich, ja für das Verfahren konstitutiv. Das wird in der »Grounded Theory« unverkennbar deutlich an der Auswahl der Untersuchungsfälle erst nach geringen, dann nach maximalen Unterschieden zum bereits untersuchten Fall (Kodierregeln und Theoretisches Sampling, vgl. Strauss/Corbin 1990). Die aus der Auseinandersetzung mit den ersten Fällen gewonnenen Konstrukte werden an weiteren Fällen bewusst dem Risiko ausgesetzt, sich als nur begrenzt gültige und zu erweiternde Konstrukte zu erweisen. Viel Nachdenken und Kreativität sollen die Forschenden darauf verwenden, Fälle zu finden, die den bisherigen Konzepten möglicherweise widersprechen könnten – die Forscher sollen gerade nicht nach »bestätigenden« Fällen suchen. Erst wenn sich trotz angestrengten Suchens keine Fälle mehr finden lassen, die die Konzepte falsifizieren können, ist eine »Sättigung« erreicht und die Untersuchung zur Theoriebildung wird (zunächst) abgeschlossen. Die Ähnlichkeit zur eben genannten Entwicklung der Physik ist deutlich. Grounded Theory beschreibt eine Kette der Theoriebildung, in dem die Theoriebildung nicht mit dem Test einer einzigen Hypothese abgeschlossen ist. Vielmehr führt jede Falsifikation einer hypothetischen Erwartung zu einer theoretischen Neukonzipierung, die wieder dem Falsifikationsrisiko ausgesetzt wird. Und da *capo al fine*, bis Sättigung erreicht ist.

Demselben Falsifikationismus – bei allen sonstigen Unterschieden – folgt die objektive Hermeneutik in ihrer Sequenzanalyse, die Ähnlichkeiten mit der line by line Analysis von Anselm Strauss aufweist. Bei der Analyse eines Handlungsprotokolls, das als Text vorliegt, geht es darum, an den ersten Textsequenzen möglichst präzise und falsifizierbar solche hypothetischen Lesarten des Textes, also falsifizierbare Fallstrukturhypothesen zu entwickeln, die dann durch den Fortgang des Textes widerlegt werden können. (Diese Falsifikationsverfahren funktionieren besser bei einigen sozialen Arrangements. Zu ihnen gehören unter anderem: Die Interpreten müssen hart am Text kreativ und streitbar Fallstrukturhypothesen erarbeiten können und möglichst wenig Neigung haben, Autoritäten nach dem Munde zu reden. Die Interpreten sollten nicht das Ende des Handlungsprotokolls kennen, wenn sie den Beginn interpretieren, weil ihnen dann bei Beginn schwerlich alle möglichen Fortsetzungen hypothetisch einfallen, die dem Text zu Anfang innewohnen. Die konkurrierenden Hypothesen dürfen nicht so vage formuliert sein, dass sie alle mit dem Text vereinbar und daher nicht falsifizierbar sind. Denn un-

überprüfbare Hypothesen sind für die interpretatorische Erschließung ziemlich wertlos, gerade deshalb, weil sie sich nicht als falsch erweisen können. (Zum Vorgehen vgl. die materialen Analysen von Allert 1998, Oevermann/Simm 1985.)

Die Sequenzanalyse wie auch die line-by-line Analyse folgen einem methodischem Prinzip, das ähnlich auch den prospektiven Verlaufsanalysen zugrunde liegt, wie es für die sogenannte quantitative klinische Forschung typisch ist. Das Prinzip der Sequenzialität ist mit dem Phänomen der Strukturierung als Individuierung eines sozialen Gebildes, einer Person in der Zeit begründbar. Aus allen möglichen Fortsetzungen einer Handlung oder eines Ereignisses realisiert sich immer nur eine Auswahl. Diese Auswahl ist nicht zufällig, sondern folgt erkennbaren Regeln. Die hypothetische prognostische Fortschreibung einer Sequenz dient der Erkenntnis *ex negativo*, welche Auswahl-Regel durch den tatsächlichen Verlauf (noch) nicht falsifiziert wurde. Sie rekonstruiert das So und nicht anders Gewordensein.

