

Was ist ein faktorieller Survey?

Christiane Atzmüller/Peter M. Steiner

Einleitung

Forschungsvorhaben im Bereich Palliative Care oder naheliegenden Bereichen aus dem Pflege- und Gesundheitsbereich, die zum Ziel haben, empirisch fundierte Kenntnisse zur Qualität der Pflege und Betreuung von Patienten bzw. zur medizinischen Indikation zu generieren, stehen aufgrund ethischer, rechtlicher und kostenbedingter Restriktionen oft vor großen Herausforderungen. Stellen wir uns etwa vor, wir möchten untersuchen, wie Hospizmitarbeiter im Rahmen der häuslichen, mobilen Palliativbetreuung mit sozialen und psychischen Belastungssituationen umgehen. Hier eine Beobachtung als Erhebungsmethode einzusetzen, um die konkreten Interaktionen zwischen Hospizmitarbeitern, Patienten und anderen Familienangehörigen zu dokumentieren, wäre aufgrund situativer Beschränkungen in der Regel nicht durchführbar und wahrscheinlich auch ethisch nicht zu verantworten, wenn anwesende Forscher solche Situationen zusätzlich belasteten. Würde man Hospizmitarbeiter mittels klassischer Frage-techniken dazu befragen, fehlt aufgrund der oft relativ abstrakten Fragenformulierungen der notwendige Kontextbezug, da anzunehmen ist, dass sich der Umgang mit Belastungssituationen in verschiedenen häuslichen oder patientenbezogenen Settings ganz unterschiedlich gestalten kann. Hinzu kommt, dass besonders auf der Ebene der persönlichen Betroffenheit z.B. auch mit Antwortverweigerungen oder sozial erwünschtem Antwortverhalten gerechnet werden muss, wenn wir uns z.B. vorstellen, dass psychische Belastungssituationen auch zu emotional überzogenen Reaktionen der Hospizmitarbeiter selbst führen können. Solche sensiblen, emotional und moralisch aufgeladenen und mitunter auch kontrovers diskutierten Inhalte sind dann schwer zugänglich.

Gerade im Pflege- und Gesundheitsbereich sind empirische Befunde, die Aufschluss darüber geben, welche Faktoren die patientenbezogene Urteilsbildung und Handlungsentscheidung durch medizinisches Fach- und Pflegepersonal beeinflussen, von hoher Bedeutung, da diese enorme Folgen für die einzelnen Patienten bzw. die Qualität des gesamten Gesundheitssystems haben (Evans et al., 2015).

Um solchen Anforderungen gerecht zu werden, bietet sich der *Einsatz von Vignetten* an. Das sind kurze, sorgfältig konstruierte Personenbeschreibungen,

hypothetische Szenarien oder Geschichten (Finch, 1987; Schoenberg & Ravadal, 2000), die zum Beispiel typische, auch in der Realität vorkommende Schlüssel-szenarien repräsentieren (Alexander & Becker, 1978; Atzmüller & Steiner, 2010; Gould, 1996) und sich aus einer Kombination von verschiedenen relevanten Vignettenfaktoren, wie z.B. Geschlecht, Alter oder Gesundheitszustand einer Person zusammensetzen und in der Regel in schriftlicher oder bildgestützter Form vorgelegt werden. Durch die Variation einzelner Vignettenelemente können unterschiedliche Vignettenszenarien mit unterschiedlichen Kontexten konstruiert werden, die dann den Respondenten zur Bewertung vorgelegt werden können. Vignetten wurden bereits in den 60er und 70er Jahren in der Sozialpsychologie und experimenteller Psychologie verwendet (ein historischer Überblick findet sich bei Evans et al. 2015).

Abbildung 1 zeigt eine kurze, fiktive Vignette, die eine Person durch die spezifische Kombination der Faktoren Geschlecht, Alter, familiärer Hintergrund, Krankheitsbild, Realitätsverweigerung bzw. -akzeptanz und stationäre bzw. ambulante Behandlung beschreibt.

Frau Holzer: 38 Jahre, ist verheiratet und hat zwei kleine Kinder.

Sie leidet an Krebs im Endstadium, trotzdem gibt sie die Hoffnung auf Heilungschancen nicht auf. Sie wird ambulant zu Hause betreut.

Abbildung 1: Beispielvignette 1

In einer möglichen Studie könnten z.B. Hospizmitarbeiter mit verschiedenen Vignetten konfrontiert und gebeten werden, für jede Vignettenperson zu entscheiden, welche palliative Betreuungsmöglichkeit als angemessen erachtet wird. Vignetten werden sowohl in Kombination mit qualitativen als auch mit quantitativen Forschungsmethoden eingesetzt und haben sich vor allem im Bereich der Pflege und Medizin bewährt, indem zum Beispiel Einstellungen, Meinungen und Wahrnehmungen von medizinischen Fachkräften oder Pflegepersonal untersucht wurde (z.B. Hughes & Huby, 2001), oder die Qualität der Pflege in ambulanten Settings (z.B. Peabody, Luck, Glassman, Dresslhaus, & Lee, 2000), forensische psychiatrische Urteile betreffend krimineller Unzurechnungsfähigkeit (z.B., Grøndal, Grønnerød, & Sexton, 2009) und Entscheidungsfindung (z.B., Wainwright, Gallagher, Tompsett, & Atkins 2010, Taylor, 2006; Wallander, 2012, Stokes & Schmidt, 2011).

Qualitative Vignetten dienen dabei vor allem als Gesprächsstimulus (siehe dazu zum Beispiel eine Untersuchung von Jackson et al. 2015 zu öffentlichen Gesundheitsfragen), um zum Beispiel eine Diskussion oder Bewertung des prä-

sentierten Vignettenszenarios anzuregen (Finch, 1987; Schoenberg & Ravdal, 2000;) oder auch Handlungsoptionen anzugeben (z.B. Barter & Renold 1999).

Quantitative Vignetten (z.B. Auspurg & Hinz, 2015; Steiner & Atzmüller, 2006; Beck & Opp, 2001, Jasso, 2006;) erlauben es hingegen, die Bedeutung einzelner Vignettenelemente systematisiert und unter kontrollierten Bedingungen zu untersuchen. Dabei werden Respondenten mit mehreren, auf der Grundlage eines experimentellen Anordnungsplans gezielt ausgewählten Vignetten (= Vignettenexperiment) konfrontiert und gebeten, an Hand einer Skala die einzelnen Vignettenszenarien hinsichtlich einer bestimmten Fragestellung zu beurteilen. Für das eingangs eingeführte Vignettenbeispiel könnte so zum Beispiel analysiert werden, welche Rollenakzeptanz der familiäre Kontext oder die Realitätsverweigerung oder -akzeptanz der Patienten für die Bewertung der Angemessenheit von palliativen Betreuungsformen spielt bzw. welche Bedeutung die übrigen Vignettenfaktoren für die Vignettenbeurteilung haben. Wird ein solches Vignettenexperiment in eine klassische Umfrage eingebettet, indem mittels eines traditionellen Fragebogens zusätzliche respondentenspezifische Inhalte, wie etwa relevante soziodemographischer Merkmale, aber auch persönliche Einstellungen und Erfahrungen erhoben werden, spricht man von einem *faktoriellen Survey*. Analyseziel ist es zu bestimmen, welche Vignettenfaktoren die Vignettenbeurteilungen maßgeblich beeinflusst haben, sowie mögliche Unterschiede zwischen den Respondenten in den Vignettenbewertungen zu erklären und damit (implizite) Urteils- und Bewertungsmuster in Form von Erklärungsmodellen zu rekonstruieren. Für das fiktive Beispiel könnte also festgestellt werden, inwieweit z.B. das Geschlecht oder der familiäre Kontext der befragten Hospizmitarbeiter die Vignettenbewertungen zusätzlich beeinflusst haben bzw. inwieweit sich unter den Hospizmitarbeitern unterschiedliche Urteils- und Bewertungsmuster finden lassen. Die Kombination von experimentellem Vignettendesign und Fragebogen nützt so sowohl der hohen internen Validität experimenteller Designs als auch der hohen externen Validität von Umfragetechniken (Atzmüller & Steiner 2010).

Der faktorielle Survey wurde erstmals 1974 von Rossi und seinen Kollegen eingesetzt (Rossi et al., 1974a, 1974b). Die Idee dazu entstand nach Rossi aufgrund eines Kommentars von Lazarsfeld in Rossis Dissertation 1951 (Rossi 1979). Rossi bezeichnet die Technik als Faktoriellen Survey: "... [it] combines ideas from balanced multivariate experimental designs with sample survey procedures" (Rossi und Anderson, 1982: 15). Wichtige Beiträge zum faktoriellen Survey wurden auch von Jasso (2006) geleistet. Unabhängig von dieser Tradition wurden in etwa zur gleichen Zeit auch bedeutende methodologische Innovationen zum experimentellen Vignettendesign von Cook (1977, 1979) und 1978 von Alexander und Becker (1978) hervorgebracht. Obwohl dieser Ansatz eine Reihe

von Vorteilen sowohl hinsichtlich technisch-statistischer als auch inhaltlicher Aspekte hat, wurde dieser Zugang erst relativ spät explizit mit dem faktoriellen Survey in Verbindung gebracht (Atzmüller & Steiner 2010; Steiner & Atzmüller, 2006; Dülmer 2007).

Die hauptsächlichen *Vorteile* von faktoriellen Surveys lassen sich wie folgt zusammenfassen: (i) Die Präsentation und Bewertung von Vignetten, die die spezifische Zusammensetzung mehrerer Faktoren gleichzeitig repräsentieren, ermöglicht die Beurteilung von konkreten und realistischen Szenarien anstatt von einzelnen, abstrakten klassischen Fragebogenfragen (Alexander & Becker, 1978). (ii) Die Methode hat hohes Potenzial zur Erforschung konditional und situativ bedingter Fragestellungen, aber auch zur Erforschung sensibler Themen, da die persönliche Ebene der Betroffenheit auf eine abstraktere Ebene verlagert wird, sodass es den Befragten leichter fällt, sensible Situationen über dritte Personen in der Vignette zu bewerten (Hughes, 1998). (iii) Die Möglichkeit, die Vignetteninhalte z.B. mittels Text, Bild oder Video zu präsentieren, eröffnet hohes Gestaltungspotenzial und macht Vignetten zu einem attraktiven Erhebungswerkzeug. (iv) Vignetten ermöglichen hoch kontextualisierte und multivariate Messungen, sodass auch die Identifizierung von Wechselwirkungseffekten möglich wird und die Untersuchung kausaler Beziehungen im Vorhinein auf der Designebene kontrolliert werden kann. (v) Die simultane Präsentation mehrerer Faktoren erschwert die normative Orientierung und folglich sozial erwünschtes Antwortverhalten. (vi) Die im Zuge der Erhebung entwickelten und mitunter recht aufwändig produzierten Vignetten können aber auch z.B. für Schulungs- oder Trainingszwecke weiterverwendet werden. Auf diese Weise können wissenschaftliche Erkenntnisse dem Erhebungsmaterial gegenübergestellt und beispielhaft komplexe Sachverhalte und Zusammenhänge veranschaulicht werden. (vii) Und schließlich hat der Einsatz von realitätsnah gestalteten Vignetten auch hohes Potenzial als Alternative zu Verhaltensbeobachtungen insbesondere bei emotional und moralisch stark aufgeladenen Themenfeldern oder wenn reale Beobachtungen aufgrund ethischer oder rechtlicher Bedenken nicht durchführbar sind. Vignetten sind zudem eine kostengünstige und praktische Methode zur Untersuchung der medizinischen Praxis im Vergleich zum Einsatz von standardisierten Patienten (Bachmann et al. 2008; Veloski et al., 2005, Evans et al. 2015). Darüber hinaus können auch ethische Aspekte umgangen werden, die bei Untersuchungen mit tatsächlichen Patienten oder in Zusammenhang mit vertraulichen Gesundheitsdaten auftreten können (Shah et al., 2007; Taylor, 2006; Evans et al. 2015).

Diese Vorteile haben auch zu einem Anstieg quantitativer Vignettenstudien in vielfältigen Anwendungsgebieten geführt, wie etwa im Bereich der medizinischen Behandlung (z.B. Parsons et al, 2014; Brenner et al., 2014) und in der

Betreuung und Pflege (z.B. Samuelsson & Wallander, 2014; Gould, 1996), oder auch zur Untersuchung der Berufsethik in Psychologie und Medizin (Macpherson & Veatch, 2010; McCarron & Steward, 2011). Zahlreiche Studien sind auch in der Psychologie zu finden (z.B. Ludwick & Zeller, 2001; Barnhill, 2014). Vignetten haben also ein hohes Potenzial zu Untersuchung kausaler Bewertungs-, Urteils- und Entscheidungsmustern, die im Pflege- und Gesundheitsbereich üblicherweise an realen Patienten nicht umsetzbar bzw. aus ethischen Gründen nicht durchführbar ist (Evans et al., 2015).

Den Vorteilen stehen aber auch gewisse *Limitationen* gegenüber: (i) Trotz des hohen Potenzials als Alternative zu Beobachtungen sind Vignetten hypothetische Konstrukte und können daher nur bedingt Aufschluss über reale Verhaltensabsichten vermitteln. Es wird empfohlen, Vignetten vorrangig zur Untersuchung von Einstellungen, Meinungen, Bewertungen oder zum Beispiel zur Untersuchung von normativen oder moralischen Urteilmustern oder Werthaltungen einzusetzen. (ii) Damit gehen auch mögliche Einschränkungen hinsichtlich der Validität und Reliabilität von Vignettenbefunden einher (Eifler, 2010). (iii) Und schließlich ist auch die Zahl an unterschiedlichen Vignettenfaktoren bzw. die Anzahl an Vignettenvariationen insofern limitiert, als mit sehr großen Vignettenpopulationen statistische Design- und Analysetools nur mehr bedingt oder vielleicht sogar überhaupt nicht die unvermischte Schätzung von interessierenden Effekten gewährleisten können. Strategien im Umgang mit diesen Herausforderungen werden im abschließenden Kapitel dieses Beitrags diskutiert. Im Folgenden soll nun der faktorielle Survey als Methode und dessen grundsätzliche Arbeitsweise vorgestellt und wichtige Hinweise vermittelt werden, die beim Einsatz und in der Planung berücksichtigt werden müssen.

Planung und Umsetzung eines faktoriellen Surveys

Das Untersuchungsdesign eines faktoriellen Surveys beinhaltet die Umsetzung eines Vignettenexperiments, das in eine klassische Befragung eingebettet wird. Wir konzentrieren uns in diesem Beitrag auf das Vignettenexperiment, indem wir auf grundlegende Überlegungen zum Vignettedesign, Stichproben- und Erhebungsdesign, sowie Analysedesign eingehen. Für die Ausarbeitung des Vignettedesigns sind folgende Schritte notwendig: 1) Konstruktion der Vignettenpopulation, 2) Reduktion der Vignettenpopulation bzw. Vignettensetbildung, 3) Festlegung der Beurteilungsaufgabe, der Antwortskalen und Wahl der Präsentationsform und 4) Pretest des Vignettedesigns.

Zur *Konstruktion der Vignettenpopulation* müssen zunächst die relevanten bzw. theoretisch begründeten oder aus entsprechenden explorativen Vorstudien

(Auspurg & Hinz, 2015; Atzmüller & Steiner, 2010) identifizierten Faktoren und zugehörigen Faktorstufen, die für die Untersuchung der Forschungsfragen herangezogen werden, festgelegt werden. Bezogen auf unser fiktives Vignettenbeispiel (Abbildung 1), waren das die 6 Faktoren (A) Geschlecht, (B) Alter, (C) familiärer Hintergrund, (D) Krankheitsbild, (E) Realitätsakzeptanz/-verweigerung und (F) ambulante/stationäre Betreuung.

Im nächsten Schritt gilt es nun die zugehörige Anzahl an Faktorstufen für jeden Faktor zu bestimmen. Wie Tabelle 1 zeigt, wählen wir für unser fiktives Beispiel zwei Faktorstufen pro Vignettenfaktor (in Klammer angeführt).

Tabelle 1: Ausgewählte Faktoren und Faktorstufen

Faktor	Faktorstufen
Geschlecht (2)	weiblich, männlich
Alter (2)	38, 75
Familie (2)	verheiratet mit zwei Kindern, alleinlebend
Krankheitsbild (2)	Krebs im Endstadium, schwer Herzkrank
Akzeptanz (2)	Realitätsakzeptanz, Realitätsverweigerung
Betreuung (2)	ambulant, stationär

Die gesamte Vignettenpopulation wird konstruiert, in dem sämtliche Variationsmöglichkeiten aller Vignettenfaktoren ausgeschöpft werden. Folglich ergibt sich die Gesamtzahl aller möglichen Vignetten aus $2^6=64$. Würden wir uns dafür entscheiden, alle 64 Vignetten jedem einzelnen Respondenten vorzulegen, könnten alle Haupt- und Interaktionseffekte innerhalb des Vignettensets geschätzt werden. Für unser Beispiel erscheint es aus inhaltlichen Überlegungen heraus durchaus plausibel, dass neben den Haupteffekten auch Interaktionseffekte zwischen den ausgewählten Vignettenfaktoren zu erwarten sind, wenn es darum geht für eine fiktive Vignettenperson zu bewerten, inwieweit eine bestimmte palliative Betreuungsmöglichkeit als angemessen erscheint oder nicht. Möglich wären z.B. Interaktionseffekte zwischen dem Alter und der Realitätsverweigerung bzw. -akzeptanz, aber auch zwischen dem Geschlecht und dem familiären Hintergrund.