An dieser Stelle kann nicht weiter auf Vorgehen und Unterschiede der Verfahren eingegangen werden (vgl. für materiale Analysen Allert 1998, Hildenbrand 1990, Oevermann/Simm 1985, für Methodenvergleiche Wagner 1999, Behrens 1980 und Behrens/Langer 2002, Meinefeld 1997). Deutlich sollte nur geworden sein, dass sich die genannten falsifikatorischen Verfahren von solchen Verfahren unterscheiden, die entweder Texte/Handlungsprotokolle nach Konzept-bestätigenden Stellen durchmustern oder die Forscher in die Lage versetzen sollen, sich induktiv wie unbeschriebene Blätter beschreiben zu lassen. Streng induktive Verfahren kann es nicht geben.

Für diese falsifikatorisch orientierten Verfahren relativiert sich offenbar die Entgegensetzung von deduktiv quantitative Hypothesen testenden und induktiv qualitative Konzepte generierenden Verfahren. Sinn machen quantitative Tests nur als Teile qualitativer Konzeptklärungen. Die Nutzbarkeit externer Evidence für interne Arbeitsbündnisse zwischen Klienten und Professionen hinge dann auch daran, wieweit die wissensproduzierenden Falsifikationsprozesse nachprüfbar dokumentiert sind. Das wäre ein entscheidendes Indikationskriterium. Die Betonung des Anspruchs auf Falsifizierbarkeit und die erkenntnistheoretische Position des fallibilistische Realismus (i.S. etwa von Lakatos 1982) sind nun allerdings nicht Allgemeingut der Ansätze, die sich als »qualitative« verstehen. Nicht nur in verbreiteten Pflegelehrbüchern, sondern auch in speziellen Methodenbüchern findet sich das glatte Gegenteil. Qualitative und quantitative Ansätze hätten unterschiedliche Gütekriterien, und gerade der Anspruch auf Falsifizierbarkeit unterscheidet sie.

3.2 Entgegensetzung der Gütekriterien qualitativer und quantitativer Verfahren?

Yvonn Lincoln und Egon Guba (1985), weithin bekannte Autoritäten in der qualitativen Evaluationsforschung, haben besonders explizit den Anspruch auf Falsifizierbarkeit abgelehnt. Für einen radikalen Konstruktivisten seien alle Darstellungen der Welt gleich gültig (vgl. Lincoln/Guba 1985). Sie können nicht an so etwas wie der »Realität« überprüft werden.

»There is no ultimate benchmark to which one can turn for justification – whether in principle or by a technical adjustment via the falsification principle.« (Lincoln/Guba 1985, 295)

Folgerichtig würde man, wenn einem die Sichtweise eines Kollegen nicht zusagt, keinen Kurs über qualitative und quantitative Methoden des Evidence-based Nursing durcharbeiten. Stattdessen würde man ausschließlich in einen Rhetorik- Kurs oder in einen Kurs für Schriftsteller gehen. Diese Vorliebe für Rhetorik findet man nicht nur im Leben, sondern tatsächlich auch in der Methoden- Literatur: Angesichts von »multiple, equally valid worldviews« verlieren für radikale Konstruktivisten Begriffe wie Fehler oder Irrtum ihren Sinn (vgl. Lincoln/Guba 1985). Wenn Forschungsergebnisse weder falsch noch richtig sein können, werden Ergebnisse immun gegen Kritik. (»Eine solche Immunisierung von Forschungstätigkeit und Forschungsergebnissen gegen Kritik mag manchem als verführerische Möglichkeit erscheinen angesichts der zahlreichen Probleme, die sich bei Beschreibung und Erklärung sozialer Probleme stellen, verlangt aber einen hohen Preis«, schreiben Susann Kluge und Udo Kelle (2001):

»Denn letztendlich verliert die empirische Sozialforschung damit die Möglichkeit, gegenüber anderen Formen von Wissensproduktion besondere Geltungsansprüche anzumelden. Die Ergebnisse einer qualitativen Studie über die Lebenswelten von Aussiedlern könnten dann letztlich keine größere Validität beanspruchen als Aussagen, die irgend jemand nach zehn Gläsern Bier am Stammtisch trifft«. (Kluge/Kelle 2001: 144f.)