Da 64 Vignetten zu viele sind, um von jedem Respondenten beurteilt zu werden, muss die Anzahl an Vignetten, die einer Person zur Bewertung vorgelegt wird, reduziert werden. Eine *Reduktion der Vignettenpopulation* ist allerdings in jedem Fall mit einem Informationsverlust und daraus resultierender Einschränkung in der Interpretation der Effekte verbunden. Je nach Größe der

gesamten Vignettenpopulation, der Stichprobengröße und der Anzahl an Vignetten, die einem Befragten zur Bewertung vorgelegt werden, sind grundsätzlich zwei Vorgehensweisen möglich, die mit unterschiedlichen Vor- und Nachteilen verbunden sind (Steiner & Atzmüller 2006): Erstens, es kann aus der gesamten Vignettenpopulation eine kleinere Vignetten-Subpopulation entnommen werden, zweitens, es wird die gesamte Vignettenpopulation in respondentenspezifische Sets aufgeteilt. Entscheidet man sich für die Auswahl einer kleineren Subpopulation, ist es möglich, diese entweder gezielt mit Hilfe eines experimentellen, sogenannten Fraktionalisierten faktoriellen Designs (Alexander & Becker, 1978; Cochran & Cox, 1950; Kirk, 1995), oder mittels einer Zufallsauswahl (Jasso, 2006; Rossi & Anderson, 1982) zu wählen. Je nach gewählter Strategie kommt es zu unterschiedlichen Restriktionen hinsichtlich der Schätzung und Interpretation von Effekten, da aufgrund von unvermeidbaren Vermischungsstrukturen (Konfundierung, Aliasing) nicht mehr alle Effekte innerhalb eines Vignettensets geschätzt werden können (eine ausführliche Darstellung findet sich bei Steiner & Atzmüller, 2006).

Der Vorteil experimenteller Designs ist nun, dass solche Konfundierung der Effekte gezielt geplant werden kann, sodass bestimmt werden kann, welche Haupt- und Interaktionseffekte unvermischt schätzbar bleiben sollen. Für kleinere Designs mit wenigen und vorzugsweise der gleichen Anzahl an Faktorstufen sind in der Literatur entsprechende Designpläne zu finden (Cochran & Cox, 1950; Kirk, 1995), bei komplexeren Designs können D-Effiziente Quotendesigns eine gute Alternative darstellen (Dülmer, 2016; Dülmer, 2007; Su & Steiner, 2016). Da es bei solchen Designs in der Regel unvermeidbar ist, Interaktionseffekte mit Haupteffekten zu vermischen, sind Fraktionalisierungen nur dann empfehlenswert, wenn jene Interaktionseffekte, die mit den Haupteffekten vermischt sind, als unbedeutend oder gleich null angenommen werden können, was bei unserem fiktiven Vignettenbeispiel kaum anzunehmen ist. Wird hingegen die Subpopulation zufällig ausgewählt, da eine solche auch bei aufwändigeren Designs mit mehreren Faktoren und ungleicher Anzahl an Faktorstufen relativ einfach umsetzbar ist, ist die Konfundierung der Effekte nicht planbar. Eine sinnvolle Interpretation von Effekten ist nur dann möglich, wenn die Konfundierungsstruktur noch vor der Datenerhebung entsprechend überprüft wird, wobei in der Regel lediglich Haupteffekte schätzbar bleiben, oft wieder nur unter der Voraussetzung, dass die mit den Haupteffekten konfundierten Effekte als null oder vernachlässigbar klein angenommen werden können. Ein möglicher Weg ist es, die Zufallsauswahl so lange zu wiederholen, bis eine zufriedenstellende Vermischungsstruktur gegeben ist.¹ Darüber hinaus ist eine Abweichung von der

1 Es ist auch möglich, für jeden Respondenten eine individuelle Zufallsstichprobe (Rossi 1979) zu ziehen (Ziehen mit Zurücklegen), wodurch die gesamte Vignettenpopulation besser ausge-

Orthogonalität² anzunehmen, da nicht alle Faktorstufen-Kombinationen gleich häufig repräsentiert sind. Das führt zu weniger effizienten Schätzungen und erfordert eine größere Stichprobe.

Wird es vorgezogen, die gesamte Vignettenpopulation in unterschiedliche Vignetten-Sets aufzuteilen, sodass die gesamte Vignettenpopulation ausgeschöpft wird, kann auch hier wieder eine experimentelle oder eine zufallsgestützte Vorgehensweise gewählt werden. Während bei einer Zufallsauswahl die Nachteile einer unkontrollierbaren Konfundierungsstruktur bestehen bleiben, ist *eine experimentelle Vorgehensweise* weitaus zielführender, wobei aus mehreren Gründen ein *konfundiert faktorielles Design mit randomisierten Blöcken* (Randomized Block Confounded Factorial, RBCF-Design, Kirk 1968, Cochran und Cox 1950) empfehlenswert ist: (i) Die Konfundierung, d.h. Vermischung von Effekten kann bewusst und im Voraus in Abhängigkeit von inhaltlichen Anforderungen und interessierenden Interaktionseffekten geplant werden. (ii) Nur höhere Interaktionseffekte sind mit dem Seteffekt (anstatt mit Haupteffekten) konfundiert, sodass Haupteffekte und relevante Interaktionseffekte unvermischt schätzbar bleiben. (iii) Aufgrund des balancierten und strikt orthogonalen Designs kann die statistische Effizienz der resultierenden Schätzungen maximiert werden.

Die nachfolgende Tabelle gibt noch einmal einen zusammenfassenden Überblick über die verschiedenen Möglichkeiten zur Reduktion der Vignettenpopulation:

schöpft wird. Das Problem unkontrollierter Vermischungsstrukturen bleibt aber weiterhin bestehen.

2 Die Orthogonalität eines Designs ermöglicht es, den Einfluss einzelner Vignettenfaktoren unabhängig voneinander zu schätzen.

Tabelle 2: Vorgehensweisen zur Reduktion der Vignettenpopulation

Entnahme einer kleinen Subpopulation	experimentelle Auswahl: kontrollierte Vermischung von Effekten, eine Vermischung von Interaktionseffekten mit Haupteffekten ist aber in der Regel unvermeidbar
	zufällige Auswahl: einfache Handhabung, aber unkontrollierte Konfundierung von Effekten
Aufteilung der gesamten Vignettenpopulation in respondentenspezifische Vignettensets	experimentelle Auswahl: kontrollierte Vermischung von Effekten, eine unvermischte Schätzung von Haupt- und relevanten Interaktionseffekten ist möglich
	zufällige Auswahl: einfache Handhabung, aber unkontrollierte Konfundierung von Effekten

Die Vignettenpopulation in unserem Beispiel mit 64 Vignetten, das sich aus 6 Faktoren zu je 2 Faktorstufen zusammensetzt, ist eine überschaubare Größe um ein experimentelles RBCF-Design zu wählen, das die unvermischte Schätzung von allen Haupt-, 2-fach und den meisten 3-fach Interaktionseffekten zulässt (Steiner & Atzmüller 2006, Atzmüller & Steiner 2010). Anhand unseres fiktiven Beispiels wollen wir diese Vorgehensweise daher demonstrieren. Bei der Umsetzung eines RBCF-Designs wird in Abhängigkeit von der Anzahl an Faktoren, Faktorstufen und der maximalen Anzahl an Vignetten pro RespondentIn, die gesamte Vignettenpopulation von m Vignetten systematisch in s gleich große Sets mit m/s Vignetten pro Set aufgeteilt, sodass Gruppen von RespondentInnen unterschiedliche Sets (Atzmüller & Steiner 2009) zur Beurteilung vorgelegt bekommen. Diese Vorgehensweise garantiert, dass die gesamte Vignettenpopulation ausgeschöpft wird und durch die notwendige Vignettenselektion möglichst wenig Information verloren geht. Darüber hinaus kann dieser Informationsverlust bzw. die dadurch entstehende Vermischungsstruktur gezielt geplant werden (Steiner & Atzmüller 2006, Atzmüller & Steiner 2010).

Als Erstes muss entschieden werden, in wie viele gleich große Sets die 64 Vignetten aufgeteilt werden sollen. Für eine optimale Aufteilung sollte die gewählte Setgröße ein ganzes Vielfaches der Vignettenpopulation darstellen, wodurch prinzipiell folgende Setgrößen möglich sind: 4 Sets mit jeweils 16 Vignetten ($4 \times 16 = 64$), 8 Sets mit jeweils 8 Vignetten ($8 \times 8 = 64$), oder 16 Sets mit jeweils 4 Vignetten ($16 \times 4 = 64$). Darüber hinaus bestimmt die Setgröße aber auch die Konfundierungsstruktur der jeweiligen Effekte: Sollen sowohl Haupt- als auch die wichtigsten Wechselwirkungseffekte unvermischt schätzbar bleiben, sollte

die Setgröße so gewählt werden, dass sie ein ganzes Vielfaches der notwendigen Faktorstufenkombinationen darstellt. Wir entschieden uns für 8 Sets zu jeweils 8 Vignetten. Damit bleiben sowohl alle Haupteffekte (8 ist ein ganzes Vielfaches von 2), als auch alle 2-fach (8 ist ein ganzes Vielfaches von 2×2) und die meisten 3-fach (8 ist ein ganzes Vielfaches von $2 \times 2 \times 2$) Wechselwirkungseffekte unvermischt schätzbar. Gelingt dies durch zum Beispiel eine höhere Anzahl an Faktoren und Faktorstufen nicht, kommt es zu einer (partiellen) Konfundierung (Steiner & Atzmüller 2010). Eine Setgröße von 8 Vignetten ist zudem im Hinblick auf die geplante Bewertungsaufgabe noch überschaubar, invalide Messungen durch Ermüdungseffekte sind hier nicht zu erwarten (Auspurg & Hinz, 2015; Sauer, Auspurg, Hinz & Liebig, 2011).

Nachdem die Größe der Vignettensets bestimmt ist, müssen die Vignetten systematisch gemäß dem RBCF-Design den einzelnen Sets zugeordnet werden. Aufgabe ist es, die Sets so zusammenzustellen, dass möglichst effiziente Schätzungen erreicht werden und alle Haupt- und alle 2-fach Interaktionen unvermischt schätzbar bleiben, indem das Design für die Setbildung möglichst balanciert und orthogonal ausgerichtet ist (Kirk 1995). Ein Design ist dann vollständig balanciert, wenn die Vignetten so auf die Sets aufgeteilt werden, dass sie gleich häufig gemessen werden. Kommt sowohl insgesamt als auch in jedem Set die gleiche Anzahl an Faktorstufen vor, ist auf diese Weise auch die Orthogonalität vollständig gewährleistet. Für die konkrete Umsetzung bedeutet das, dass darauf geachtet werden muss, dass jene Faktorstufenkombinationen, die für die Schätzung jedes Haupteffektes und jeder 2-fach Interaktion usw. notwendig sind, gleich häufig in jedem Set vorkommen müssen. Wenn nicht mehr jede k -te Interaktion höherer Ordnung in jedem Set vorkommen kann, dann sollten für diese nun mit dem Seteffekt konfundierten Interaktionseffekte die Faktorstufenkombinationen so balanciert wie möglich auf die Vignettensets verteilt werden. Damit sind lediglich Interaktionseffekte höherer Ordnung, die üblicherweise von geringerem Interesse oder vernachlässigbar klein oder null sind, mit dem Seteffekt konfundiert. Und da der Seteffekt lediglich potenzielle Unterschiede in den Beurteilungen von Set zu Set reflektiert, ist dieser üblicherweise nicht von Interesse (siehe dazu ausführlich Steiner & Atzmüller 2006; Atzmüller & Steiner 2010; Su & Steiner, 2016). Ausgearbeitete Konfundierungspläne können in der Literatur gefunden werden (Kirk, 1995; Cochran & Cox, 1950), wobei darauf geachtet werden soll, dass die Anzahl an Faktoren möglichst klein bleibt und idealerweise jeder Faktor die gleiche Anzahl an Faktorstufen aufweist, um zu komplexe Designs zu vermeiden (Kirk 1995). Einfache Designs können aber auch selbst mittels systematischer Zuordnung erstellt werden, die so lange variiert wird, bis eine optimale Lösung bezüglich der Konfundierungsstruktur erreicht wird.

Abbildung 2 zeigt ein mögliches RBCF (=Randomized Block Confounded Factorial)-Design, das für unsere fiktive Studie eingesetzt werden könnte: Die insgesamt 64 Vignetten wurden zu 8 Vignettensets mit jeweils 8 Vignetten aufteilt. Die Bezeichnung „Set“ gibt die jeweilige Vignettensetzugehörigkeit an, die Bezeichnungen *A*, *B*, *C*, *D*, *E* und *F* repräsentieren die sechs Faktoren mit entsprechend jeweils zwei Faktorstufen (*a*1, *a*2, etc.). Jede Zeile innerhalb der Vignettensets stellt die konkrete Zusammensetzung einer Vignette dar, also z.B. die erste Zeile im Set 1: *a*0 (=weiblich), *b*0 (=38 Jahre), *c*0 (=verheiratet mit zwei Kindern), *d*0 (=Krebs im Endstadium) *e*0 (=Realitätsakzeptanz) und *f*0 (=ambulante Betreuung).

Set	A	B	C	D	E	F	Set	A	B	C	D	E	F	Set	A	B	C	D	E	F	Set	A	B	C	D	E	F
1	a0	b0	c0	d0	e0	f0	2	a1	b1	c0	d0	e0	f0	3	a1	b0	c1	d0	e0	f0	4	a0	b1	c1	d0	e0	f0
1	a1	b1	c1	d1	e0	f0	2	a0	b0	c1	d1	e0	f0	3	a0	b1	c0	d1	e0	f0	4	a1	b0	c0	d1	e0	f0
1	a1	b0	c1	d0	e1	f0	2	a0	b1	c1	d0	e1	f0	3	a0	b0	c0	d0	e1	f0	4	a1	b1	c0	d0	e1	f0
1	a0	b1	c0	d1	e1	f0	2	a1	b0	c0	d1	e1	f0	3	a1	b1	c1	d1	e1	f0	4	a0	b0	c1	d1	e1	f0
1	a0	b1	c1	d0	e0	f1	2	a1	b0	c1	d0	e0	f1	3	a1	b1	c0	d0	e0	f1	4	a0	b0	c0	d0	e0	f1
1	a1	b0	c0	d1	e0	f1	2	a0	b1	c0	d1	e0	f1	3	a0	b0	c1	d1	e0	f1	4	a1	b1	c1	d1	e0	f1
1	a1	b1	c0	d0	e1	f1	2	a0	b0	c0	d0	e1	f1	3	a0	b1	c1	d0	e1	f1	4	a1	b0	c1	d0	e1	f1
1	a0	b0	c1	d1	e1	f1	2	a1	b1	c1	d1	e1	f1	3	a1	b0	c0	d1	e1	f1	4	a0	b1	c0	d1	e1	f1
Set	A	B	C	D	E	F	Set	A	B	C	D	E	F	Set	A	B	C	D	E	F	Set	A	B	C	D	E	F
5	a0	b1	c0	d0	e0	f0	6	a1	b0	c0	d0	e0	f0	7	a1	b1	c1	d0	e0	f0	8	a0	b0	c1	d0	e0	f0
5	a1	b0	c1	d1	e0	f0	6	a0	b1	c1	d1	e0	f0	7	a0	b0	c0	d1	e0	f0	8	a1	b1	c0	d1	e0	f0
5	a1	b1	c1	d0	e1	f0	6	a0	b0	c1	d0	e1	f0	7	a0	b1	c0	d0	e1	f0	8	a1	b0	c0	d0	e1	f0
5	a0	b0	c0	d1	e1	f0	6	a1	b1	c0	d1	e1	f0	7	a1	b0	c1	d1	e1	f0	8	a0	b1	c1	d1	e1	f0
5	a0	b0	c1	d0	e0	f1	6	a1	b1	c1	d0	e0	f1	7	a1	b0	c0	d0	e0	f1	8	a0	b1	c0	d0	e0	f1
5	a1	b1	c0	d1	e0	f1	6	a0	b0	c0	d1	e0	f1	7	a0	b1	c1	d1	e0	f1	8	a1	b0	c1	d1	e0	f1
5	a1	b0	c0	d0	e1	f1	6	a0	b1	c0	d0	e1	f1	7	a0	b0	c1	d0	e1	f1	8	a1	b1	c1	d0	e1	f1
5	a0	b1	c1	d1	e1	f1	6	a1	b0	c1	d1	e1	f1	7	a1	b1	c0	d1	e1	f1	8	a0	b0	c0	d1	e1	f1

Abbildung 2: RBCF-Design: 8 Vignettensets zu je 8 Vignetten

Die Konfundierungsstruktur kann durch eine einfache Kreuztabellierung der Faktoren für jedes separate Vignettenset überprüft oder mit Hilfe des frei zugänglichen Statistikprogramms R mit der Funktion `alias()` dargestellt werden.

Es zeigt sich, dass in jedem Set die jeweiligen Faktorstufen für Faktor *A*, *B*, *C*, *D*, *E* und *F* gleich häufig vorkommen. Damit bleiben alle Haupteffekte unvermischt schätzbar. Auch bei den 2-fach Interaktionseffekten kommen die entsprechenden Faktorstufenkombinationen in jedem Vignettenset gleich häufig vor, wodurch die 2-fach Wechselwirkungseffekte unvermischt schätzbar bleiben. Für das fiktive Vignettenbeispiel mit 8 Sets zu je 8 Vignetten ist das ein optimales Design, da alle inhaltlich relevanten Haupt- und Wechselwirkungseffekte unvermischt schätzbar bleiben. Für komplexere Designs mit wesentlich mehr Vignettenfaktoren als in unserem Beispiel und mit einer unterschiedlichen Anzahl an Faktorstufen, wo die Erstellung eines konfundiert faktoriellen Designs eine zu hohe Herausforderung darstellt (Steiner & Atzmüller 2006), können konfundierte D-effiziente Designs (Dülmer, 2016; Su & Steiner, 2016) eingesetzt werden, die mit Hilfe von Software-Paketen wie SAS oder R erstellt werden können. Ein methodischer Vergleich der unterschiedlichen Reduktionsstrategien hat allerdings gezeigt, dass ein balanciertes, konfundiert faktorielles Designs insgesamt das effizienteste Design mit der höchsten internen Validität darstellt und damit einem idealem Design entspricht (Dülmer, 2016; Su & Steiner, 2016).