Wir schätzen sowohl die Bildung von Alltagstheorien in Trinkgelagen (»Symposien«) als auch die Erfahrungen freisetzende und Zungen lösende Wirkung von Alkohol hoch, stimmen aber Kluge und Kelle zu, dass dabei bestenfalls Experten-Ratings und keine externe evidence erzeugt wird. (Uta Gerhardt nennt das, was wir als Rhetorik bezeichnen, »Didaktik« (Gerhardt 1985: 238). Didaktik ist mehr.) Wir raten keineswegs davon ab, einen Rhetorik-Kurs zu besuchen. Aber die Prüfung von Vorverständnissen an der Realität bedarf anderer Methoden als der Methoden rhetorischer Überrumpelung. Jeder historische Bericht ist auch eine Erzählung. Aber nicht jede Erzählung ist ein historischer Bericht. Es gibt – im Unterschied zur Ansicht des radikalen Konstruktivismus – einen Unterschied zwischen Science und Science-Fiction – gerade wenn keine Methode sich von selbst versteht, sondern selber begründet werden muss (zu dieser Reflexivität vgl. Lakatos 1982, Gadamer 19723).

Dabei legen wir Wert auf die Feststellung, die Kelle und Kluge an dieser Stelle nicht hervorheben, dass auch Fiction wie Romane, Theaterstücke und Filme ebenfalls offensichtlich einem Validitätskriterium des Realismus unterliegen. Die Story, die Dialoge, die Gesten müssen – wenn auch gerne etwas verfremdet und exotisiert – hinreichend den Regeln unserer Lebenswelt entsprechen, dass wir den Eindruck haben, genau so könnte es sich abspielen. Sonst würden wir Schauspieler als schlecht, nämlich viel zu unglaubwürdig und gestelzt ausbuhnen, Bücher als langweilig, weil viel zu konstruiert beiseitelegen und den Fernseher umschalten. Es ist keineswegs beliebig, was wir uns erzählen

lassen. Die Regeln der Glaubwürdigkeit bei Fiction sind so streng, dass mancher Zeitungsbericht über ein »tatsächliches Ereignis« nie als Roman durchginge. Gerade wenn wir wissen, dass etwas nur ausgedachte Fiction ist, muss es besonders glaubwürdig sein. Wegen der Strenge dieser Glaubwürdigkeitsregeln eignen sich Romane, Theaterstücke und Filme auch als Gedankenexperimente, an denen wir durchspielen können, was wir von der Welt wissen und wünschen. Die differenzierte Diskussion über unsere Lebensprobleme führen wir vielleicht weniger in wissenschaftlichen Untersuchungen als in belletristischen Gedankenexperimenten. Auch an Werken des Surrealismus ließe sich bei etwas mehr Platz zeigen, dass ihre Wirkung gerade daher rührt, dass sie uns im Kern bekannt und vertraut vorkommen.

Wenn wir zurückkommen auf die für mich wesentliche Differenz von interner Evidenz, wie sie sich im Arbeitsbündnis zwischen Klienten und Professionen herstellt, und der externen Evidence nicht selbst gemachter Erfahrung, auf die wir bei unseren Entscheidungen im Arbeitsbündnis zurückgreifen, lässt sich die große Bedeutung von Fiction würdigen. Wahrscheinlich ziehen wir Belletristik und Filme sogar häufiger heran als Forschungsberichte, wenn wir uns über uns selber verständigen und unsere Dilemmata gedankenexperimentell durchspielen wollen. In Romanen finden wir im glücklichen Fall etwas von uns besser ausgedrückt, als wir es uns selbst hätten sagen können. Aber die meisten von uns würden doch genau wissen wollen, ob der Bericht von einer erstaunlichen Heilung ein Roman ist oder ein Forschungsergebnis. Wir unterscheiden zwischen der Glaubwürdigkeit, Plausibilität und Wahrscheinlichkeit von Fiction und der Validität von Untersuchungen. Allerdings – und das unterscheidet uns vom naiven Realismus, den Lincoln und Guba (1985) zu recht kritisieren – lässt sich durch Forschungsergebnisse nie die ganze und unverbrüchliche Wahrheit feststellen. Schon die beliebte Formulierung »Diese Forschung bestätigte die Hypothese X« ist unhaltbar naiv realistisch. Das Beste, was wir sagen und wissen können, ist, eine Ansicht habe sich durch eine Untersuchung (noch) nicht als falsch erwiesen. (»At best, one can know what has not yet been ruled out as false«, schreiben Cook/Campell 1979, S. 37) Diese Einsicht bezeichnet Lakatos 1982 mit dem in der Tat griffigen Begriff »fallibilistischer Realismus«. Alles Wissen ist vorläufig – das gilt für jede evidence-basierte Leitlinie und jeden Standard. Wir können uns täuschen, aber wir können unsere Täuschungen durch Untersuchungen auch als falsch erweisen.

Lincoln und Gubas Begriff der »holistischen Generalisierung« kommt hingegen das Verdienst zu zu zeigen, was passieren kann, wenn man sich vom Falsifikationismus abkoppeln, aber Forschungsberichte statt Romane schreiben will. Statt zufällig ausgewählter Stichproben erlaubten, schreiben sie, solche Untersuchungsgruppen eine »holistische Generalisierung«, bei denen

»any part or component is a perfect sample in the sense that it contains all the information about the whole that one might ever hope to obtain [...] imperfect (blurred) information from any source can be improved (clarified), if one has the appropriate filters so doing« (Lincoln und Guba 1985: 128).

Das ist weit mehr, als jeder Statistiker je zu hoffen wagte. Aber wer hat diese angemessenen Filter, wie können solche perfekten samples gezogen werden, die alle Informatio-

nen über ihre Ausgangsgruppe enthalten? Empfehlungen, sobald sie konkret werden, ähneln eher der konventionellen »quantitativen« Sozialforschung, z. B. die empfohlene Suche nach Gegenevidenz (also Falsifikation?). Auch der »member check« (in Deutschland bekannt als kommunikative Validierung), bei dem »data, analytic categories, interpretation, and conclusions are tested with members whom the data were originally collected« (Lincoln/Guba 1985, S. 314) setzt als Test eine eher naiv realistische erkenntnistheoretische Position voraus als eine radikal konstruktivistische. Ein solcher Check ist nämlich nur aussagekräftig, wenn mindestens drei Identitäten vorausgesetzt werden sollen

- die Sichtweisen der Untersuchten zur Zeit der Untersuchung und nach der Interpretation
- die analytischen Kategorien und Interpretationen der Forscher und die Sicht der Untersuchten
- die Selbstdarstellung der Untersuchten gegenüber den Untersuchern und ihre eigene Sicht.

So weitgehende Übereinstimmungen anzunehmen wäre naiv. Wenn aber die Sprecher und Erzeuger der produzierten Texte/Daten nur weitere nachträgliche Interpreten dieser Daten sind (wie von Schütz am Selbstverstehen als Fremdverstehen gezeigt), würde der member check keine herausgehobenere hypothetische Lesart des Textes erzeugen können als jeder andere Interpret.

Dieses Schwanken zwischen naiv realistischen und radikal konstruktivistischen Positionen scheint mir nicht zufällig. Die Autoren führen, schreiben Kelle und Kluge, »die realistische Erkenntnistheorie, die sie quasi aus der Vordertür herausgeworfen hatten, durch verschiedene Hintertüren wieder ein und präsentieren letztendlich konventionelle Wege der Validitätssicherung« (Kelle/Kluge 2001:141). Man könnte begrüßen, dass sie wenigstens durch die Hintertür hereinkommen. Aber das Reflexionsniveau sinkt. Die Gegenüberstellung von »Validität« und »Glaubwürdigkeit« ist dann mit dem Risiko verbunden, dass die Bereiche wechselseitig vor Reflexion geschützt werden.

So unterschiedliche qualitativ- quantitative Methoden wie die Grounded Theory, die objektive Hermeneutik, Experimente und viele andere, haben den Anspruch gemeinsam falsifikatorisch vorzugehen (wie ich im letzten Abschnitt zu zeigen versuchte). Dieser Anspruch wird nicht immer erfüllt. Aber eine naiv realistische oder radikal konstruktivistische Untersuchung ohne den Anspruch auf ein falsifikatorisches Vorgehen wird schwerlich evidence erzeugen können. Die radikal konstruktivistische Rhetorik kann dann den Gartenzaun abgeben, hinter dem der naive Realismus blüht.