Nachdem eine passende Vignettensetbildung auf experimenteller Basis vorgenommen wurde, müssen die *Beurteilungsaufgabe*, die *Antwortskalen* und die *Wahl der Präsentationsform* festgelegt werden. Als Antwortskalen sind vor allem Rating-Skalen üblich, wodurch regressionsanalytische Analyseverfahren anwendbar sind. Alternativ dazu sind aber auch Kombinationen denkbar, zum Beispiel könnte dem Ratingverfahren ein Ranking (Steiner, Atzmüller & Su, 2016; Steiner, Atzmüller & Wroblewski, 2009) vorausgehen, um eine differenzierte Bewertung zu erleichtern. Die Anwendung von kategorialen Skalen erfordert andere Analyseverfahren, wie zum Beispiel Logit- oder Probit-Modelle (Auspurg & Hinz 2015).

Üblicherweise und aus Gründen der leichteren Administrierbarkeit werden die Vignetten häufig in Textform präsentiert. Es sind aber auch bild- oder audigestützte Präsentationsformen möglich. Sowohl die Antwortskala, als auch die Wahl einer passenden Präsentationform hat aber großen Einfluss auf die Stabilität der Messung und folglich auch auf die interne Validität des Vignettenexperiments. Wir werden darauf noch näher im nächsten Kapitel eingehen.

Schließlich ist es wichtig, das Vignettendesign noch einem ausführlichen *Pretest* zu unterziehen, bei dem neben der Relevanz und Sinnhaftigkeit der gewählten Vignettenfaktoren auch Aspekte wie Verständlichkeit oder Ermüdungserscheinung getestet werden müssen. Außerdem muss geprüft werden, wie viele Vignetten den RespondentInnen zumutbar sind bzw. wie hoch der Schwierigkeitsgrad der Bewertungsaufgabe ist, was auch von der befragten Zielgruppe abhängt, da z.B. ältere Menschen möglicherweise schneller überfordert sind als

jüngere. Erfahrungsgemäß zeigt sich, dass die Vignettenbewertung oft sehr genau genommen wird, sodass entsprechend Zeit zur Beantwortung der Vignetten mit eingeplant werden muss. Die Pretestdaten können aber auch für die Durchführung einer Poweranalyse herangezogen werden, um eine effiziente Stichprobengröße zu bestimmen (Auspurg & Hinz, 2015; Steiner, Atzmüller & Su, 2016).

Nun muss noch das den experimentellen Designbedingungen entsprechende *Stichprobendesign und die Zuordnung der Vignetten zu den Respondenten* geklärt werden. Die Stichprobengröße sollte so angelegt werden, dass jede einzelne Vignette bzw. jedes einzelne Vignettenset gleich häufig und jede Vignette zumindest fünfmal gemessen wird (Auspurg & Hinz 2015). Möglich ist hier eine Quotenstichprobe z.B. mit disproportionaler Schichtung der Zielpopulation durch relevante soziodemographische Variablen, sodass in jeder Schicht genau gleich viele Respondenten ausgewählt werden, um auch bei den Respondentmerkmalen eine hohe Effizienz und interne Validität zu erzielen (Steiner & Atzmüller, 2006; Steiner, Atzmüller & Su, 2016). Möglich ist aber auch eine geschichtete Zufallsauswahl. Die Zuteilung der in unserem Beispiel acht Vignettensets zu den Respondenten muss innerhalb einer Quotierungszelle bzw. Schicht zufällig erfolgen. Idealerweise wird auch die Vignettenreihenfolge innerhalb eines Sets z.B. durch Mischen, falls die Vignetten in Form von Kärtchen vorgelegt werden, randomisiert.

Nach erfolgter Datenerhebung sollte dann ein entsprechender Datensatz vorliegen, der bezogen auf unser Beispiel mit 8 Vignetten pro Person 8 Zeilen enthält, also pro Vignette eine Zeile, wo dann die Bewertung der jeweiligen Person eingegeben wird.³ Diese Datenmatrix (siehe dazu auch Auspurg & Hinz, 2015) stellt die Grundlage für das *Analysedesign* dar.

Das Hauptinteresse der Vignettenanalyse liegt in der Rekonstruktion von Bewertungsmodellen, die die Vignettenbewertungen erklären, um so jene Einflussfaktoren zu spezifizieren, die die Beurteilung der Vignetten maßgeblich beeinflusst haben. Die Bewertung einer Vignette hängt sowohl von den Vignettenfaktoren (=Level-1-Variablen), als auch von der Person, die die Vignette beurteilt (=Level-2-Variable), ab. Die Vignettenbewertung stellt die abhängige Variable dar, die auf einer entsprechenden Skala gemessen wird.

3 Hilfreich ist es, einen Datensatz für die Vignettenbewertungen zu erstellen, der auch die zu jeder Vignette zugehörigen Information der Faktorstufenkombinationen enthält und einen separaten Datensatz, wo die zusätzlichen respondentenspezifischen Informationen enthalten sind, die mittels des zusätzlichen Fragebogens erhoben wurden. Mit Hilfe einer gemeinsamen Vignetten-ID können die beiden Datenfiles zusammengefügt werden, sodass eine Datenmatrix mit 8 Zeilen entsteht, wobei die respondentenspezifischen Informationen sich bezogen auf die Anzahl der Vignetten x-fach wiederholen.

Da jeder Respondent mehrere Vignetten zur Beurteilung erhält, werden pro Respondenten (Block) mehrere Messungen durchgeführt, daher sind die Messungen über alle RespondentInnen betrachtet nicht unabhängig voneinander, was in der in der statistischen Analyse durch die Modellierung des Blockeffektes berücksichtigt werden muss.

Betrachtet man die Varianzzerlegung eines solchen Modells und dabei den Anteil der nicht erklärten Varianz in Form der Residuen, so gibt es demnach einerseits innerhalb einer Person einen Anteil an unerklärter Variation der Vignettenbeurteilungen, der nicht durch ein solches Regressionsmodell erklärt werden kann und andererseits einen Anteil an unerklärter Variation in den Vignettenbeurteilungen zwischen den einzelnen Personen.

Dieser Umstand kann bei Vorliegen balancierter Daten durch ein Varianz-analytisches Modell berücksichtigt werden. Kommt es bei der Erhebung zu Abweichungen, weil z.B. einzelne Vignettensets nicht von gleich vielen Respondenten beurteilt oder die Vignettensets ungleich häufig verteilt wurden, wird ein Hierarchisches Lineares Modell (Snijders & Bosker, 2012; Raudenbush & Bryk 2002) herangezogen. Eine ausführliche Diskussion der Besonderheiten, die bei der Datenanalyse zu beachten sind, findet sich bei Steiner & Atzmüller 2006 sowie bei Auspurg & Hinz 2014. Für die analytische Rekonstruktion der zugrunde liegenden Bewertungsmodelle empfiehlt sich ein sukzessiver Modellaufbau, indem zunächst mit dem „Empty Model“, das den konstanten Term (Intercept) aber keine Vignettenfaktoren als erklärende Faktoren enthält, geklärt wird, inwieweit es überhaupt Bewertungsunterschiede zwischen den Respondenten gibt. Im „Fitted Model“ werden dann zunächst ausschließlich die Vignettenfaktoren (= Level 1 Ebene) in das Modell aufgenommen und dann schrittweise um Respondentenvariablen (= Level 2 Ebene), die mit Hilfe des Fragebogens erhoben wurden (wie z.B. Geschlecht oder familiärer Hintergrund der Befragten, aber auch Einstellungskonstrukte wie z.B. persönliche Werthaltungen) erweitert. Neben den Haupteffekten können so auch Interaktionseffekte zwischen den Vignettenfaktoren und Respondentenvariablen (Cross-Level Interaktionseffekte) geschätzt werden.

Herausforderungen und Strategien im Umgang mit Limitationen

Trotz der vielen Vorteile, die der Einsatz von experimentellen Vignettendesigns anstelle einfacher Zufallsauswahltechniken in faktoriellen Surveys bietet, scheinen die Anforderungen bei der Erstellung des experimentelle Designs für Forscher oft eine zu große Hürde zu sein. Zum einen, da für das spezifische Forschungsvorhaben (zum Beispiel mit einer ungleichen Anzahl an Faktorstufen) in der

Literatur (z.B. Kirk, 1995) keine passenden Designpläne gefunden werden bzw. keine ausreichenden Kenntnisse vorhanden sind, um bestehende Designpläne entsprechend zu adaptieren. In diesem Fall ist es empfehlenswert für die Designerstellung mit Statistik-Experten bzw. Personen, die experimentelle Designs in ihrem Fachgebiet anwenden (z.B. Psychologie, aber auch Agrarwirtschaft), zu kooperieren. Hilfreich können aber auch Computer-Algorithmen für Softwareprogramme wie R, Stata oder SAS sein. Unter folgendem Link: https://study.sagepub.com/auspurg_hinz stehen hilfreiche SAS-Makros zur Verfügung, auf derselben Website finden sich zahlreiche andere Informationen z.B. auch zur Erstellung von D-Effizienten Designs (Auspurg & Hinz, 2015).

Zum anderen scheint die Vignettenpopulation aufgrund vieler Faktoren und Faktorstufen oft zu groß für die Ausarbeitung von Designs mit experimenteller Setbildung (RBCF-Designs) zu sein. Die Vorzüge eines solchen Designs sollten aber dazu anregen, alternative Designüberlegungen anzustellen, die es ermöglichen, die gesamte Vignettenpopulation zu reduzieren. So kann zum Beispiel ein zweistufiges Design (Atzmüller & Steiner, 2010) überlegt werden, indem in einer ersten explorativen Phase z.B. mit Hilfe einfacher Zufallsauswahltechniken oder fraktionalisierten Designs die relevantesten Vignettenfaktoren identifiziert werden, die dann in einem zweiten, kleineren experimentellen Design auf Ihre Kausalbeziehungen hin untersucht werden. Denkbar sind aber auch serielle Designs mit mehreren, kleinen Vignettenexperimenten, wo z.B. gezielt der Vignettenkontext (z.B. bezogen auf unser fiktives Vignettenbeispiel werden anstelle Erwachsener Kinder als Patienten präsentiert) zwischen den einzelnen Vignettendesigns variiert wird. Auf diese Weise kann eine breite Palette an relevanten Faktoren abgedeckt werden und so auch die Generalisierbarkeit der Ergebnisse erhöht werden (Atzmüller & Steiner, 2010).

Sollen solche zweistufigen oder seriellen Designs zum Beispiel aus Ressourcengründen vermieden werden, kann aber auch eine sinnvolle Aufsplitterung innerhalb eines Designs überlegt werden. So wurde zum Beispiel in einer Studie zur Bewertung geschlechtsspezifischer Einkommensunterschiede (Steiner, Atzmüller & Wroblewski, 2009) der 2-stufige-Faktor Geschlecht zum „between-subjects-factor“, indem die Vignettenpopulation von insgesamt 162 auf zwei Subpopulationen mit je 81 Vignetten (3x3x3x3-Design) reduziert und den Respondenten entweder nur Frauen- oder nur Männervignetten vorgelegt wurden. Neben dem Effekt der Reduktion der gesamten Vignettenpopulation mit gleicher Anzahl an Faktorstufen, wodurch die Designerstellung erleichtert wurde, umging man so auch Effekte sozial erwünschten Antwortverhaltens. In einer anderen Studie zur Gewaltwahrnehmung durch Jugendliche (Atzmüller & Kromer 2013) wurde eine Vignettenpopulation, die mit insgesamt 768 Vignetten zunächst zu groß für ein experimentelles Design erschien, in 3 Blöcke mit je 256

Vignetten (4x4x2x2x2x2-Design) aufgesplittet. In einer weiteren Untersuchung zu delinquentem Verhalten (Atzmüller & Kromer, 2014) sprachen zunächst inhaltliche Überlegungen für ein 8x9x2x2-Design, die heterogene Anzahl an Faktorstufen war in diesem Fall suboptimal für die Umsetzung eines passenden experimentellen Designs. Darüber hinaus hätte das Design auch unplausible Vignettenkonstellationen⁴, also Kombinationen, die in der Realität kaum oder nie vorkommen (im unserem fiktiven Beispiel etwa weibliche Patientinnen mit einer Krankheit, die nur bei Männern auftritt) beinhaltet. Die Vermeidung unplausibler Vignetten ist deshalb wichtig, da das experimentelle Vignettendesign auf der Balanciertheit der Daten aufbaut, d.h., dass jede Vignette gleich oft gemessen werden soll, und daher der Ausschluss unplausibler Vignetten aus Validitätsüberlegungen nicht möglich ist (Steiner & Atzmüller 2006). Es war aber möglich, das Design aus inhaltlichen Überlegungen heraus in homogene Subgruppen zusammenzufassen, sodass das Gesamt-design in 7 kleinere und gut handhabbare 2x4x2x2 Sub-Designs gesplittet werden konnte. Auf diese Weise konnten auch unplausible Vignettenkonstellationen vermieden werden. Es zeigt sich also, dass es durchaus bei „ungünstigen“ Voraussetzungen möglich ist, geeignete experimentelle Designlösungen zu finden.

Es lässt sich zusammenfassen, dass faktorielle Surveys mit sorgfältig geplanten experimentellen Designs und mit der gleichen Anzahl an Messungen für jede Vignette vor allem von einer erleichterten statistische Analyse und einer sinnvollen Interpretation der Effekte und damit einer maximalen internen Validität profitieren. Es sollte daher ausreichend Zeit für die Ausarbeitung des Vignettendesigns eingeplant werden in dem klar festgelegt werden muss, welche Haupt- und Wechselwirkungseffekte unvermischt schätzbar bleiben sollen. Erst dann kann darüber entschieden werden, welche Strategie zur Reduktion der Vignettenpopulation angemessen erscheint. Eine Entscheidung für eine einfache Zufallsauswahl rein aus Gründen der einfacheren Handhabung sollte in wissenschaftlichen Forschungskontexten aber jedenfalls vermieden werden. Neben der (i) internen Validität sind auch Aspekte der (ii) Konstruktvalidität, also inwieweit Vignetten tatsächlich das messen, was zu messen beabsichtigt wird, der (iii) externen Validität, inwieweit also Schlussfolgerungen auf andere Settings oder Respondenten-Stichproben gerechtfertigt, sind und der (iv) Reliabilität zu beachten, inwieweit also Fehler auf der Ebene der experimentellen Designerstellung, auf Messebene und auf der Stichproben-Ebene vermieden werden können, um präzise und effiziente Schätzungen in ausreichender Güte zur Überprüfung von Hypothesen zu erzielen. Es ist wichtig, solche Validitäts- und Reliabilitätsaspek-

4 Sind solche Kombinationen nicht vermeidbar, muss geklärt werden, wie in der Datenanalyse damit umgegangen wird bzw. inwieweit das die Bewertung beeinträchtigt (Atzmüller & Steiner, 2010; Auspurg & Hinz 2015).

te in der Planung von faktoriellen Surveys zu beachten und entsprechende Designüberlegungen anzustellen.

Kritisiert wird in diesem Zusammenhang häufig die Künstlichkeit von Vignetten, indem betont wird, dass hypothetische Szenarien Phänomene in der Realität nicht ausreichend repräsentieren und daher Bedenken hinsichtlich der Validität der Ergebnisse und Schlussfolgerungen bestehen (z.B. Gould 1996; Hughes, 1998; Eifler, 2010). Werden Vignetten aber möglichst realitätsnah und glaubhaft (Finch, 1987) gestaltet, erhöht sich nicht nur die Konstruktvalidität, sondern Vignetten stellen insbesondere bei ethisch und rechtlich bedenklichen Settings eine gute Alternative zur Verhaltensbeobachtung dar, wenn auch das hypothetische Verhalten in einem Experiment von realen Leben abweichen kann (Ludwick et al., 2004). Studien etwa mit Krankenschwestern (Lunza, 1990) oder Ärzten (Langley, Tritchler, Llewelly-Thomas, & Till, 1991) weisen aber darauf hin, dass Respondenten bei hypothetischen Szenarien ähnlich antworten wie in realen Szenarien bzw. mit Vignettendaten reales Verhalten gut prognostiziert werden kann. Wenn Vignetten also dazu verwendet werden, jene Faktoren zu analysieren, die Bewertungs-, Urteilmustern bzw. Entscheidungsprinzipien zugrunde liegen, dann geht es weniger um die Frage, ob die Vignetten-Welt der realen Welt entspricht, sondern eher darum, ob die kognitiven und verhaltensbezogenen Prozesse auf eine Art aktiviert werden, wie sie auch im realen Leben aktiviert werden (Evans et al., 2015).

In diesem Zusammenhang spielt der Modus der Stimulus-Präsentation eine bedeutende Rolle, da eine möglichst genaue, konkrete und detaillierte Vignettenpräsentation die Glaubwürdigkeit und Realitätsnähe unterstützt (Alexander & Becker, 1978; Atzmüller & Steiner, 2010). Je nach Forschungsfrage können Vignetten zum Beispiel als Textvignetten (z.B. verbalisierten Formen wie „Geschichtenanfänge“, aber auch stichwortartige Kurzdarstellungen), Bildvignetten (z.B. visualisierte Abbildungen „Bildgeschichte“; Cartoons (Verweis Vignetten Gewaltintervention), Zeichnungen, Videovignetten oder Audiovignetten eingesetzt werden. Je nach Präsentationsform muss aber bedacht werden, in welchem Setting (z.B. Papierfragebogen, computergestützt) die Vignetten vorgelegt werden, wobei hier die Möglichkeiten der im experimentellen Design vorgesehenen Randomisierungsschritte berücksichtigt werden müssen.

Die folgenden Abbildungen zeigen mögliche Präsentationsmodi: Vignette in Dialogform (Abbildung 3a) und in narrativem Stil (3b), die auch als Audiovignette (Nachrichtenbeitrag) zur Untersuchung von Gewalt unter Jugendlichen umgesetzt wurde. Des Weiteren eine Bildvignette (Abbildung 4a) im Cartoon/Piktogramm-Stil zur Untersuchung der Wahrnehmung polizeilicher Einvernahmen mit Kindern und eine Video/Fotovignette (Abbildung 4b) zur Untersuchung von Mobbing unter Jugendlichen.