3.3 Gleichsetzung von Gütekriterien mit Techniken

Wenn man bestimmte Validierungstechniken mit dem Gütekriterium gleichsetzt, können qualitative und quantitative, aber auch Beobachtungs- und Experimentalstudien niemals in ihren spezifischen Stärken verglichen werden. Eine eindimensionale Rangfolge der Qualität von Studien, wie sie in der Evidence based Medicine zur Folklore gehört, könnte es nur geben, gäbe es nicht mehr als eine Täuschungsmöglichkeit. Wenn es mehrere

Täuschungsgefahren, also mehrere Gefährdungen der Validität gibt, gibt es nicht eine Rangfolge guter und weniger guter Studiendesigns, sondern so viele Rangordnungen, wie nicht gemeinsam zu kontrollierende Fehlermöglichkeiten. Nicht alle Fehlerquellen können von uns durch dieselbe Vorsichts-Maßnahme berichtigt werden.

Dass wir uns täuschen können, aber Täuschungen auch falsifizieren können, ist der erkenntnistheoretische Kern eines fallibilistischen Konzepts von Gütigkeit oder »Validität«. Wenn es nicht nur eine Möglichkeit der Selbsttäuschung gibt, bedarf es einer Auflistung von Fehlerquellen (sogenannte Fehlertheorien) und der Methodologien der Fehlervermeidung. Solche Methodologien sind sowohl in der qualitativen als auch in der quantitativen Forschung begonnen worden – wenn auch in der Regel getrennt voneinander und gern in wechselseitiger Ignoranz und Revierabgrenzung. Wechselseitige Ignoranz und Revierabgrenzung ist irreführend, anstrengend und zeitraubend, wenn der erkenntnistheoretische Kern doch derselbe ist. In der quantitativen Forschung lassen sich die statistische Hypothesenprüfung und die Quasi-Experimentation (Cook/Cambell 1979), in der qualitativen Methodologien einige Versuche in der Grounded Theory und Auswertungsprozeduren der objektiven Hermeneutik als Beispiele für Fehlerlisten und Falsifikationsstrategien nennen.

Maßnahmen gegen verzerrte Stichproben, die auf unterschiedlichen Dimensionen wirken und in der Regel nicht gleichzeitig umzusetzen sind, sind z.B.

- zufällige Zuordnungen zu Experimenten
- repräsentative Stichproben ganzer Bevölkerungen oder Vollerhebungen,
- hochstandardisierte Messinstrumente und Befragungs-Instrumente,
- gering standardisierte Mess- und Befragungsinstrumente, um Informationen darüber zu erhalten, wie Befragte die Fragen verstanden (vgl. Behrens 1983: 204 ff, 244).

Diese Maßnahmen sind in der Regel nicht gleichzeitig zu haben. Zwar könnte einer sagen, eine gute Studie müsse alle vier Maßnahmen enthalten. Eine Studie mit allen 4 Maßnahmen sei der Gold-Standard, an der sich alle anderen Studien zu messen hätten. Wenn eine Studie den Gold-Standard erfülle, brauche man die anderen gar nicht mehr zu lesen. Offensichtlich lassen sich in der Praxis aber so gut wie nie alle vier Maßnahmen gleichzeitig durchführen. So ist es zwar denkbar, dass repräsentative Stichproben von Bevölkerungen – z.B. aller Europäer über 18 Jahren – zufällig sich einer Experimental- und einer Vergleichsgruppe zuordnen ließen. Praktisch kommt das fast nie vor. Experimentalstudien, insbesondere sog. randomisierte Kontrollstudien (»RCT«.), werden fast nur mit solchen Kranken oder Pflegebedürftigen durchgeführt, die Patienten oder Klienten bestimmter Kliniken und Pflegeeinrichtungen geworden sind. Diese repräsentieren keineswegs hinreichend auch die Kranken und Pflegebedürftigen, die nicht in die Kliniken und Einrichtungen der Untersuchung gingen. Selbst wenn die Alters- und Geschlechtsverteilung der Patienten der Untersuchungs-Klinik und der Untersuchungs-Einrichtung mit den anderen übereinstimmte, wäre keineswegs von einer unverzerrten, repräsentativen Stichprobe zu reden. Genau das Merkmal »Aufsuchen gerade dieser Kliniken und