PETRA und IVKO beschimpfen sich:

PETRA: He, du Fettsack!

IVKO laut: Was ist mit dir!

PETRA: laut: Schau dich an, Opfer, hast keinen Spiegel?

IVKO schreit: Was willst du, deine Mutter ist eine Hure!

Wie schlimm findest du so eine Situation?

nicht schlimm ☐ ☐ ☐ ☐ ☐ sehr schlimm

Abbildung 3a

Gestern wurde beobachtet, wie ein 16-jähriger Bursch in Begleitung seiner Freunde einen Gleichaltrigen mit einem Messer bedrohte und ihn aufforderte, das Smartphone herzugeben. Gegenüber der Polizei meinte der Beschuldigte wörtlich: „Das war doch alles nur Spaß! Ich hab gar nicht damit gerechnet, dass er mir das Smartphone wirklich gibt.“

Was denkst du, wenn du das hörst?

nicht schlimm ☐ ☐ ☐ ☐ ☐ sehr schlimm

Abbildung 3b

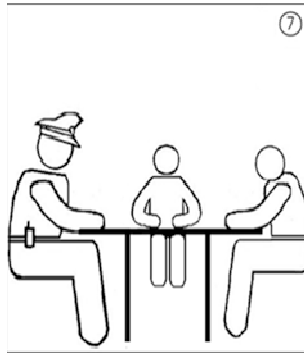
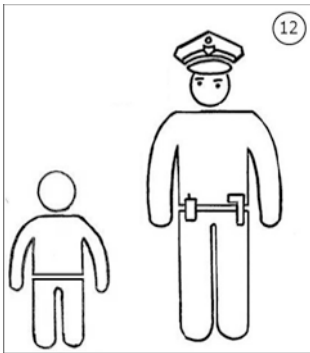


Abbildung 4a



Wie schlimm findest du so eine Situation?
 nicht schlimm o o o o o sehr schlimm

Abbildung 4b

Werden Fragebögen in Papierform gedruckt, müssen möglicherweise verschiedene Fragenbogenvarianten mit zufällig variiert Vignettenreihenfolge ausgegeben werden. Eine alternative Möglichkeit besteht darin, die Vignetten in Form von separaten Kärtchen vorzulegen. Insgesamt erhält die Befragung dadurch den Charakter eines Kartenspiels, wodurch auch die Randomisierung der Vignettenreihenfolge durch Mischen der Karten erleichtert wird, um systematische Reihenfolgeeffekte im Antwortverhalten zu vermeiden. Eine computergestützte Präsentation erleichtert hier möglicherweise die Vignettenpräsentation, darüber hinaus können auch alternative Vignetten-Präsentationsmodi wie zum Beispiel Video- oder Audiovignetten (Atzmüller & Kromer 2013, Atzmüller & Kromer 2014) eingesetzt werden. Erfahrungsgemäß erhöht eine solche innovative Darbietung die Aufmerksamkeit und die Bereitschaft zur Befragungsteilnahme überdurchschnittlich und kommt zudem möglichen Lese- und Konzentrationsschwierigkeiten entgegen. Da Textvignetten einfacher und kostensparender umgesetzt werden können, wird der überwiegende Teil an Vignettenstudien derzeit mittels Textvignetten umgesetzt. Allerdings können insbesondere emotionale Inhalte über die Textebene weniger treffsicher transportiert werden. Es ist aber auch denkbar, nur einige wenige, gezielt ausgewählte Videovignetten zu produzieren, um eine ausreichend detaillierte Kontextinformation zu übermitteln und anschließend auf der Textvignettenebene weitere Differenzierungen vorzunehmen (Atzmüller & Kromer 2014). Da die Art und Weise der Stimuluspräsentation auch die Vignettenbewertung beeinflusst, muss allerdings die Präsentationsform sorgfältig bedacht werden.

Conclusio

Ein faktorieller Survey ist aufgrund des besonderen Erhebungssettings mit systematisch variierten Vignetten für den Bereich Palliative Care eine äußerst vielversprechende Methode, einerseits als Alternative zu Verhaltensbeobachtungen, andererseits, um jene Faktoren festzumachen, die medizinischen Beurteilungen bzw. patientenbezogenen Behandlungsentscheidungen zugrunde liegen. Um das Potenzial dieser Methode, nämlich die hohe interne Validität experimenteller Designs mit der hohen externen Validität von quantitativen Befragungen tatsächlich nutzen zu können, müssen wesentliche Gütekriterien experimenteller Forschungsdesigns, die Orthogonalität, Balanciertheit und Randomisierung von nicht kontrollierbaren Einflussfaktoren, eingehalten werden. Darüber hinaus müssen geeignete Maßnahmen gesetzt werden, um die Validität und Reliabilität der Messdaten zu maximieren.

Um diesen Anforderungen gerecht zu werden, empfehlen wir wann immer möglich, explizit ein experimentelles Vignettendesign einzusetzen, idealerweise ein konfundiert faktorielles Design mit randomisierten Blöcken (RBCF-Design), das in diesem Beitrag ausführlich dargestellt wurde und im Vergleich zu anderen Vorgehensweisen ein ideales Design darstellt. Die damit verbundenen Herausforderungen und teilweise auch Restriktionen insbesondere bei der Planung des Vignettendesigns sind bei geeigneten Designüberlegungen gut handhabbar, verlangen allerdings eine tiefergehende Auseinandersetzung mit der Methode bzw. ist möglicherweise eine Zusammenarbeit mit Personen, die mit experimentellen Settings zumindest im statistischen Sinne versiert sind, anzuraten. Als Alternative sind auch D-Effiziente Designs denkbar, die allerdings trotz der zur Verfügung stehenden computergestützten Makros immer noch relativ anspruchsvoll in der Umsetzung sind.

Können Studierende oder Personen, die einen faktoriellen Survey zum ersten Mal umsetzen möchten, diese Herausforderungen nicht bewerkstelligen, sind auch einfach durchzuführende Zufallsauswahltechniken denkbar, erfordern aber unbedingt die Überprüfung der Vermischungsstrukturen der interessierenden Faktoren und gegebenenfalls mehrfache Wiederholungen der zufälligen Vignettenauswahl. Anderenfalls ist eine seriöse Interpretation der Effekte nicht möglich. Solche Studien sollten auch klar als explorative Untersuchungen ausgerichtet sein, da maximal Haupteffekte schätzbar sind. Es könnte aber durchaus ein zweistufiges Design angedacht werden, indem im Rahmen der Zufallsauswahl als bedeutsam identifizierte Haupteffekte in einem sehr kleinen experimentellen Design hinsichtlich möglicher Interaktionseffekte untersucht werden, möglicherweise mit einem so kleinen oder mehreren kleinen Designs, bei denen auf eine notwendige Reduktion der Vignettenpopulation verzichtet werden kann

(= Vollständig Faktorielle Designs, z.B. $3 \times 3 = 6$ Vignetten; $2 \times 2 \times 2 = 8$ Vignetten; $2 \times 2 \times 3 = 12$ Vignetten usw.; 6, 8, 12 Vignetten usw. sind in der Regel pro Person zumutbar) und damit die Umsetzung leicht durchführbar ist.

Wir hoffen, dass es uns mit diesem Beitrag gelungen ist, stärker geeignete Designelemente bei der Planung eines faktoriellen Surveys zu berücksichtigen, die den Validitäts- und Reliabilitätskriterien experimenteller Designs gerecht werden.

Literatur

- Alexander, C. S. & Becker, H. J. (1978). The Use of Vignettes in Survey Research. *Public Opinion Quarterly* 42(1), 93-104.
- Atzmüller, Ch. & Kromer, I. (2013). *Peer Violence: Gewalt unter Jugendlichen aus der Perspektive von Mädchen und Burschen*. Forschungsbericht für das Bundesministerium für Wissenschaft, Forschung und Wirtschaft (BMFWF), Wien, Österreich.
- Atzmüller, Ch. & Kromer, I. (2014). *Peer Delinquency: Wahrnehmung und Bewertung typischer Jugenddelikte aus der Sicht Jugendlicher als Grundlage für Präventionsmaßnahmen*. Forschungsbericht für das Bundesministerium für Verkehr, Innovation und Technologie (BMVIT), Wien, Österreich.
- Atzmüller, Ch. & Steiner, P. M. (2010). Experimental Vignette Studies in Survey Research. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences* 6(3), 128-138.
- Auspurg, K. & Hinz, T. (2015). *Factorial Survey Experiments*. Thousand Oaks: Sage.
- Bachmann, L., Mühleisen, A., Bock, A., Riet, G., Held, U., & Kessels, A. (2008). Vignette studies of medical choice and judgment to study caregivers' medical decision behavior: systematic review. *BMC Research Methodology*, 8, 8-50.
- Barnhill, J. W. (Ed.). (2014). *DSM-5 clinical cases*. Washington, D.C: American Psychiatric Publishing.
- Beck, M., & Opp, K-D. (2001). Der faktorielle Survey und die Messung von Normen. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 53(2), 283-306.
- Barter, Ch. & Renold, E. (1999). The use of vignettes in qualitative research. *Social Research Update* 25, University of Surrey.
- Brenner, M., Drennan, J., Treacy M. P., & Fealy, G. M. (2014). An exploration of the practice of restricting a child's movement in hospital: a factorial survey. *Journal of Clinical Nursing* 24, 1189-1198.
- Cochran, W. G. & Cox, G. M. (1950). *Experimental Designs*. New York: Wiley & Sons.
- Cook, F. L. (1979). *Who Should Be Helped? Public Support for Social Services*. London: Sage Publications.
- Cook, F. L. (1977). *Differences in Public Support for Seven Social Welfare Groups: Description and Explanation*. Doctoral Dissertation, University of Chicago.
- Dülmer, H. (2016). The Factorial Survey: Design Selection and its Impact on Reliability and Internal Validity. *Sociological Methods & Research*, 45(2), 304-347.