Einrichtungen« und dessen Gründe kann die untersuchte von der nicht untersuchten Gruppe wesentlich unterscheiden. Insofern werden fast alle RCTs auf der Basis verzerrter Stichproben durchgeführt, wenn man etwas über alle Gleichkranken oder Gleichpflegebedürftigen einer gesamten Bevölkerung erfahren oder sagen wollen. Gemessen an diesem Gold-Standard sind RCTs keine vollkommenen Umsetzungen unverzerrter Stichproben. Umgekehrt lassen sich ganze Bevölkerungen bzw. ihre repräsentativen Stichproben nur selten zu vertretbaren Kosten auf Experimente ein. Erhebungen individueller Verläufe in Bevölkerungen liegen durchaus vor – wie z.B. im sozioökonomischen Panel Deutschlands und vieler anderer Länder. Da diese Erhebungen z.T. öffentlich zur Verfügung stehen, kann jeder sie mit einiger Mühe, aber geringeren finanziellen Kosten auch selbst auswerten. Aber man kann hier quasi-experimentell in multivariablen Analysen nur den Einfluss derjenigen Ereignisse messen, die ohne Zutun eines Forschers auftraten und erhoben wurden. Man muss die Welt als *experimentum mundi* deuten und nicht sich selber als Experimentator (Behrens/Langer 2002).

RCTs und multivariable Analysen individueller Verläufe ganzer Bevölkerungen sind also zwei Umgangsweisen mit der Täuschung durch verzerrte Stichproben, die sich ergänzen und leider faktisch nur sehr selten zusammen genutzt werden können. Da sie sich ergänzen, kann auch nicht unabhängig von der Problemstellung ein für alle Mal gesagt werden, welche besser ist. Beide sind für Therapiestudien, Diagnosestudien, Nebenwirkungsstudien usw. unverzichtbar.

Aber nicht nur RCTs und multivariable Analysen individueller Verläufe ganzer Bevölkerungen, sondern auch hochstandardisierte und offenere Messinstrumente sind Maßnahmen der Irrtumsvermeidung, die sich ergänzen, aber sich selten gleichzeitig nutzen lassen.

Hochstandardisierte Messinstrumente sind allgemein bekannt von Fragebögen mit festen Antwortvorgaben zum Ankreuzen, offenere von Tonbandinterviews, bei denen die Befragten selber ihre Antworten formulieren müssen und können. Bei Fragebögen zum Ankreuzen ist es offenbar leichter sicherzustellen, dass überall derselbe Fragebogen genutzt wird, und die Antworten können über einen Scanner schon codiert automatisch eingegeben werden. Bei Tonbandprotokollen selbst formulierter Sätze hingegen wird dafür eher erschließbar, wie die Befragten die Frage überhaupt verstanden und in ihren Kontext eingeordnet haben, während die Forscher den Kreuzchen kaum entnehmen können, wie der Befragte die Frage verstanden hat (deshalb lassen sich geschlossene Fragebogen auch leichter durch Interviewer fälschen als offene Interviews). Daher haben Hoag und Allerbeck 1981, wie wir eingangs sagten, recht, wenn sie schreiben: »es wäre ein Irrtum, zu glauben, dass die gezeigten Situationseffekte durch Verzicht auf Standardisierung und Quantifizierung verschwinden« (Hoag und Allerbeck 1981: 425). Unzutreffend ist aber ihr nächster Satz: »Der einzige Unterschied wäre, dass sie bei qualitativer Datenerhebung nicht mehr erfassbar wären«. »Im Unterschied zu standardisierten Verfahren werden Situationseffekte bei all den ›qualitativen‹ Verfahren, die Tonbandaufnahmen verwenden (und das sind die meisten) wenigstens protokolliert«. In diesen auswertbaren Handlungsprotokollen »schlagen sich auch Äußerungen nieder, die, eben weil sie von den Sprechern nicht bewusst kontrolliert werden können, Ausdruck durchgehender Haltungen sind und rekonstruierbar machen, wie die Befragten Fragen verstanden und zum

Anlass von Strukturierungen machten. Man mache die Probe: jeder kann ein standardisiertes Interview fälschen; aber um ein qualitatives Interview zu fälschen, müssen schon ein großer Dichter und ein großer Schauspieler zusammenkommen.« (Behrens 1983: 244, vgl. neuerdings ebenso Kelle/Kluge 2001 für Interviews mit Altenheim-Bewohnern und mit straffälligen Jugendlichen).

Diese vier Beispiele zeigen, dass es nicht ein Studiendesign oder ein Interviewverfahren gegen alle Verzerrungsgefahren gibt. Die Maßnahmen zur Reduktion der jeweiligen Verzerrung erwiesen sich in der Praxis als zwar nacheinander (für eine solche Abfolge von Prüfungen vgl. ausführlich Behrens/Langer 2002), aber nicht gut gleichzeitig einsetzbar. Daher gibt es auch mehr als eine Rangordnung von guten und weniger guten Studien. Man muss sich entscheiden, welche Verzerrung man für die am meisten zu vermeidende hält. Für unterschiedliche zu vermeidende Verzerrungen ergeben sich unterschiedliche Hierarchien von Designs.

Ein gemeinsamer Bezugspunkt, ein allgemeines Gütekriterium für die unterschiedlichen Techniken der Validitätssicherung lässt sich im Falsifikationismus erkennen, der auf einen fallibilistischen Realismus (Lakatos 1982), aber auch auf andere erkenntnistheoretische Positionen zurückgeht. Jede Übertragung der operationalisierten Kriterien einer Validitätssicherungstechnik auf eine andere führt hingegen dazu, dass diese als völlig defizient erscheint, eben weil diese auf eine andere Validitätsbedrohung gerichtet war als jene.

Cook und Campbell haben 1979 eine (hauptsächlich auf experimentelle und quasiexperimentelle Studien bezogene) Taxonomie von »threats of validity«, validitätsbedrohender Fehler, aufzustellen versucht und zu jeder Bedrohung Strategien benannt, die die jeweilige Fehlerquelle beherrschbar machen sollte. Auch für die qualitative Feldforschung gibt es seit einiger Zeit implizit oder explizit das Programm einer solchen Taxonomie (vgl. Schon, Hopf 1994, Flick 1991, 1994, 2000, Kelle/Kluge 2001, Meinefeld 1997).

4 Schlussbemerkung

In diesen einführenden Bemerkungen ging es mir nicht darum, die 20 bis 30 Verfahren qualitativer Interviews vorzustellen oder gar ihre Qualität zu prüfen. Vielmehr muss jeder einleitende Aufsatz sich mit der Frage auseinandersetzen, ob ein Überblick entlang unterschiedlicher Indikationen (Indikationsstellungen sind jeder und jedem im Gesundheitswesen im Alltag selbstverständlich) für Untersuchungsverfahren bereits üblich ist. Das ist weniger der Fall, als ein Außenstehender erwarten würde.

Daraus ergab sich die Notwendigkeit, Hindernisse auf dem Weg zu einer indikationsbezogenen Übersicht auszumachen. Als solche Hindernisse erwiesen sich die reviermäßige Entgegensetzung von Verfahren, wo es sich nur um unterschiedliche Abschnitte der selben Verfahrenskette handelt sowie zu geringe Klarheit über die unterschiedlichen Ziele verschiedener Verfahren und vor allem

- die Entgegensetzung der Gütekriterien quantitativer und qualitativer Verfahren, wo quantitative Verfahren doch nur als Teile qualitativer Verfahren Sinn machen und

- die Gleichsetzung von Gütekriterien mit Techniken, als gäbe es nur eine Fehlermöglichkeit.

Solche Profilierungen, die der tatsächlichen Praxis der Verfahren gar nicht entsprechen (vgl. auch Meinefeld 1997), sind historisch dem Interesse an Schulbildung geschuldet. Sobald die Verfahren entfaltet sind, kann – und muss – dazu übergegangen werden, quantitative und qualitative Interview- und naturalistische Verfahren nicht mehr als wechselseitig völlig unverständliche, aber in sich völlig homogene Schulen darzustellen, sondern zwischen ihnen nach ihrer Indikation, also nach ihrer Eignung zur Bewältigung von Täuschungsmöglichkeiten und zur Produktion spezifischer Evidenz zu wählen.