- Dülmer, H. (2007). Experimental plans in factorial surveys. Random or quota design? *Sociological Methods & Research* 35(3), 382-409.
- Eifler, S. (2010). Validity of a Factorial Survey Approach to the Analysis of Criminal Behavior. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 6(3), 139-146.
- Evans, S. C., Roberts, M. C., Keeley, J. W., Blossom, J. B., Amaro, Ch. M., Garcia, A. M., Stough, C. O., Cantera, K. S., Robles, R., & Reed, G. M. (2015). Vignette methodologies for studying clinicians' decision-making: Validity, utility, and application in ICD-11 field studies. *International Journal of Clinical and Health Psychology* 15, 160-170.
- Finch, J. (1987). Research Note: The Vignette Technique in Survey Research. *Sociology* 21, 105-114.
- Gould, D. (1996). Using vignettes to collect data for nursing research studies: How valid are the findings? *Journal of Clinical Nursing* 5, 207-212.
- Grøndal, P., Grønnerød, C., & Sexton, J. (2009). A comparative case vignette study of decision making in forensic psychiatric cases. *International Journal of Forensic Mental Health* 8, 263-270.
- Hughes, R. (1998). Considering the vignette technique and its application to a study of drug injecting and HIV Risk and safer behaviour. *Sociology of Health & Illness* 20(3), 381-400.
- Hughes, R., & Huby, M. (2001). The application of vignettes in social and nursing research. *Methodological Issues in Nursing Research* 37, 382-386.
- Jackson, M., Harrison, P., Swinburn, B., & Lawrencel, M. (2015). Using a Qualitative Vignette to Explore a Complex Public Health Issue. *Qualitative Health Research* 25(10), 1395-1409.
- Jasso, G. (2006). Factorial Survey Methods for Studying Beliefs and Judgments. *Sociological Methods & Research* 34(3), 334-423.
- Kirk, R. E. (1995). *Experimental Design: Procedures for the Behavioral Sciences* (Third Edition). Pacific Grove: Brooks/Cole Publishing Company.
- Langley, G. R., Titchler, D. L., Llewellyn-Thomas, H. A., & Till, J.E. (1991). Use of written cases to study factors associated with regional variations in referral rates. *Journal of Clinical Epidemiology* 44, 391-402.
- Ludwick, R., Wright, M. E., Zeller, R. A., Dowding, D. W., Lauder, W., & Winchell, J. (2004). An improved methodology for advancing nursing research: Factorial surveys. *Advances in Nursing Science*, 27(3), 224-238.
- Ludwick, R., & Zeller, R. A. (2001). The factorial survey: An experimental method to replicate real world problems. *Nursing Research*, 50(2), 129-133.
- Lunz, M. L. (1990). A methodological approach to enhance external validity in simulation based research. *Issues in Mental Health Nursing* 11, 407-422.
- Macpherson, C. C., & Veatch, R. M. (2010). Medical student attitudes about bioethics. *Cambridge Quarterly of Healthcare Ethics* 19, 488-496.
- McCarron, M., & Stewart, D. (2011). A Canadian perspective on using vignettes to teach ethics in psychology. *Canadian Psychology* 52, 185-191.
- Parsons, C., McCorry, Murphy, Byrne, S., O'Sullivan, D., O'Mahony, Passmore, P., Patterson, S., & Hughes, C. (2014). Assessment of factors that influence physician deci-

- sion making regarding medication use in patients with dementia at the end of life. *International Journal of Geriatric Psychiatry* 29, 281–290
- Peabody, J. W., Luck, J., Glassman, P., Dresslhaus, T., & Lee, M. (2000). Comparison of vignettes, standardized patients, and chart abstraction: A prospective validation study of 3 methods for measuring quality. *Journal of the American Medical Association* 283, 1715–1722.
- Raudenbush, S. W., & Bryk A. S. (2002). *Hierarchical Linear Models. Applications and Data Analysis Methods. Second Edition*. London: Sage Publications.
- Rossi, P. H., & Anderson, A. B. (1982). The Factorial Survey Approach: An Introduction. In: P. H. Rossi & S. L. Nock (Eds.), *Measuring Social Judgments: The Factorial Survey Approach*. Beverly Hills: Sage Publications.
- Rossi, P. H. (1979). Vignette Analysis: Uncovering the Normative Structure of Complex Judgments. In: R. K. Merton, J. S. Coleman & P. H. Rossi (Eds.), *Qualitative and Quantitative Social Research. Papers in Honor of Paul F. Lazarsfeld* (pp. 176–186) New York: Free Press.
- Rossi, P. H., Sampson, W. A., Bose, C. E., Jasso, G., & Passel, J. (1974a). Measuring Household Social Standing. *Social Science Research* 3, 169–190.
- Rossi, P. H., Waite, E., Bose, C. E., & Berk, R. E. (1974b). The seriousness of crimes: Normative structures and individual differences. *American Sociological Review* 39(2), 224–237.
- Samuelsson, E. & Wallander, L (2014). Disentangling practitioners' perceptions of substance use severity: A factorial survey. *Addiction Research and Theory* 22(4), 348–360.
- Sauer, C., Auspurg, K., Hinz, T., & Liebig, S. (2011). The application of factorial survey in general population surveys: The effects of respondent age and education on response times and response consistency. *Survey Research Methods* 5, 89–102.
- Schoenberg, N. E., & Ravdal, H. (2000). Using vignettes in awareness and attitudinal research. *International Journal of Social Research Methodology* 3, 63–74.
- Shah, R., Edgar, D., & Evans, B. J. W. (2007). Measuring clinical practice. *Ophthalmic and Physiological Optics* 27, 113–125.
- Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. Second edition. London: Sage Publishers.
- Steiner, P. M., & Atzmüller, Ch. (2006). Experimentelle Vignettendesigns in faktoriellen Surveys. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 58(1), 117–146.
- Steiner, P. M., Atzmüller, Ch., & Su, D. (2016). Designing Valid and Reliable Vignette Experiments for Survey Research: A Case Study on the Fair Gender Income Gap. *Unpublished Manuscript*.
- Steiner, P. M., Atzmüller, Ch. & Wroblewski, A. (2009). *Wahrnehmung, Bewertung und Messung geschlechtsspezifischer Einkommensunterschiede: Eine methodologische Untersuchung mittels Vignettenexperiment, einem traditionellen Fragebogen und Registerdaten*. Forschungsbericht für den Oesterreichischen Jubiläumsfonds, Wien , Österreich.
- Stokes, J., & Schmidt, G. (2011). Race, poverty and child protection decision making. *British Journal of Social Work* 41, 1105–1121.

- Su, D., & Steiner, P. M. (2016). An Evaluation of Experimental Designs for Constructing Vignette Sets in Factorial Surveys. *Unpublished Manuscript*.
- Taylor, B. J. (2006). Factorial surveys: Using vignettes to study professional judgment. *British Journal of Social Work* 36, 1187-1207.
- Veloski, J., Tai, S., Evans, A. S., & Nash, D. B. (2005). Clinical vignette-based surveys: A tool for assessing physician practice variation. *American Journal of Medical Quality* 20, 151-157.
- Wainwright, P., Gallagher, A., Tompsett, H., & Atkins, C. (2010). The use of vignettes within a Delphi exercise: A useful approach in empirical ethics. *Journal of Medical Ethics* 36, 656-660.
- Wallerand, L. (2012). Measuring social workers' judgements: Why and how to use the factorial survey approach in the study of professional judgments. *Journal of Social Work*, 12(4), 364-384.
- Welch, L. C., Lutfey, K. E., Gerstenberger, E., & Grace, M. (2012). Gendered uncertainty and variation in physicians' decisions for coronary heart disease: The double-edged sword of "atypical symptoms." *Journal of Health and Social Behavior* 53, 313-328.

Ärztliche Werthaltungen gegenüber
nichteinwilligungsfähigen Patienten

Ein Faktorieller Survey

Schnell, M.W.; Schulz, C.; Atzmüller, C.; Dunger, C.
(Hrsg.)

2017, IX, 141 S. 16 Abb., Softcover

ISBN: 978-3-658-16565-9