Literatur

- Allert, Tilmann (1998): Die Familie. Fallstudien zur Unverwüstlichkeit einer Lebensform. Berlin/New York
- Barton, Allen H./Lazarsfeld, Paul F. (1955): Some Functions of Qualitative Analysis in Social Research. In: Frankfurter Beiträge zur Soziologie. Frankfurt am Main: Europäische Verlagsanstalt S. 321–361
- Behrens, Johann, Gero Langer (2002), Vertrauensbildende Entzauberung. Qualitative und quantitative Methoden für Evidence basierte Pflegepraxis, Bern: Huber
- Behrens, Johann (1983), »Bedürfnisse« und »Zufriedenheiten« als Statussymbole und Anrechte. Lehren aus einem Panel, in: Hondrich, Karl Otto, Vollmer, Randolph, Bedürfnisse im Wandel: Theorie, Zeitdiagnose, Forschungsergebnisse, Opladen: Westdeutscher Verlag
- Cook, Thomas D./Campbell, Donald T. (1979), Quasi-experimentation: Design & Analysis Issues for Field Settings. Boston: Houghten Mifflin
- Corbin, Juliet/Hildenbrand, Bruno (2000) Qualitative Forschung in: Rennen-Allhoff, B., Schaeffer, D. (Hrsg.) Handbuch Pflegewissenschaft, Weinheim: Juventa S. 159 – 184
- v. Ferber, Liselotte, Behrens, Johann (1996), Public Health Forschung mit Gesundheits- und Sozialdaten. Memorandum zur Nutzung von Gesundheits- und Sozialdaten. St. Augustin: Asgard
- Flick, Uwe (1994) Der qualitative Forschungsprozess als Abfolge von Entscheidungen, in: Flick, Uwe et al (1994). S. 148–186
- Flick, Uwe et. al. (Hrsg.) (1994), Handbuch Qualitative Sozialforschung, München und Weinheim: Psychologie Verlags Union
- Gadamer, Hans-Georg (19723) Wahrheit und Methode – Grundzüge einer philosophischen Hermeneutik. Hübigen: Mohr 3. Auflage
- Hopf, Christel (1978) Die Pseudo-Exploration – Überlegungen zur Technik qualitativer Interviews in der Sozialforschung. Zeitschrift für Soziologie 7, S. 97 – 115
- Hopf, Christel, Weingarten, E. (Hrsg.) (1979) Qualitative Sozialforschung, Klett-Cotta, Stuttgart
- Hopf, Christel (1994), Qualitative Interviews in der Sozialforschung - - ein Überblick, in Flick, Uwe et. al (Hrsg.) (1994), S. 177–182
- v. Kardorff, Ernst (2000) Zur Verwendung qualitativer Forschung in: Flick, U., von Kardorff, E., Steinke, I. (Hrsg.) Qualitative Forschung – Ein Handbuch, Rowohlt Reinbek, S. 615 – 623

- Kelle, Udo, Susann Kluge (2001) Validitätskonzepte und Validierungsstrategien bei der Integration qualitativer und quantitativer Forschungsmethoden, in: Susann Kluge, Udo Kelle (Hrsg.) Methodeninnovation in der Lebenslaufforschung. Weinheim und München: Juventa
- Lakatos, Imre (1982) Die Methodologie der wissenschaftlichen Forschungsprogramme. Philosophische Schriften, Band 1. Wiesbaden: Vieweg
- Meinefeld, Werner (1997) Ex ante Hypothesen in der Qualitativen Sozialforschung: zwischen »fehl am Platz« und »unverzichtbar«. Zeitschrift für Soziologie 26, S. 22 – 30
- Merton, R. K., Kendall, P. L. (1979) Das fokussierte Interview, in: Hopf, C., Weingarten, E. (Hrsg.) Qualitative Sozialforschung, Klett-Cotta Stuttgart, S. 171 – 204
- Oevermann, Ulrich, Schuster, Leo, Simm, Andreas (Hg.)(1985) Zum Problem der Perseveranz in Delikttyp und modus operandi. BKA-Forschungsreihe Band 17. Wiesbaden
- Raspe, Heiner (2000), Grundlagen und Theorie der evidenzbasierten Medizin (EbM), in: Regina Kunz, Günter Ollenschläger, Heiner Raspe, Günther Jonitz, Friedrich-Wilhelm Kolkman (Hrsg.), Lehrbuch Evidenzbasierte Medizin in Klinik und Praxis, Köln: Deutscher Ärzte Verlag
- Strauss, Anselm L./Corbin, Juliet (1990) Basics of Qualitative Research. Grounded Theory Procedures and Techniques. Newbury Park: Sage
- Wagner, Hans-Josef (1999): Rekonstruktive Methodologie: Georges Herbert Mead und die qualitative Sozialforschung, Opladen